Title	Augmented Reality Using Visible Light Communication
Sub Title	
Author	Khanal, Prem Bhushan(Haruyama, Shinichiro) 春山, 真一郎
Publisher	慶應義塾大学大学院システムデザイン・マネジメント研究科
Publication year	2015
Jtitle	
Abstract	
Notes	
Genre	Thesis or Dissertation
URL	http://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40002001-00002015-0004

Master’s Dissertation 2015

Augmented Reality Using Visible Light Communication

Prem Bhushan Khanal

（Student ID Number：81334533）

Supervisor
Professor Shinichiro Haruyama

September 2015

Graduate School of System Design and Management,
Keio University
Major in System Design and Management

SUMMARY OF MASTER’S DISSERTATION

Student Identification Number	81334533	Name	Prem Bhushan Khanal

Title

Augmented Reality Using Visible Light Communication

Abstract

In this thesis we propose the use of visible light communication to implement Augmented Reality System.

Augmented Reality (AR) is an interactive, real time system that combines virtual information over real graphical information. Visible light communication uses light in visible light spectrum between 400 to 800 THz as a medium of communication. Existing methods of achieving augmented reality using visible light communication either use a custom made image sensors to detect light position or do not overlay virtual information over real environment accurately. In addition some existing solutions make use of color markers which are useless in the dark. We propose an AR system using Light Emitting Diodes (LEDs) installed on target objects as markers to display the virtual information accurately. This type of AR system is particularly useful to mark large objects like exterior of a building or its sections for example each floor or each window. Each light source transmits information which can be read using image sensor in a camera and determine the exact position where virtual information needs to be displayed. The virtual information can be predetermined or may depend on the information received in the form of light. An algorithm has been developed to detect the position of LEDs and project a predetermined virtual information in camera live feed.

A prototype has been developed to evaluate the algorithm using custom designed programmable LED lights, Point Gray Grasshoper3 camera, NVidia Terga K1 GPU running Ubuntu 14.04 to aid intensive computation and C++ with OpenCV library for image processing.

Key Word (5 words)

Augmented Reality, Visible Light Communication, LED markers, Visualization of large buildings, Outdoor

Acknowledgement

I would like to start by thanking SDM for providing me the wonderful opportunity to study system thinking, design thinking and project management concepts as part of my master's degree. I am grateful to my supervisor Prof Shinichiro Haruyama for guiding me and giving me challenges which I never knew I could pull off. I would like to thank Assistant Professor Hideaki Uchiyama from Kyushu University for his insightful lectures on computer vision and OpenCV. I would like to thank all the past researchers who have done research in the field of Visible Light Communication (VLC) and Augmented Reality (AR). Studying past research work has been inspiring. I would like to thank all the teachers at SDM for teaching us in a friendly and fun environment. I thank all my friends and seniors at SDM for interesting discussions, knowledge sharing and group works. I am thankful to all the people who spared time to fill in my survey and helped me to complete my research work. I am thankful to my wife Deepa for supporting me in her own way.

Finally I would like to thank my secondary supervisor Prof Takasi Maeno for reviewing my thesis and guiding me for the improvements.

Table of Contents

List of Tables:

List of Figures

1 Introduction

1.1 Augmented Reality

Augmented Reality (Hence forth AR) is basically superimposing virtual information over real information. It is a live, direct or indirect view of a physical, real world environment whose elements are augmented by computer generated sensory input such as sound, video, graphics or GPS data. AR has been defined to have following three properties (Azuma, et al., 2001)

a) Combines real and virtual objects in real environment
b) Runs interactively, and in real time; and
c) Registers (aligns) real and virtual object with each other

Researchers have not constrained AR just into visual information. It encompasses other senses like hearing, touch and smell as well. AR not only augments reality with additional information but also it can be used to remove some objects from the real environment. This division of AR is sometimes termed as 'Mediated or diminishing reality' (Azuma, et al., 2001). This research focuses on visual AR system only.

AR is often confused with virtual reality (VR) system. The distinction has been well explained by (Milgram & Kishino, 1994) in following Virtuality Continuum diagram.

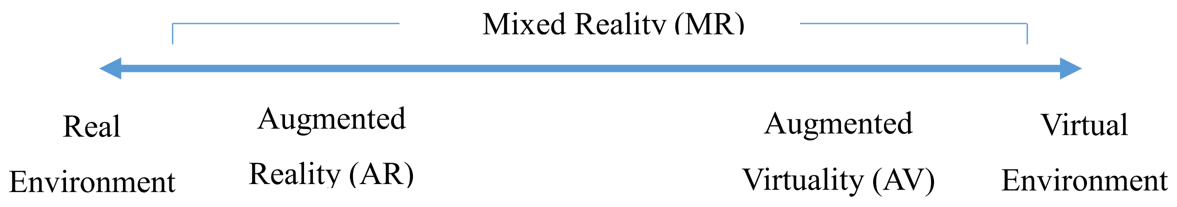


Figure 1: Simplified representation of a "Virtuality Continuum"

In augmented reality some virtual information is added to real environment for the specified purpose where as in augmented virtuality some real information is added to virtual environment.

AR is not a new concept as the first AR system dates back to late 1960s when Sutherland used a see through head mounted display to present 3D graphics (Sutherland, 1968). AR technology has come a long way since then. Many successful commercial applications use the concept of AR.

AR enriches user experience by displaying the information that user can not perceive directly. The virtual information can help users to perform their real world tasks easily. (Azuma R. T., 1997) listed following six classes of potential application of AR

a) Medical Visualization
b) Maintenance and repair
c) Annotation
d) Robot path planning
e) Entertainment
f) Military aircraft navigation and targeting

These potential applications were envisioned by Azuma back in 1997. With advent of powerful tablets and cellular phones, now we can see widespread usage of AR in many other sectors

a. Advertising
b. Mapping
c. Education
d. Monitoring
e. construction

These are some additional areas where AR is being used effectively or researchers are trying to develop AR system to enhance user experience. Figure below shows a typical use of AR in a LEGO shop where the buyers can see the actual objects that can be built from LEGO on screen.

Figure 2: Typical Application of AR in a Lego Shop
(http://ilovenewwork.com/2009/10/02/lego-ar/, n.d.)

As we can see the buyer can see a helicopter on top of the box if she brings the box in front of the camera. This is an example of really successful usage of AR.

Sekai camera introduced a revolutionary concept that is social networking using augmented reality system. Users of Sekai (World) camera can tag real world objects with digital post-it notes (CrunchBase, 2015). Information pertaining to the time and place is displayed on screen while looking through the camera view on phone or PDA devices. Basically Sekai camera offers two services

a. User generated content features
b. Social communication features

User Generated Content Features: Firstly It allows users to tag locations with multimedia information (text, photos, videos or voice) termed as Air Tag. Secondly users can look for the relevant air tags specifying search criteria like time, type and distance termed as Air Filter. Lastly multimedia bookmarks that could be viewed when users are not at the location where the tags were posted termed as Air Pocket.

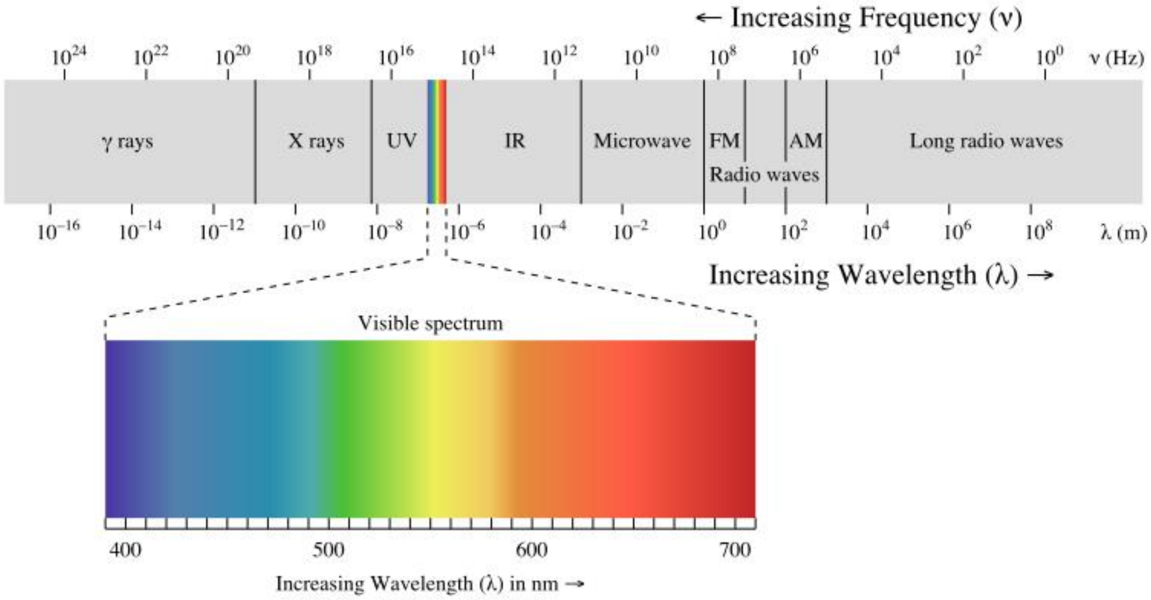
Social Communication Features: One's own permanent timeline where one can view self as well as friend's activities termed as Sekai Life. User's profile with some custom message could be attached above users termed as Air Profile. Tweets could be viewed as floating air tag termed as Air Tweet.

Sekai camera is unique and cool application of Augmented Reality worth mentioning. Although it tags real users and places with virtual information but this application doesn't require precise alignment of real and virtual objects.

Figure 3: Typical Sekai Camera View (TechChurch, 2010)

1.2 Visible Light Communication

Any communication system has three basic components, sender, receiver and medium of data/information transmission. Visible Light Communication (VLC) uses light in visible light spectrum between 400 to 800 THz (380 to 780 nm) as medium of information transmission (Wikipedia, Visible Light Communication, 2015).



Light Emitting Diodes (LEDs) are predominantly used in visible light communication because they require low power, have high switching speeds and longer lifetime as compared to incandescent lamps (apparently up to 40 times) (Haruyama, 2013). The receiver in visible light communication is either photodiode or an image sensor. Image sensor inside a camera is itself a set of large number of photodiodes and can be used to receive the information sent in the form of light. A photodiode detects the signal intensity of visible light while an image sensor not only detects the incoming signal intensity but also the angle of arrival of light emitted from a visible light transmitter such as white LED light (Yamazato & Haruyama, 2014).

Visible light communication has a lot of limitations which can be exploited to be useful in some novel niche. Following table lists the pros and cons of VLC (though I don’t claim it to be a comprehensive list):

Table 1: Advantages and Disadvantages of Visible Light Communication

S.N	Disadvantages of VLC	Advantage of VLC
1	It requires line of sight between LED light and receiving sensor. It can be used up to about 100 meters only.	The property that it requires line of sight can be used in favor of VLC in security related applications and in graphical user interfaces that combine visual imagery with visible light communication.
2	Bandwidth is limited to 10 mbps to few 100 mbps	The band above 3 THz is not regulated and is free to be used by anyone.
3	Range and flexibility are really limited as compared to other wireless communication.	Unlike other wireless communication, the source of transmission could be located easily.
4	People would like to control luminosity of the lights and may not like the idea to turn on the lights even during the day time which could be the major drawback in using LED lights for communication.	LED light has become an integral part of infrastructure. Adding communication function would make easy and multipurpose ubiquitous communication system.
5		Unlike RF spectrum which is supposed to give rise to neurological disorders on contrary visible lights do not have any evident negative effects in human health.

As evident VLC can be a ubiquitous communication medium which could be used in conjunction of other communication technologies like wired or radio wave communication.

2 Meta-Analysis

In this section I would analyze the research work that has already been done in the field of Augmented Reality and Visible Light Communication and present the gaps which this research is trying to fulfill. Let us start with research works on augmented reality till date.

2.1 Augmented Reality Research Works and Techniques

2.1.1 Augmented Reality Techniques

Azuma conducted a milestone survey in 1997 (Azuma R. T., 1997) to demonstrate the technology behind AR systems, their application areas and limitation. The survey identified a comprehensive list of application areas of augmented reality and underlying technology to realize the AR system. Basically AR utilizes two technologies to perceive the real environment. The first is known as optical technology which uses a head mountable see through display (henceforth HMD). HMD uses semi reflective optical combiners to display augmented information. And real environment is perceived allowing controlled proportion of light to the user eyes. Figure below displays a typical Optical AR device.

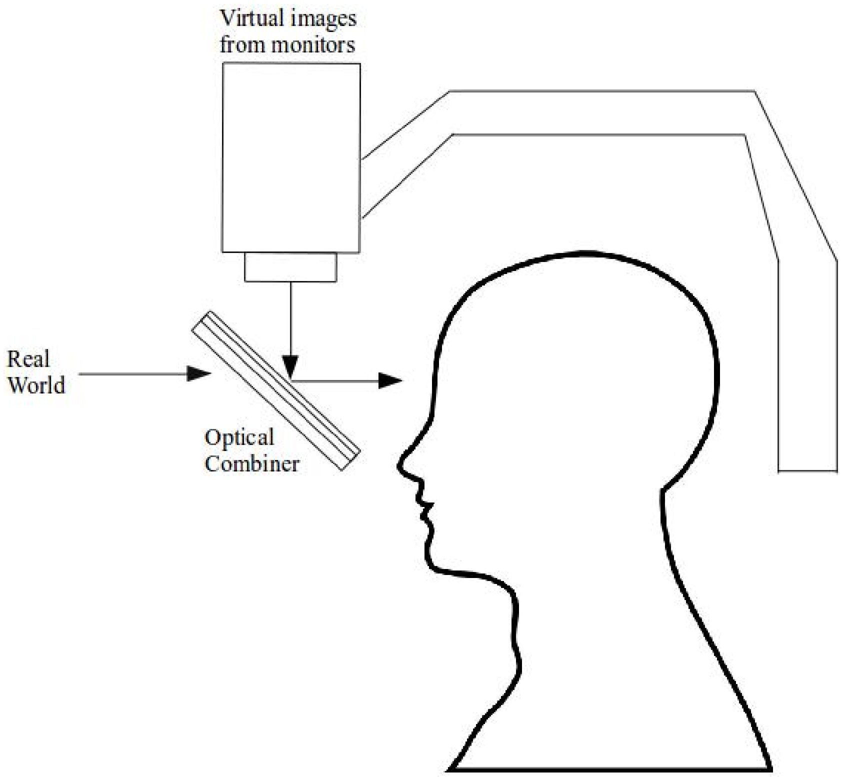


Figure 5: Optical see through HMD design (AR, 2015)

The second is video technology which uses two different streams of information. Real environment is perceived through a camera and multimedia information is augmented on the video stream in real time. The final graphics is displayed on some display, be it a closed HMD or a monitor screen.

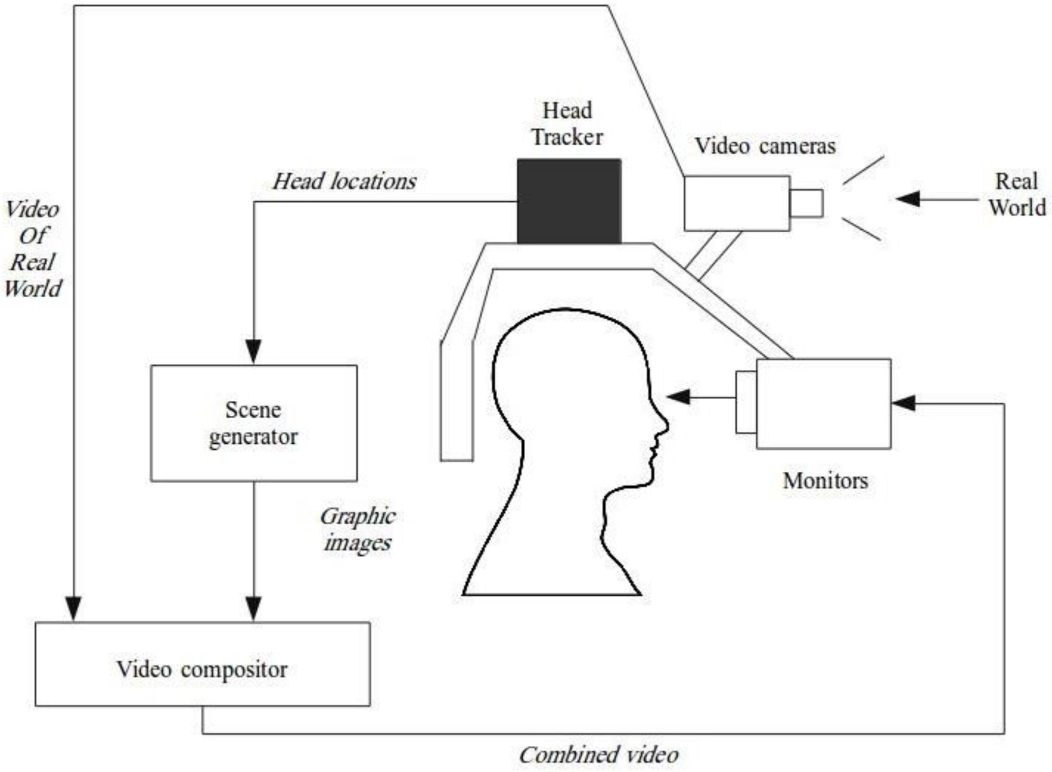


Figure 6: AR using Video Technology and Opaque HMD

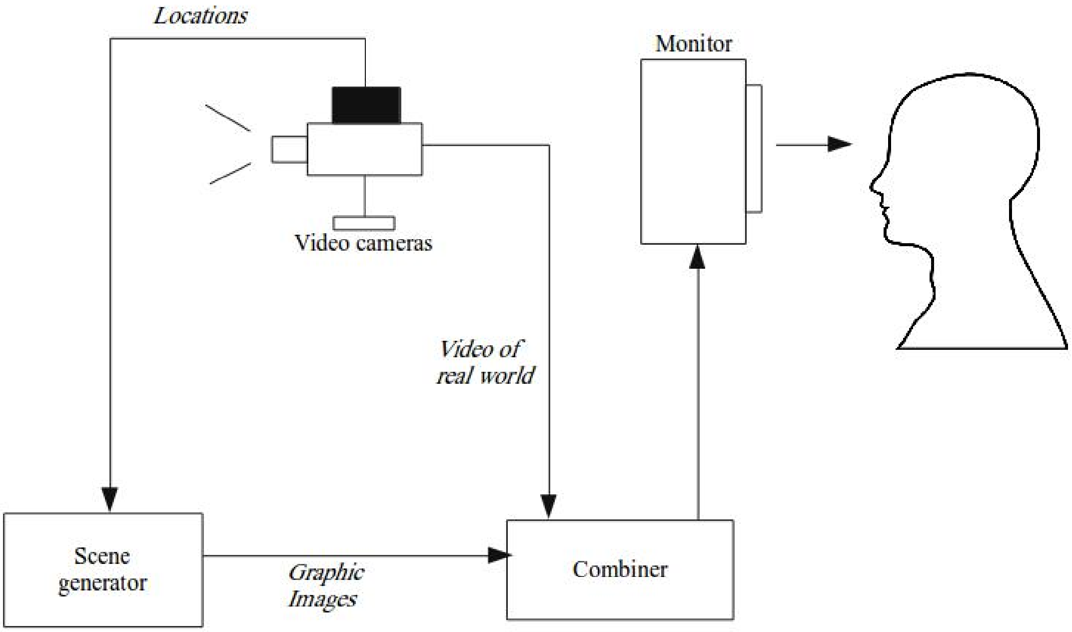


Figure 7: AR using Video Technology and Display on Monitor

Both techniques have their own pros and cons. Optical technology is simple, safe, real and less complicated whereas video technology is more complicated but flexible and virtual information could be blended in real information seamlessly. This research focuses on video technology only because of the flexibility it provides. The information in the form of visible light could be read using image sensor of camera by manipulating the instream at pixel level.

2.1.2 Distinction between Virtual and Real

In augmented reality system a clear distinction should be made what is virtual and what is real. We would adapt the definition by (Milgram & Kishino, 1994) which has been tabulated as follows:

Table 2 Real V/S Virtual Objects

SN	Real	Virtual
1	Real Object Vs Virtual Object	
	Real Objects have an actual objective existence	Virtual objects exist in essence or effect but not formally or actually
2	Direct Vs Non-direct Viewing	
	Direct viewing of real object is through glass or air. Non-direct viewing is sampling data about an object and resynthesize or reconstruct data on some display medium.	Virtual objects cannot be sampled directly and thus can only be synthesized directly.
3	Real Vs Virtual Image	
	Any image that has luminosity at the location at which it appears to be located can be classified as real image	Any image that doesn't have luminosity at the location at which it appears can be classified as virtual image.

2.1.3 Tracking and Registration Techniques

One of the major problems in augmented reality solutions is registration of virtual object over real object accurately. This becomes more complicated in video technology as the three dimensional object co-ordinates, two dimensional real image co-ordinates and two dimensional virtual information need to be aligned with each other. Registration requires tracking the object or area of interest in environment. Tracking technologies could be grouped into following three categories (You, Neumann, & Azuma)

a. active-target
b. passive-target
c. Inertial

Active-target: Active-target systems incorporate powered signal emitters and sensors are placed in a prepared and calibrated environment. Examples of such systems use magnetic, optical, radio and acoustic signals.

Passive-target: Passive-target systems use ambient or naturally occurring signals. Example of such systems include compasses sensing earth's magnetic field and vision systems sensing intentionally placed fiducials (for example circles and squares) or natural features.

Inertial: Inertial Systems are completely self-contained, sensing physical phenomena created by linear acceleration and angular motion.

Two or more tracking techniques could be combined to increase the accuracy in principle but researches have shown it may increase point of errors in turn deteriorating accuracy. Identifying the exact position in real environment where the virtual object needs to be placed is another challenge. Existing technologies for registration are:

a. Image Processing: This is the most common marker less technique used in many AR solutions. Basically key points and features of some object or place in real environment are identified and virtual object co-ordinate system is converted to world co-ordinate system. Then virtual object is superimposed in real environment.
b. Markers: External markers are used to locate the real environment. Bar Codes, QR codes are popular markers to identify the real world objects. Some applications found to be using color markers as well.
c. Signals: GPS signals, Magnetometer, are used to identify the places and objects for registration.

AR technology has come a long way with advancements in above listed technologies.

2.2 Existing AR solutions using visible light communication

In this section previous research works and products using visible light communication to realize AR system are studied. The gap is analyzed at the end

2.2.1 Casio Picalico

Casio developed an interesting and flexible system which can be used for both simple visible light communication and AR solution. Like in any communication system the transmitter can be anything which can change colors in defined frequencies. Be it color LED or be it a computer or mobile application. The information is sent by color markers that use three colors (red, green and blue) and black off. Per item in the information is represented by 24 pulses. The receiver is camera in smart phone or computer (Casio, 2015).

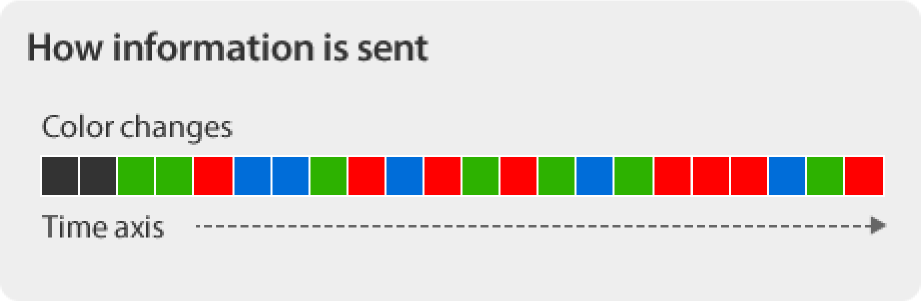


Figure 8: Information in Picalico (Casio, 2015)

There are many advantages of this solution like it can receive information from multiple items in the same field of view. And then it allows the user to choose the desired information among the received information. Picalico allows contents to be superimposed in real scene. It is a multipurpose free software developed by Casio which is claimed to have many applications including machine repair and malfunction notification. The only issue with Picalico for AR solution is that the information is not aligned to the real world object.

2.2.2 AR System Using Colored Spheres

A solution using colored spheres has been proposed for visualization of large buildings (Robler, Rogge, & Hentschel, 2011). This is a marker based solution which requires installation of colored balls. The color of the balls need to be highly saturated and should be in contrast with the environment. The spheres have been proposed as markers mainly because of following two reasons:

a. they project circle on image sensor regardless of perspective and

b. distance between each sphere (vertex of the marker) could be adjusted in ratio of target location and virtual object size.

In order to accurately locate the spheres Color Invariance Edge Detection (CIED) approach has been adapted. CIED considers color information which would increase the robustness against changing light conditions. Then the sphere can be easily located by applying a threshold to obtain a binary contour image. An algorithm has been developed to accurately predict the position of sphere. These position become four corners of marker in 2D image space.

Henceforth the camera pose is estimated using ARToolkitPlus, since their positions in real world coordinates are known. Once the pose estimation is done, transformation matrix is applied to the virtual scene in order to obtain the same perspective as the camera. Finally the resulting image is superimposed with real environment. The demonstration worked at a maximum distance of 25 meters with ideal circle of sphere to be 50 contour pixels. The authors claim that the limit could be overcome by using larger color spheres or by increasing the resolution of the image sensor.

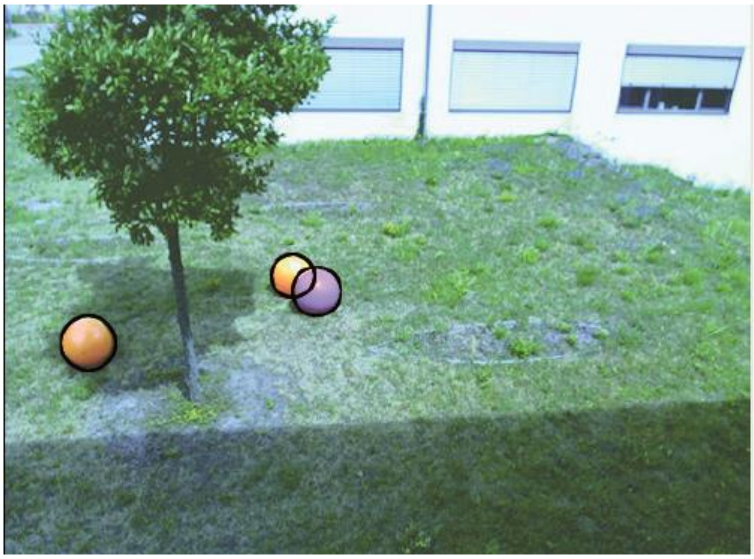


Figure 9: AR System Using Colored Spheres (Robler, Rogge, & Hentschel, 2011)

2.2.3 ID Cam System

ID Cam uses a high speed image sensor to capture images like an ordinary camera and detects the ID of a beacon emitted over a long distance (Matsushita, et al., 2003). The blinking beacon transmits a time series data that doesn’t change over distance unlike visual code which is a space data. Each pixel in image sensor grid is equipped with receivers capable of decoding beacon data. Together beacon and specialized image sensor system have been termed as “ID camera system" by the authors. Although the ID camera system does not mandate the synchronization between blinking pattern of beacon and camera frame rate, the blinking frequency should always be less than camera capturing speed. The authors emphasize on following three major features of ID cam:

a. Robustness for disturbance light
b. Simultaneous recognition of multiple beacons
c. Recognition of beacon over long distance.

Robustness for disturbance light: Each pixel is equipped with special circuitry to detect beacon which means even if one pixel fails the other will detect the beacon as one beacon would expand over many pixels. This failover mechanism also serves for the purpose of noise filter as well because view angle for each pixel is small for ID camera and the unaffected pixels can recognize beacon accurately.

Simultaneous recognition of multiple beacons: The ID of multiple beacons could be identified simultaneously even if two or more beacons have the same id. It is claimed to recognize as many as 255 beacons simultaneously.

Recognition over long distance: The output of beacon is relative to the distance over which it is to be recognized. Larger beacons could be recognized over a longer distance and vice versa.

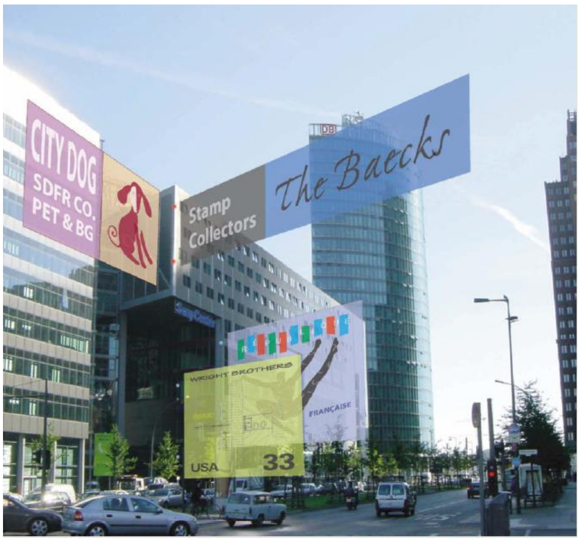


Figure 10: ID Cam Simulation (Matsushita, et al., 2003)

In order to realize the AR system using ID camera the authors recommend replacing conventional camera in PDAs with ID camera. Thus it could operate in two modes normal image capture mode and ID recognition mode. The beacons need to be installed in target environment and users could enjoy AR system in ID recognition mode of the camera. The authors claim that since the circuitry of ID recognition system is simple it is not difficult to design small ID cams for PDAs.

2.2.4 Encoded LED System for Optical Trackers

In this research work an algorithm has been developed to detect LED position without synchronizing blinking frequency and camera frame rate (Naimark & Foxlin, 2005). The authors utilize the fact that the pixel brightness decreases monotonically with distance from the centroids. Centroid is the exact LED location in the image which contains saturated pixels with values 255. In time series camera frames with brightness levels $L_i<L_j<L_k$; s_k is the number of saturated pixels for L_k. Then brightest s_k pixels are excluded in the neighborhood of centroid in all three pixels. For the frame L_k this removes all and only the saturated pixels remain where as in other frames saturated ones and high brightness non-saturated ones remain. A working algorithm has been developed based on this logic which can decode blinking LEDs with reasonable accuracy from 0.5 to 3 meters in range. Practically a single LED could be detected with 90 to 92% of accuracy.

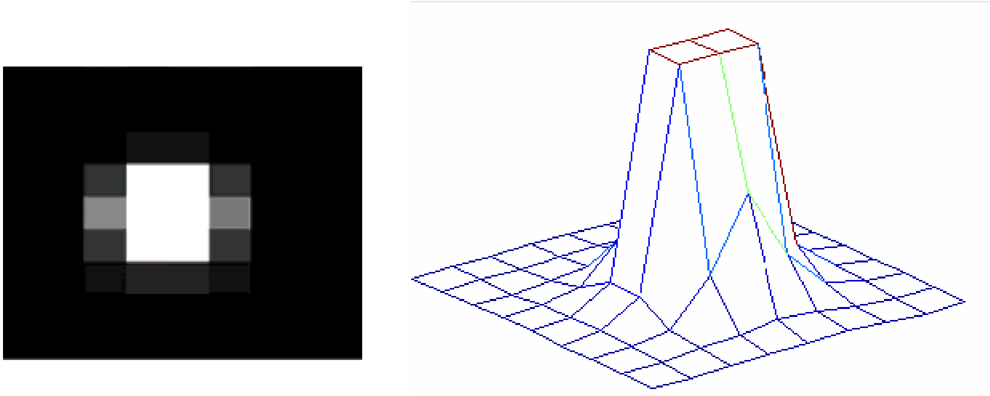


Figure 11: 6 Pixel Saturated LED. Close-up (left), 2-D grayscale map (right) (Naimark & Foxlin, 2005)

2.3 Problems in Existing solutions

Last section gives an overview of the past research works that have been done in the field of augmented reality using visible light communication. In this section we analyze the shortcomings of these systems. Picalico is a flexible system but virtual information is not aligned with real environment. Colored Sphere method would not work in the dark and low light conditions. ID Cam system demands a specially designed ID camera. Encoded LED system suffers from lack of accuracy if the distance is more than 3 meters.

Hence the proposed Augmented Reality system using visible light communication

a. Should not use flashing lights
b. Should align virtual objects with real environment accurately.
c. Should work in all lightning conditions
d. Should be able to use general purpose high speed camera
e. Should work over a longer distance.

On the foundation of these findings the research objective has been formulated in next section.

3 System Design

3.1 Research problem (Social)

Not much advancement has been seen when it comes to the displaying floor-wise information in the exterior of the building. Traditional hording boards and sign boards are troublesome to install, make building look ugly if installed too many and are difficult to update frequently. Alternatives like LCD displays or other types of digital displays require special installation and maintenance and are expensive as well. Hence displaying virtual information can be one of the alternatives. A survey was conducted among thirteen people living and working in eight different countries. People were asked if they knew methods of displaying information on the exterior of a building other than sign boards. Following table shows their responses:

Table 3: Survey on Information Display Methods on Building Exteriors

SN	Methods of displaying information on exterior of a building other than sign board
1	LED Display
2	Interactive Display
3	Posters
4	Digital TV
5	Projecting Film, Video, Picture, Text on the wall
6	Digital Display Boards
7	LED Media Facades
8	Projection Mapping

31% people responded that they don’t know any other alternative to displaying information using sign boards. As can be seen same kind of traditional methods for information display are used around the globe. AR can be an alternative to display contextual information and advertisement on the exteriors of large artifacts e.g. buildings or monuments. The need to use a camera (whether cell phone camera or some special camera or some see through device) would increase some risks of distraction and accident but the potential usefulness outweighs the risks. If a discount scheme on a product could be seen standing outside of a shopping mall, or a barber shop could located inside a multistory building in a foreign land where nobody understands the language you speak, it would be a savior.

3.2 Research Goal and Objective

The goal of this research is to develop an AR system using visible light communication based on video technology (refer to Figure 7). This goal could be further divided into following measurable objectives based on definition of AR system.

a. To use LED lights as markers for AR system. To be specific install LED lights in real environment and mark an area in real image plane (live video feed on display screen) defined by LED lights.
b. To locate LED lights in real environment. To be specific identify the pixel coordinates of LED lights on real image plane (live video feed on display screen).
c. To overlay a custom virtual information over LED markers accurately on real image plane in real time. To be specific a rectangular image (virtual information) should augment live feed from camera (on display screen), so that the vertices of the rectangular image align with LED lights in real image plane. To be further precise each vertex of rectangular image should be within the of pixel points occupied by corresponding LED light on live feed image plane. Measurable definition of Real-time is important here. Real-time computer systems are able to deal with and use new information immediately and therefore influence or direct the actions of the objects supplying that information (Cambridge, 2015). To be compliant with this definition, once the positions of LEDs are identified the information would be used to augment immediately next frame of camera feed, the information won’t be used for the past frames of camera feed.

These objectives are consistent with definition of Augmented Reality System in section 1.1. Real object and virtual objects in objectives (a), (b) and (c) are consistent with the definition in section 2.1.2.

3.3 User Needs

Before designing any system understanding of what end users expect from the system is important. A survey was conducted where users were explained what AR system is. The explanation included a conceptual picture which explained how AR system using visible light communication would look like. Then users were asked their expected characteristics and functionality of the system. Following table list the complete list of user responses.

Question: What should be the characteristics of AR system using visible light communication?

Table 4: List of User Requirements

SN	User Needs
1	It should change its behavior according to surrounding condition. For example day time and night , High and low temperature
2	There should be an option to add the sound to moving pictures or movie and change the contents
3	It should resemble the real world objects
4	Displaying in color is always helpful but we should not forget about the color blind people as well
5	It should always be less bright than the actual object and should not distract the background
6	I always think it as a 3D layer above the actual surface and very near to eye
7	It should always be easy to turn on and off and should have options to turn on/off certain features like text only, with full description etc. A perfect system will be that one which can display more and less information according to zoom in and out
8	The information should be transparent (user can set the opacity level) and not overlapping with the main image
9	The information should be accurate (specific to which exact point of image that we want to see)
10	The information should not override privacy information
11	Visible from 10 meters or less
12	Information should be visible across the street (distance depends on street size)

SN	User Needs
13	Information panels should not overlap (would be very messy)
14	There should be categories, for example display only restaurants, only hairdressers etc.
15	Information should be displayed in color
16	If the shop/company has a promotional video introducing their company, there should be a link. Maybe it can even auto play, depending on user preference
17	It should be real-time, interactive and 3D
18	In near future lots of information will be in AR system. The device/system should be able to offer filters
19	Dynamic display –change of elevation/dimension in reference to user position/location
20	Adjustable size of information display area
21	Personalized messaging. Suppose a university uses AR system and AR information is on white board but only intended student could see it
22	The system should be able to prompt for user input
23	It should limit user input to a few selected channels.
24	It should be able to emulate sensory feedback such as sound, touch, vibration etc.
25	Visual perception is at the core of VR. Thus, it should be able to display visual representation pertaining to the depth, accuracy, field of view, and critical fusion frequency
26	The system should have a zoom in and zoom out functionality
27	Should overlay video ads

These user needs were obtained from people living in five different countries, hence could be generalized. Some of these needs are out of scope of this research work which would be excluded while drawing system boundary in context diagram. Others will create the foundation of functional and non-functional requirement.

3.4 System Boundary

It is really important to define what is in the scope of this research and what is not in the scope of this research. We would define the system boundary based on the objectives of this research and user needs obtained from the survey.

a. As objective clearly defines this research focuses on video technology, see through panels are out of scope of this research. The technique developed by this research to locate LEDs in real environment could be used by see through display panels as well but the deployment strategy would be completely different.
b. This research would focus on two dimensional virtual information only. Some additional steps are mandatory to make this system work with three dimensional overlay.
c. The system should work in both indoors and outdoors.
d. The system would rely on camera image sensor to locate LED lights. A lot of research work has already been done to identify the camera pose and user tracking which could be combined with this research to expand the scope or for better results.
e. This system relies on information transmitted in the form of light using LED lights. The scope of this research includes the data structure of transmitted information. Which means the system would check if the data transmitted is compliant to the data structure defined in the system, if yes, it would decode data and display virtual information based on defined logic else it will just ignore the light source. The circuitry of LED light, configuration technique, programming techniques are out of scope of this research.

Once logical and functional system boundary is identified, context diagram of the system could be drawn as shown in Figure 12 below.

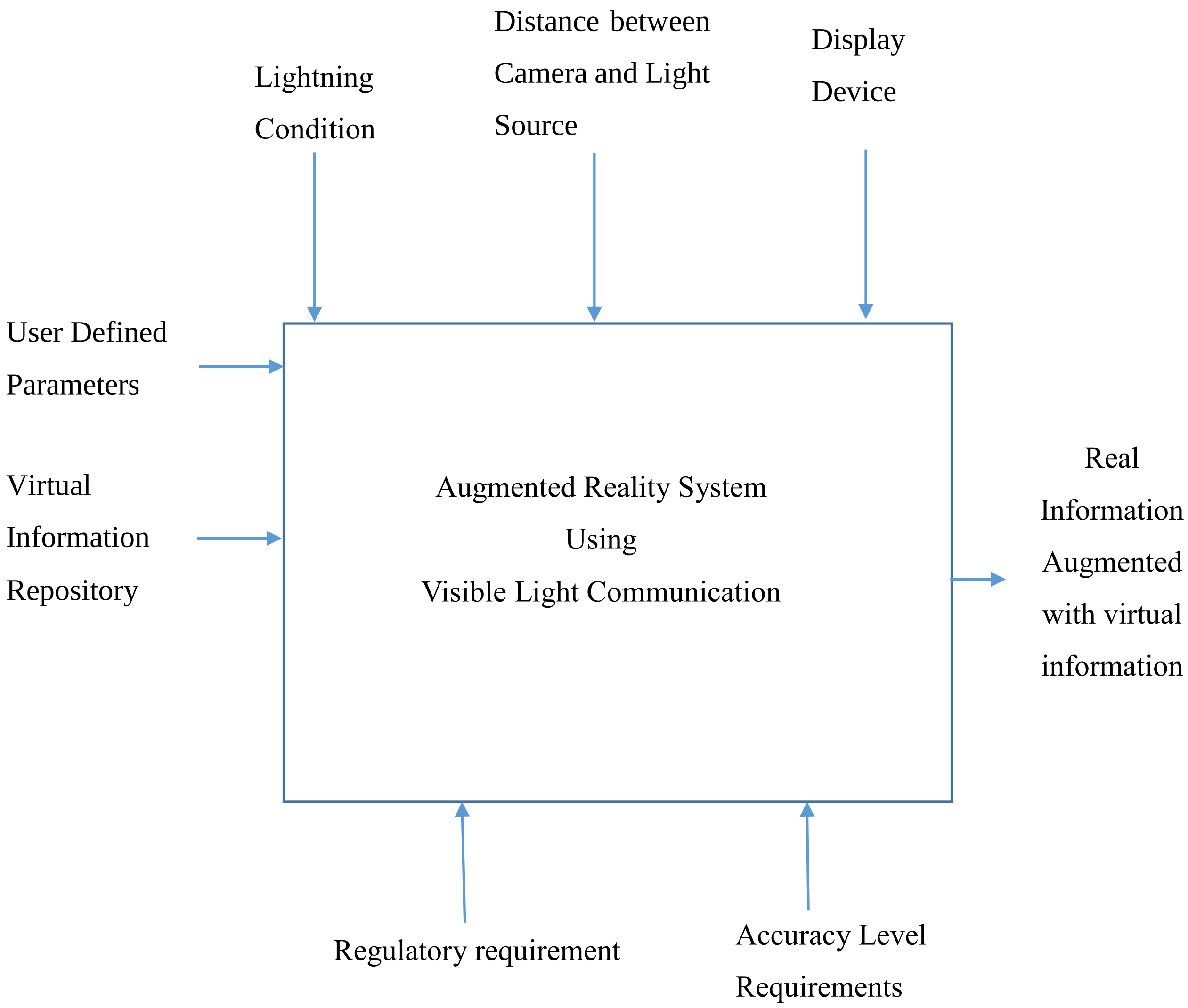


Figure 12: Context Diagram

As can be seen in the context diagram, the system takes input from users which would control core system behavior. Virtual information repository could be a local or remote repository compliant to the system specification. Lightning condition affect both accuracy and operating distance of the system. Farther the distance between light source and camera, lower is the probability of identifying the location of lights on real image plane. Display device influences the perceived quality of both real and virtual information as it is the sole interface between end user and system. The system should be compatible with certain regulatory requirement like the lights should not be flashing at some frequency range which is more likely to trigger photosensitive epilepsy. Photosensitive epilepsy is a medical condition in which seizures are triggered by visual stimuli that form pattern in time and space, such as flashing lights, bold regular patterns or regular moving patterns (Wikipedia, Photosensitive Epilipsy, 2015). Some applications may require high level of accuracy and for some lower level of accuracy is sufficient. Accuracy level requirement of would influence the system design itself. Finally the system produces the output for the end user by superimposing virtual information over real information.

3.5 System Requirements and Assumptions

Till now precise system objectives have be identified, a user survey has been conducted to find out user needs, system boundaries have been defined and a context diagram has been drawn. This section relies on all aforementioned sections to derive requirements and assumptions. Broad objectives have been divided into multiple objectives where needed and multiple requirements have been derived from a single objective. Out of scope user needs have been filtered out. Alike user needs have been merged.

Basically requirements and assumptions have been categorized in three major categories:

a. Functional Requirement/Assumption:

   Functional requirements are the things the system is expected to do right. Similarly assumptions are the intrinsic system behavior of system that are considered to exist.

b. User Requirement/Assumption:

   User requirements are the things that the stakeholders expect the system to do. Similarly assumptions are user perceived intrinsic system behavior.

c. Contextual Requirement/Assumptions:

   Contextual requirements are the things system is expected to do by external entities e.g. regulators. Similarly the assumptions are environment perceived intrinsic system behavior.

Table 5: Objectives, Requirements and Assumptions Mapping

Objective	Functional Requirement	User Requirement	Contextual Requirement	Functional Assumption	User Assumption	Contextual Assumption
Use LED lights as markers	LED lights must be blinking at predefined frequency	Adjustable size of information display area	LEDs should not be flashing at some frequency range which is more likely to trigger photosensitive epilepsy			
	Data Structure of the information transmitted should be known in advance	Information panel should not overlap			It should resemble real world objects	
Locate LED lights in real environment	Camera frame rate should be in multiples of LED frequency	Information should be visible across the street (visible from 10 meters or less)		LEDs must be visible	Should use a general purpose camera	
		Should work in dark as well as daytime		There should be no movement between camera and LED lights		
Overlay virtual information over LED markers	Location of all four LEDs must be known	Multimedia virtual information should be available	Virtual information should not override privacy information	Should function in real time	Should work for color blind people as well	

Objective	Functional Requirement	User Requirement	Contextual Requirement	Functional Assumption	User Assumption	Contextual Assumption
Overlay virtual information over LED markers	Storage location of virtual information must be known	Information should be displayed in color				
	The virtual information would adapt to the depth and angular displacement of lights in real environment.	Option to control the opacity of virtual information				
		Should be easy to turn on and off the information				
		Option to filter the specific information only				
		Information should be adjusted to user's location				

These requirements and assumptions lay the foundation of functional system design. System verification would be carried out based on these requirements and assumptions

3.6 Subsystem Diagram

This system basically has three subsystems, the LED light subsystem, which defines the area where virtual information needs to be displayed and transmits information in the form of light. High speed camera is another subsystem which receives the information from LED lights as well as surroundings (real environment). The third subsystem is computing unit, it would process the information received, and display reality augmented with virtual information.

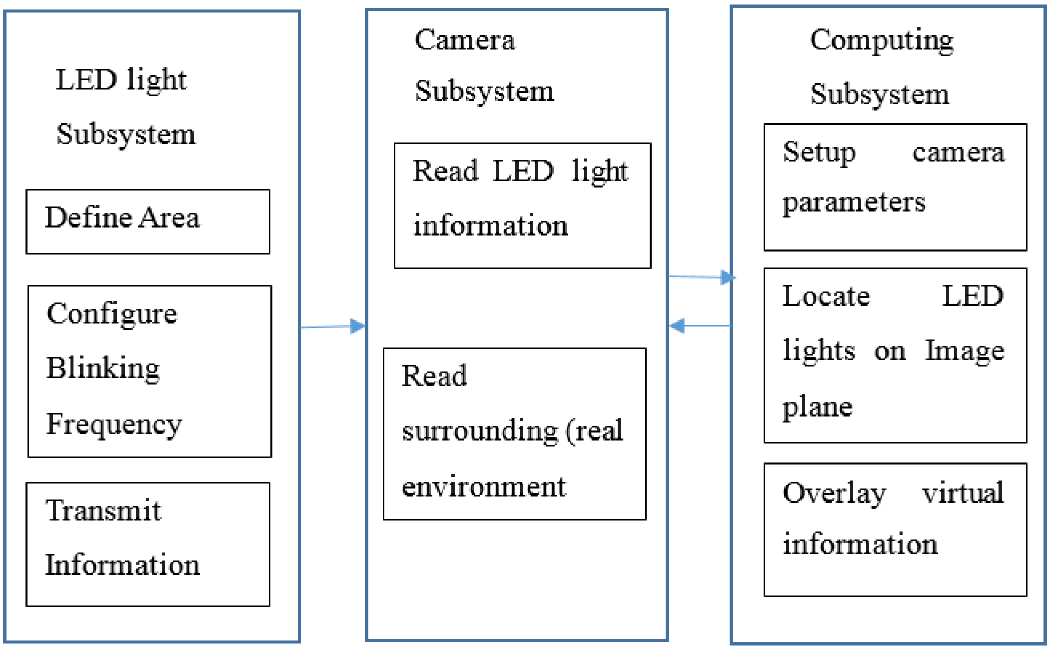


Figure 13: Subsystem Diagram

3.7 Functional Design

Based on requirements and assumptions listed in previous section the functional flow diagram of the system is developed. The system should exhibit following functionalities:

a. Identify target area
b. Install LED lights on target area
c. Transmit id in the form of light
d. Synchronize camera frame rate and LED frequency
e. Focus camera on target area/object
f. Locate LED lights in live camera feed
g. Fetch virtual information
h. Filter information based on user preference
i. Overlay virtual information on camera feed

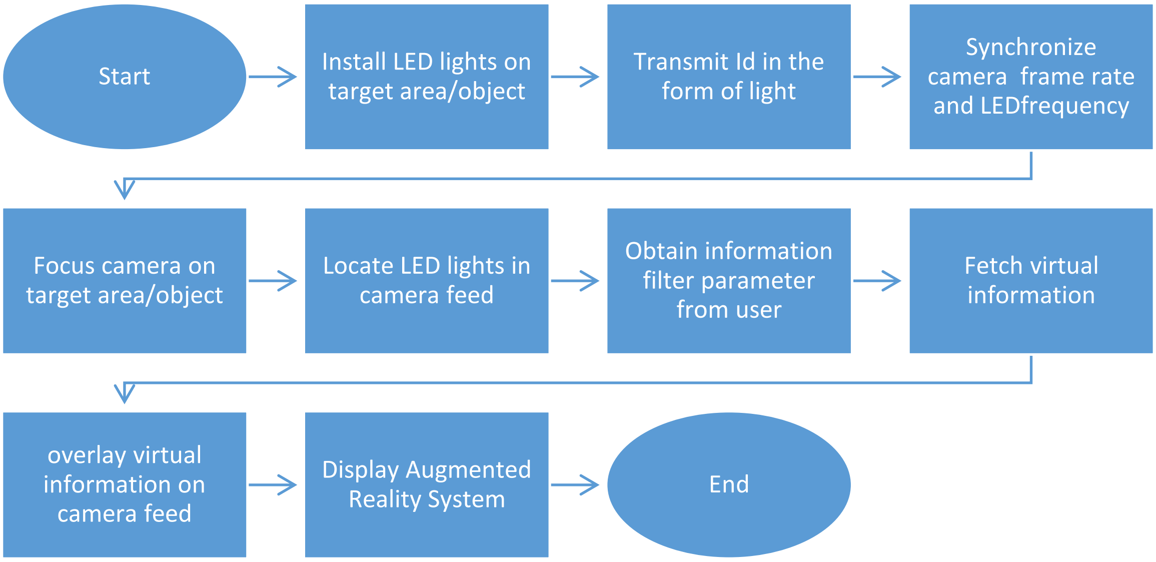


Figure 14: Functional Flow Block Diagram

To start with, LED lights should be installed on target area. This system is flexible enough to cover a large area or a small area depending on where LEDs are installed. Each LED should be blinking at certain frequency which is not detectable by human eyes. They transmit id or data in predefined form continuously. When camera is focused on that area, live feed from camera is processed to locate LEDs in picture frame. Meanwhile users can choose what kind of information they are interested in and if user criteria matches the context, virtual information is displayed over live camera feed.

3.8 Conceptual System

Figure below shows the conceptual sketch of the system we intend to develop. A person is standing in front of a building with his PDA camera pointed towards the building. The building has LED lights installed as markers for AR system on the third and fourth floors. The camera recognizes the lights and displays an additional information on screen, on top of live camera feed. Here the real environment has been augmented by additional virtual information. This has benefits to both the parties. The business establishment in the building which would like to attract more customers or provide information to more and more people without installing some permanent sign board or digital display gets benefited. On the other hand the customer who is keen to know what is inside the building or what it has to offer without going in gets benefited as well. This shows one of the possible application areas of AR using visible light communication.

Figure 15: Conceptual Sketch Diagram of the System

4 Algorithm Development

This section discusses on how the system could be realized computationally. If LEDs are placed in an environment where no other object is brighter than LEDs we can simply locate them checking the brightest spot on the scene. This principle could be used in darker environments and where identifying individual light is not important. This work develops an algorithm to detect the location of LED lights in two dimensional image space by checking the blinking pattern and then overlay the virtual information at the position detected. The limit in number of LEDs comes from

a. The blinking pattern of LEDs must be in sync with camera frame rate.
b. Two or more LED's should not overlap in image space and are separated by at least K_p pixel space. Where K_p is the diameter of largest LED light in image space.
c. The shadow from one LED on another will limit the accuracy of identification of blinking pattern.

Blinking frequency of LEDs should be known in advance to recognize them. Frame is the momentary status of real environment that a camera is capable of capturing and frame rate is number of frames that a camera is capable of capturing in one second. Camera frame rate could be set programmatically to be multiple of LED blinking frequency. Technical design trade-offs are discussed in next section. First of all a set of constant values need to be defined. Few examples of constant values are the number of frames needed to be captured to detect the LED blinking pattern, what would be the ratio of blinking pattern and camera frames per second (henceforth fps). If there are more than one lights the blinking pattern should be defined for each.

First of all the camera frame rate should be set programmatically to ensure the frame rate. Once the frame rate is set, N number of frames are quickly captured. N should be carefully chosen. If accuracy is of top priority N should be at least equal to frames that could be captured in one second by the camera (fps). Practically we observed that if the frame rate is higher the accuracy of detecting light increases in bright environment. Figure 17 shows the ideal and desirable condition of synchronization between camera frame rate and blinking pattern of LEDs. As there is no any hardwired connection between blinking lights and camera the practical scenario can be as shown in Figure 18.

Once the frame rate, frequency of LED lights and number of frames to be processed programmatically are chosen, difference between successive frames is computed which would only display the changes between two frames. This algorithm will not work if there are moving objects in between camera and lights. Once the difference is computed the difference image undergoes a binary thresholding. Binary thresholding is basically done in images to change all the values above threshold value to highest value and all the values below threshold value to 0. Following expression represents the binary thresholding (CV, 2015).

$$\text{dest}(x, y) = \begin{cases} \text{maximumValue} & \text{if } \text{src}(x, y) > \text{threshold} \\ 0 & \text{otherwise} \end{cases}$$

This means if the intensity of pixel at position (x, y) is higher than threshold it would be set to maximumValue else it would be set to 0. Figure 16 shows what exactly is happens to pixel intensity values after thresholding.

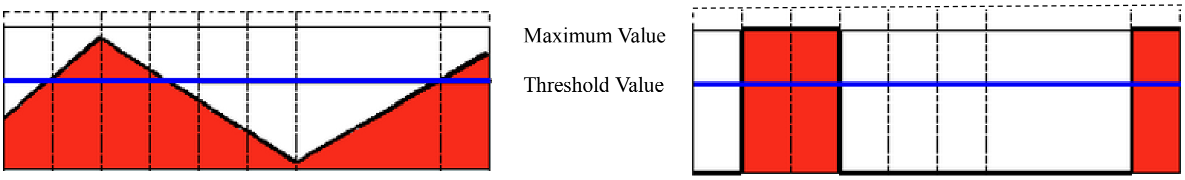


Figure 16: (left) original image showing pixel intensities before thresholding (right) Pixel intensities after thresholding (CV, 2015)

This thresholding in our case would help us get rid of shadows or other noises which may occur because of changing lightning condition which appear in the frame difference image. Now remaining pixels are the candidate positions for our LED lights which are further enhanced by blurring the edges.

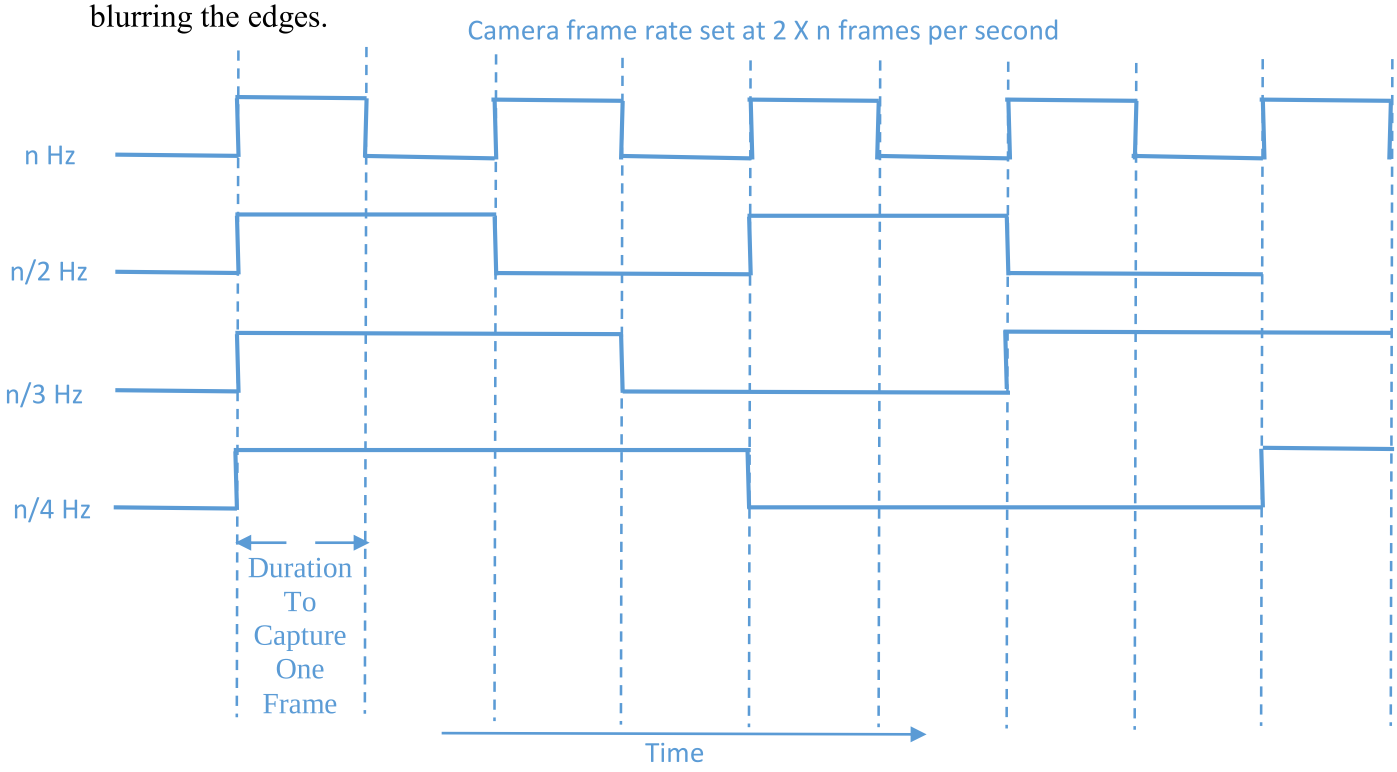


Figure 17: Ideal condition of camera frame rate and LED blinking synchronization

Once the difference image is processed all the remaining contours are possible LED locations. The size of contour solely depends on the projection of LED light on image sensor. Means if the lights are far away the small contours are formed and if the lights are nearer larger contours are formed. Hence we just consider the centers of those contours for our computation. This will save significant computation overhead.

Next problem is how to separate actual lights with noise and how to anticipate the slight movement of camera. As we are solely relying on camera to locate LED lights camera has be relatively steady. But the location of lights and shadows changes slightly even with fixed camera. Now some way needs to be found to determine the frequency of each LED light using contours in processed difference images.

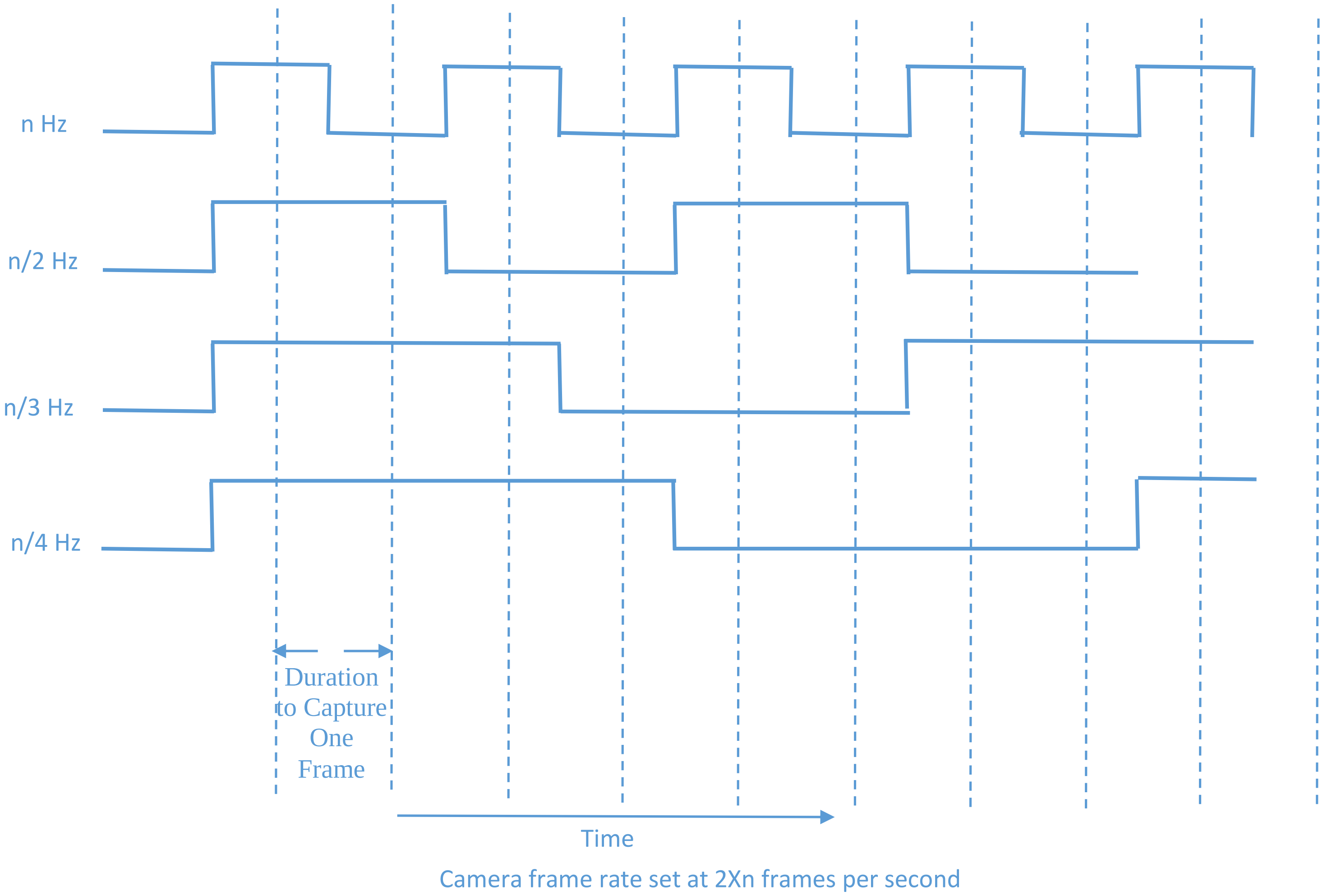


Figure 18: Practical Scenario where the frequency and frame rate are related but the timing of frame capture is not in sync

We needed some clustering algorithm to group the center of identified contours but mere grouping would not work because we had to count the frequency as well. First of all Kmeans algorithm was considered but it was discarded because it would consider both candidate points and noise while computing mean. Then RTree data structure was considered for the purpose of grouping possible candidate points and computing frequency of LEDs.

R-tree data structure was proposed by Antonin Guttman in 1984 for indexing multidimensional information (Guttman, 1984). R-tree is basically a balanced tree which means all the leaves of the tree are at the same height. It groups nearby objects and represent them by minimum bounding rectangle at next higher level of tree. Figure 19 shows a typical R-tree for two dimensional rectangles. As could be seen R-tree supports overlapping rectangles which can create noise in our case. Minimization of both coverage and overlap is critical to the performance of R-trees. Overlap means that, on data query or insertion, more than one branch or the tree needs to be expanded (due to the way data is being spilt in regions which may overlap). A minimized coverage improves pruning performance, allowing to exclude whole pages from search. Hence a variant of R-tree known as R*-tree has been considered for the algorithm. The key part of balancing algorithm is node splitting algorithm (Boost, 2015). This variant of R-tree incorporates reduction of area, margin and overlap of the directory rectangles. This makes R*-tree robust against ugly data distribution. Figure 20 Depicts a typical R*-tree. Now once the data structure to organize contour centers have been identified all the contour centers are inserted into R*-tree. The trick is retrieving them efficiently and in our case avoiding noise candidates.

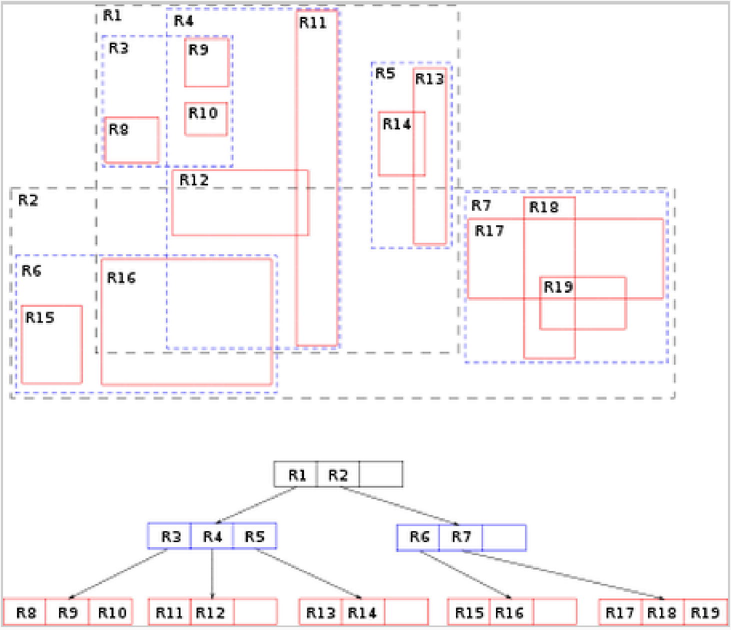


Figure 19: Simple Example of an R-tree for 2D rectangles (Wikipedia, Wikipedia, 2015)

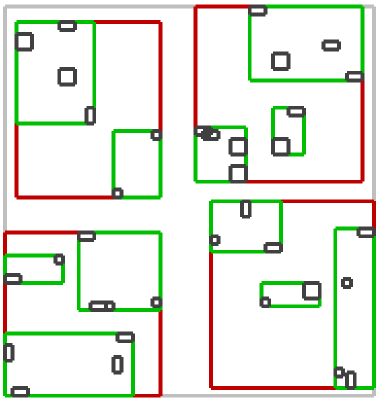


*Figure 20: Typical R*Tree Data Distribution (Boost, 2015)*

A parameter d has been defined to specify the acceptable distance between neighboring points. This parameter is critical for successfully identifying the light source accurately. It indirectly specifies the size of LED light in image plane. Then each contour center is queried for its neighbors within distance d. This approach takes too much of processing because once neighbors of a point are identified, the neighbors of each neighbor are also computed. For example let us say there are four neighboring points P1, P2, P3, and P4. We search for the neighbors of P1 and find other three points. Now again we search for neighbors of P2 and find other three points. Here one possibility is that there may be points P5 and P6 which are not neighbors of P1, but are neighbors of P2 at specified distance d. But practically at this stage where thresholding and blurring has already been done to the images this possibility is less. The tradeoff should be considered and the design choice must be made. When testing this scenario in prototype, computing neighbors of neighbor did not improve much accuracy at which lights are identified but it took a lot of processing time. Hence this algorithm excludes computing neighbors of neighbor once they are identified. R*-tree is queried for the points in neighborhood excluding neighbors of neighbor.

As could be seen in Figure 17 and Figure 18 there is a relationship between camera frame rate and frequency of LEDs. Finally this relationship would be exploited when identifying the LED position. Here is the calculation logic for the LED which has frequency n. Camera frame rate is set at nX2 means ideally the LED is on in one frame and off in another frame. N frames have been captured and difference between two frames have been computed and processed. The number of neighboring points will be equal to N/2. If frame rate is three times the frequency the number of neighboring points will N/3 and so on. Some frames might be missed during capture and in practical scenario some frames may not show any difference at all. Hence a tolerance must be defined which would be based on the noise present in the environment. The neighboring points and their count has already been identified from R*-tree. Now the points that have a neighbor count in the range of actual value ± tolerance are the final candidate points. These candidate points could be further checked for the luminosity value in unprocessed frame just before finalization to increase the accuracy. Practically we observed that this step is not needed if both binary threshold in difference image and distance predicate in R*-tree are set to the optimal value for the operating environment.

This algorithm has only minor performance degradation if the number of lights is increased. No restriction on number of lights have been placed. If the blinking pattern of one light does not affect other light with shadows and other interference both the lights could be detected independently. If detecting blinking pattern is not adequate for some specific application the data could be read using the same algorithm with following changes

a. N (number of frames captured initially to locate lights) should be greater than or equal to frames per second (fps)
b. The order information must be preserved for the contour centers which could be easily done by identifying the frame number while checking for contour centers. And R*-tree works without any problem for multidimensional data.
c. Once the LED points are identified by checking just the blinking frequency all the neighboring points queried from R*-tree for those LED points are ordered according to frame number. Now let us say a point (x, y) has neighbors (x, y) at frame 1, (x+p, y+q) at frame 3 and so forth. The data could be extracted as 1 (frame 1) 0 (frame 2 is absent) 1 (frame 3 is present) and so on. Where distance between (x, y) and (x+p, y+q) is less than or equal to d (distance predicate for R*-tree query).
d. A Cyclic Redundancy Check (CRC) on data must be performed to ensure data has been read correctly.

At this stage the LED positions are known in the image coordinate system. Although this algorithm doesn’t restrict on the number of LEDs, the shape that LEDs form must either be known in advance or should be interpretable from LED data received. Hence forth we would presume that four LEDs have been installed in rectangular shape for the sake of explanation simplicity. At this point we have located the positions of LED lights and approximated their blinking frequency. Before moving on to next step in algorithm, some discussion on how camera projects real world three dimensional objects into a set of pixels that constitute an image is needed. Following image displays coordinate systems in camera world. Camera Orientation has its own coordinate system and image plane has its own coordinate system and real world object has its own coordinate system. Camera basically projects real world object on Image plane from the perspective of camera coordinate system.

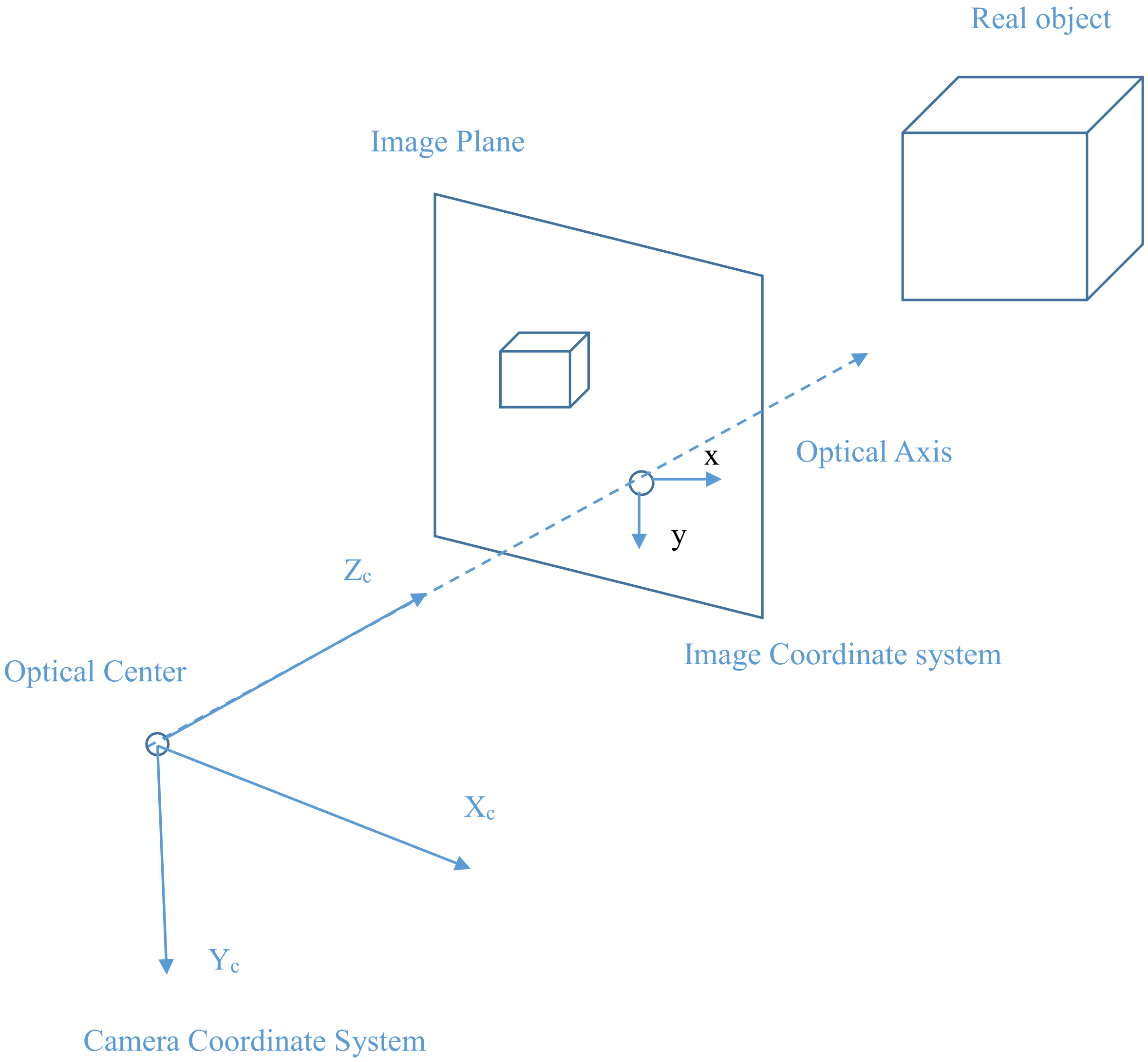


Figure 21: Camera and Image Coordinate System

As explained above the positions of LEDs have been computed in image plane coordinate system. As mentioned earlier for simplicity we are assuming there are four LEDs installed. LEDs are located in real environment, so we need to overlay virtual information over LEDs to realize augmented reality system. Again for the sake of simplicity we would like to display a rectangular image with its vertices at exactly LED positions in image plane. A three dimensional virtual object display would require different approach. The scope of this research is to display two-dimensional information. Following figure displays the current problem:

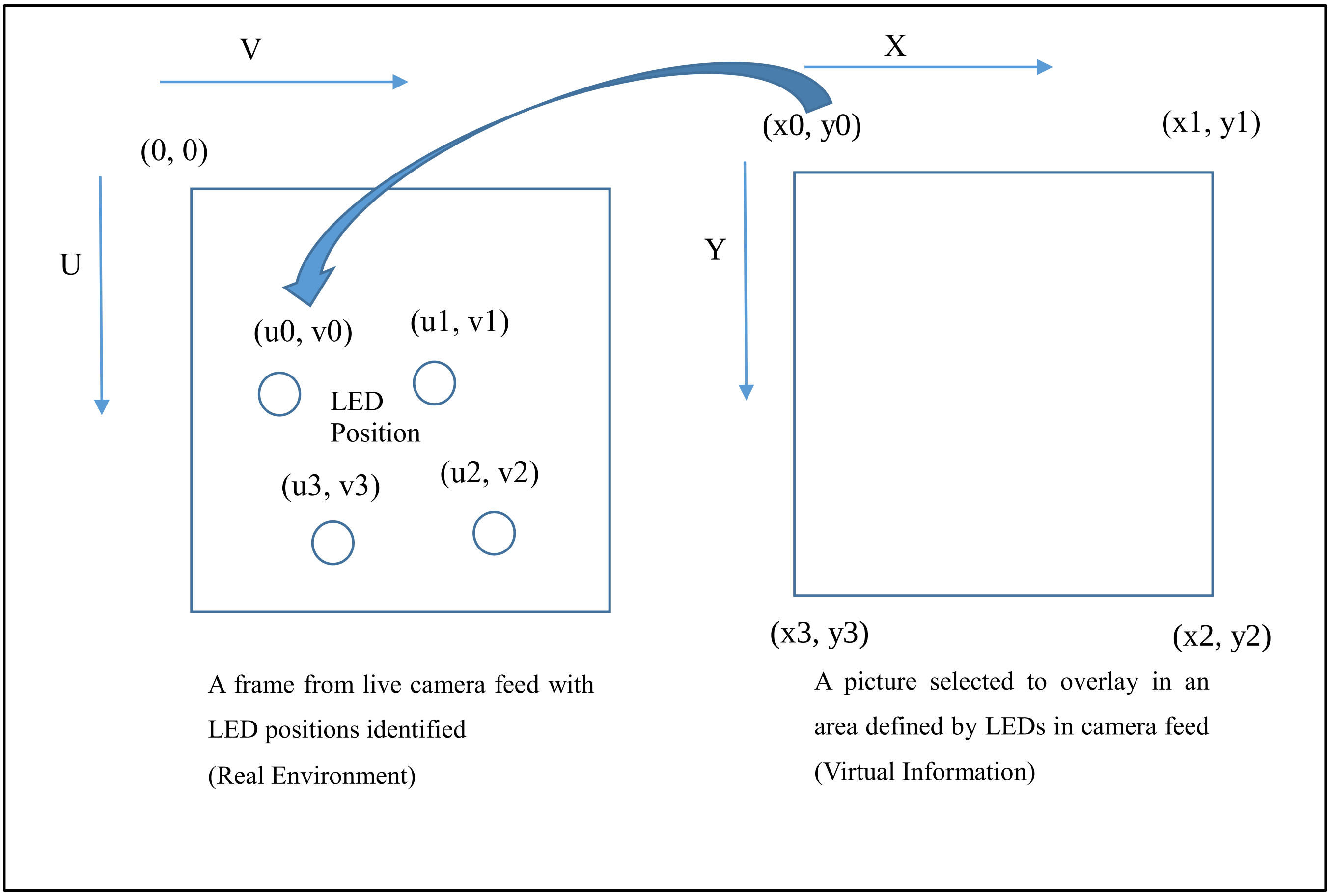


Figure 22: Coordinate Systems for LED Lights in camera feed plane and Image Overlay

Now we have two sets of four Cartesian coordinates from real world and from virtual world. A perspective transformation is computed to fit virtual object into real environment. The goal is to transform the four points (xi, yi) to four points (ui, vi) for i = 0, 1, 2, 3.

The projective transformation (also known as a collinearity or homography) can map any four points in the plane to any other four points. Rotations, translations, scaling, and shears are all special cases (Prince, 2012). The transformation matrix is as follows:

$$\lambda \begin{bmatrix} ui \\ vi \\ 1 \end{bmatrix} = \begin{bmatrix} a0 & a1 & a2 \\ b0 & b1 & b2 \\ c0 & c1 & c2 \end{bmatrix} \mathrm{X} \begin{bmatrix} xi \\ yi \\ 1 \end{bmatrix}$$

In Cartesian coordinates the homography is written as:

$$u_i = \frac{a_0 x_i + a_1 y_i + a_2}{c_0 x_i + c_1 y_i + c2}$$

$$v_i = \frac{b_0 x_i + b_1 y_i + b_2}{c_0 x_i + c_1 y_i + c2}$$

The eight unknown coefficients could be calculated by following matrix (considering c2 = 1):

$$\begin{bmatrix} x0 & y0 & 1 & 0 & 0 & 0 & -x0u0 & -y0u0 \\ x1 & y1 & 1 & 0 & 0 & 0 & -x1u1 & -y1u1 \\ x2 & y2 & 1 & 0 & 0 & 0 & -x2u2 & -y2u2 \\ x3 & y3 & 1 & 0 & 0 & 0 & -x3u3 & -y3u3 \\ 0 & 0 & 0 & x0 & y0 & 1 & -x0v0 & -y0v0 \\ 0 & 0 & 0 & x1 & y1 & 1 & -x1v1 & -y1v1 \\ 0 & 0 & 0 & x2 & y2 & 1 & -x2v2 & -y2v2 \\ 0 & 0 & 0 & x3 & y3 & 1 & -x3v3 & -y3v3 \end{bmatrix} \begin{bmatrix} a0 \\ a1 \\ a2 \\ b0 \\ b1 \\ b2 \\ c0 \\ c1 \end{bmatrix} = \begin{bmatrix} u0 \\ u1 \\ u2 \\ u3 \\ v0 \\ v1 \\ v2 \\ v3 \end{bmatrix}$$

Once the projective transformation matrix is computed the virtual information (image) is superimposed over real information (live camera feed). Finally this algorithm creates an AR system which reads LED markers in real environment and overlays virtual information over LED markers.

To summarize first of all LED lights are installed onto real environment (building or other objects). Custom camera frame rate is set such that frame rate is more than two times and in multiples of frequency of LED light. Then a number of frames are captured quickly, the frames are processed to identify LED position. A use case where four LEDs are installed has been considered to explain further concept. Once four LEDs are identified, homography between coordinate systems of camera frame instance and virtual information is computed. Finally the real environment on camera feed is augmented with virtual information exactly at the position where LEDs are located. Following flow chart depicts this algorithm pictorially.

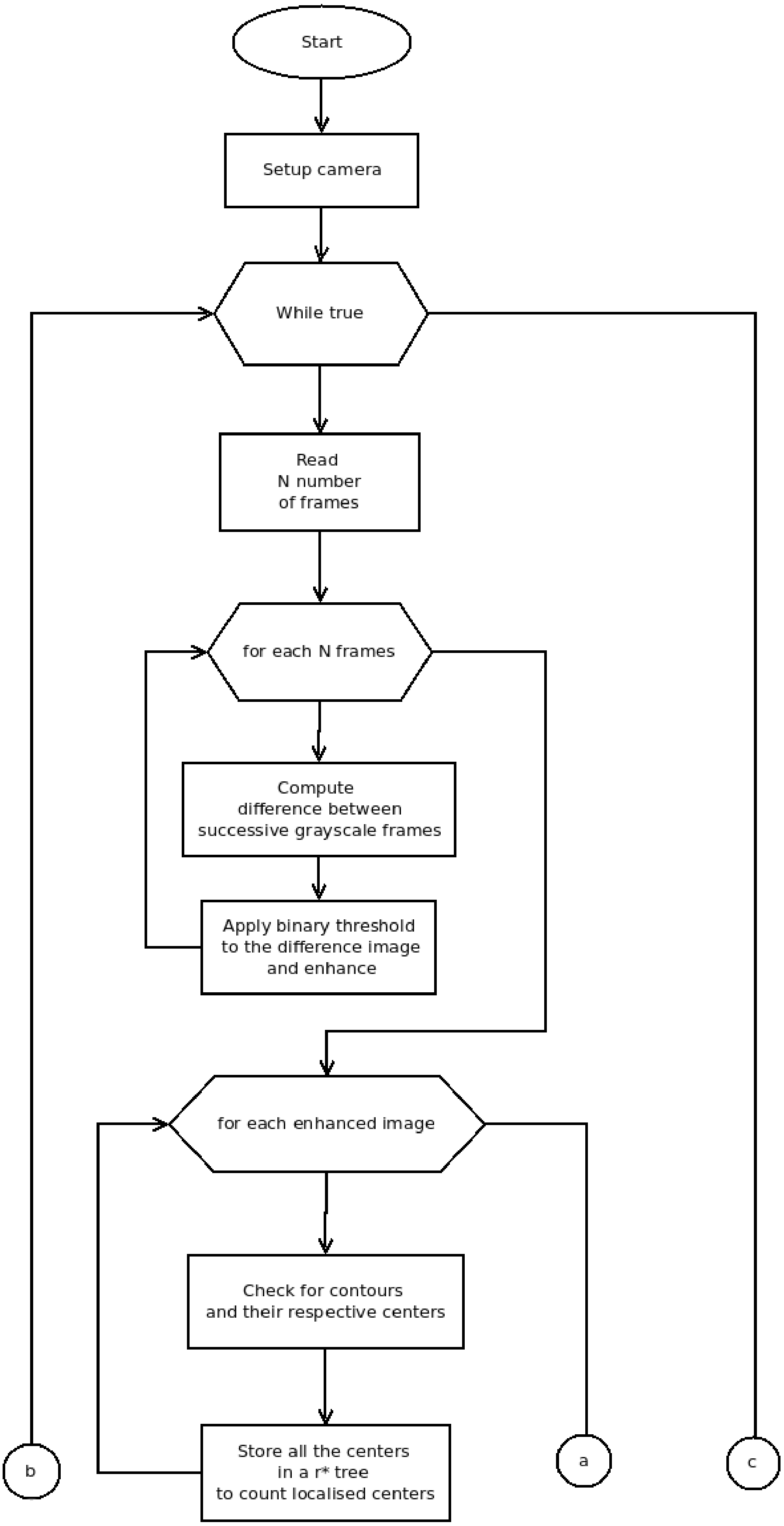


Figure 23: Computational Flow Chart

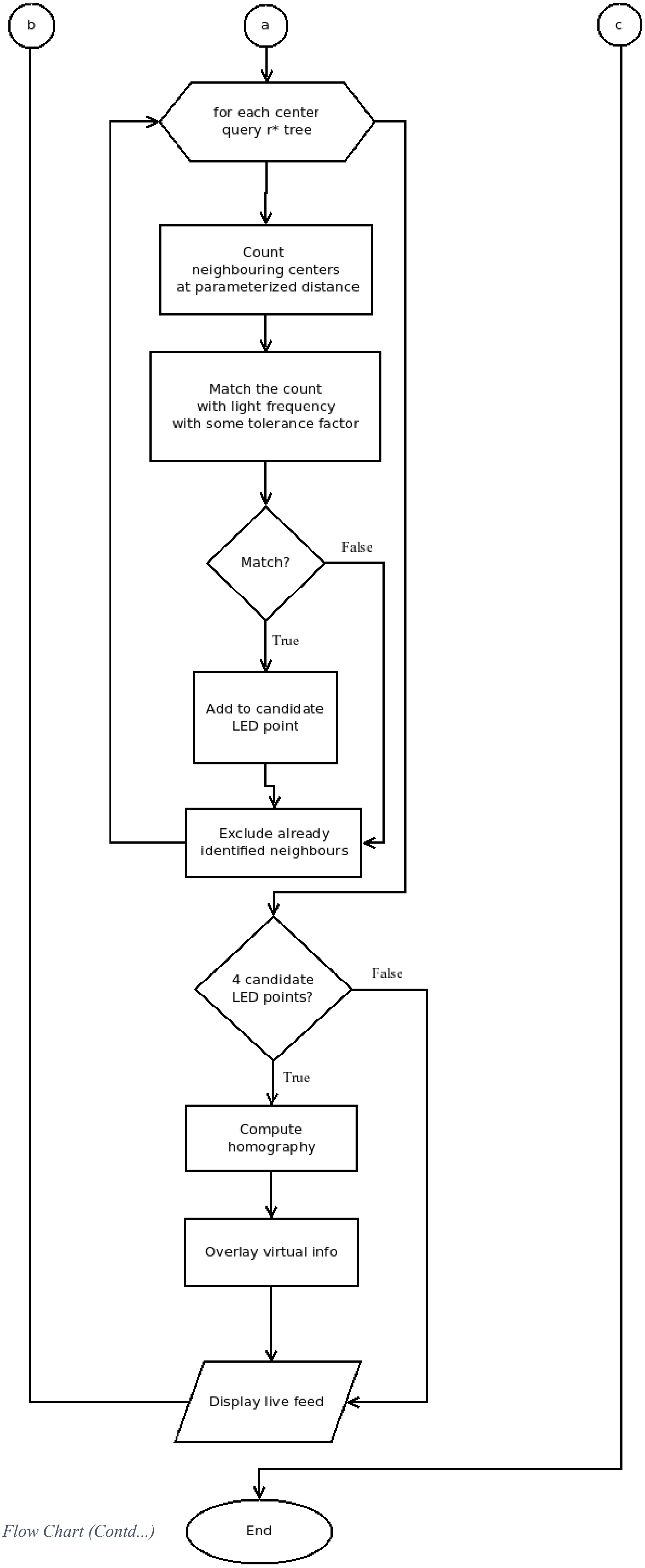


Figure 24: Computational Flow Chart (Contd...)

5 Prototype Design

5.1 Technology Tradeoffs

In this section, the reason behind selection of a physical component or a particular technology while designing the prototype has been discussed.

5.1.1 LED Lights

In visible light communication light is the medium through which information is transmitted. LED lights are the first choice of visible light communication because of their switching speed, longevity, low power requirements and minimum heat generation during lightning. In order to realize AR system using VLC the lights need to be ubiquitous. The white LED is considered as a strong candidate for future lightning technology (Komine & Nakagawa, 2003). As shown in Figure 25, by 2020 the market share of LED lights will become 64% of global lighting product market (Company, 2012).

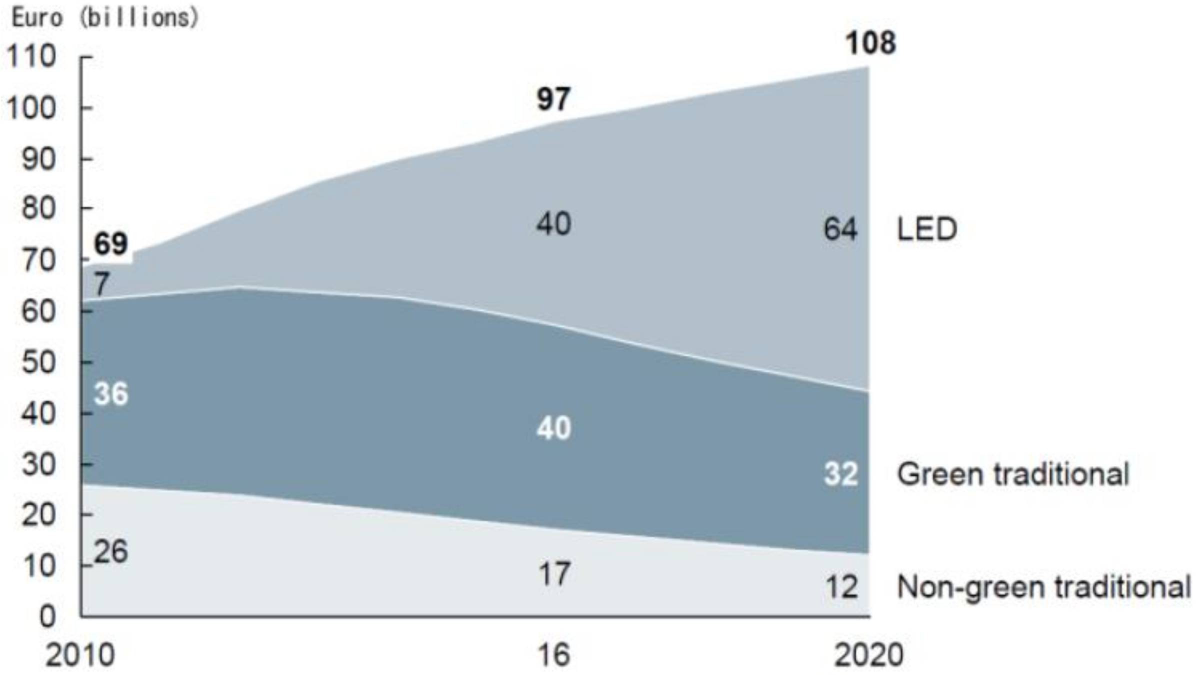


Figure 25: Global Lighting Product Market Trend

Basic data modulation could be introduced in the form of blinking LED using simple circuitry. The blinking pattern is not visible to human eyes as the frequency is higher than the eyes can perceive. Hence white LED light with a circuitry to enable custom change in frequency is used as light beacon in our prototype.

5.1.2 Camera

As discussed in the algorithm (Section 4) the camera should be able to quickly capture N number of images and the frame rate should be in multiple of LED frequency (at least 2 times). The blinking is visible to human eye if the LED frequency is lower than 45Hz. Hence
We needed a high speed camera which could be configured programmatically for both image size and frame rate. In addition we needed a high data rate through which the frames could be displayed on screen. PointGrey Grasshopper 3 could render 90 (practically 82) frames of resolution 2048X2048 per second and the configuration could be changed programmatically to increase the speed depending on image width, image height and packet size. It provides USB 3.0 interface for fast transfer between computer and camera. Hence it was part of our design to develop our prototype.

5.1.3 GPU

The AR system requires extensive computing which demands a high performance GPU to render video and information overlay in real time. The GPU needs to be compatible with camera Application Programmer's Interface (API) as well. Hence for this purpose NVIDIA Tegra K1 was used to aid extensive computing. This GPU is equipped with 192 cores and 2048 threads are allowed per multiprocessor core. As it is a light weight, low power consumption GPU it could be used for PDAs (Personal Digital Assistants) in future. NVIDIA Tegra K1 is running on Ubuntu 14.04 for arm platform.

5.1.4 C++

C++ is fast, efficient and platform independent language which can be used as both low level and high level programming language. As C++ is a general purpose language which is light weight and provides a lot of flexibility for memory management hence we decided to use C++ for prototype development. Some C++11 features like lambda and record query have been used

5.1.5 OpenCV

Open Source Computer vision library is an open source computer vision and machine learning library developed by Intel Corporation (opencv.org, 2015). The library has more than 2500 optimized algorithms in the field of computer vision and machine learning. It has been developed in C++. OpenCV is one of the most widely used libraries in computer vision both commercially and in the field of research. OpenCV 2.4.10 has been used to develop this prototype as it is the latest version available for ARM platform running on GPU.

5.1.6 Boost

This algorithm is using R*-tree to organize and retrieve spatial data efficiently. We had two choices either to develop the data structure from the scratch or use third party libraries. To avoid reinventing wheels, we decided using a third party library. We needed a library which works in both Linux and windows platform, which is open source and freely available and developed in and tightly integrated with c++. Boost libraries are open source and could be used freely for both academic and commercial purpose hence it was used to implement R*-tree in the algorithm. Boost version 1.57.0 has been used as it was the latest version available at the time of prototype development.

Once all the physical components of the system selected and platform has been decided a prototype was developed and evaluated. A typical system deployment area is shown in figure below:

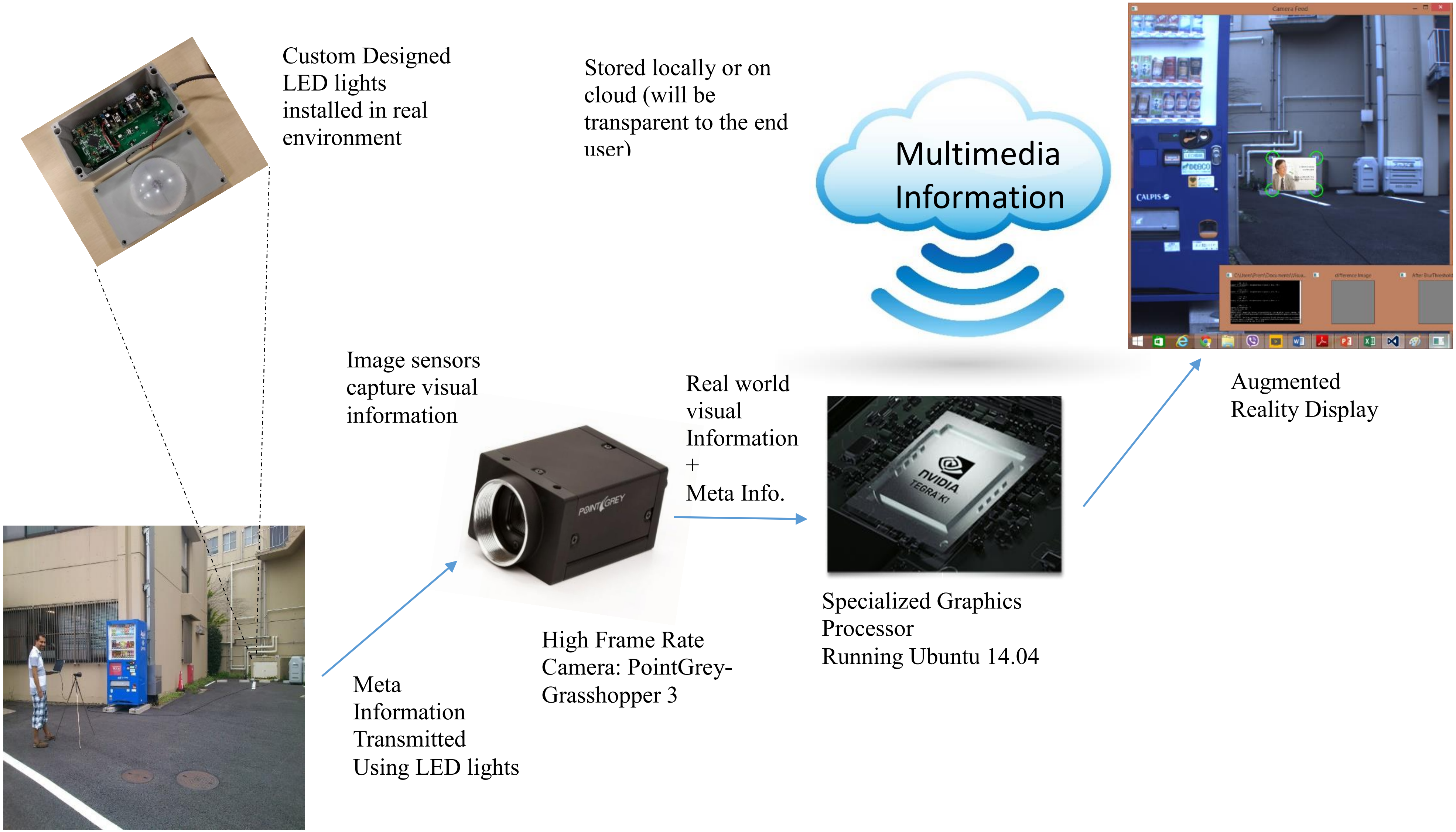
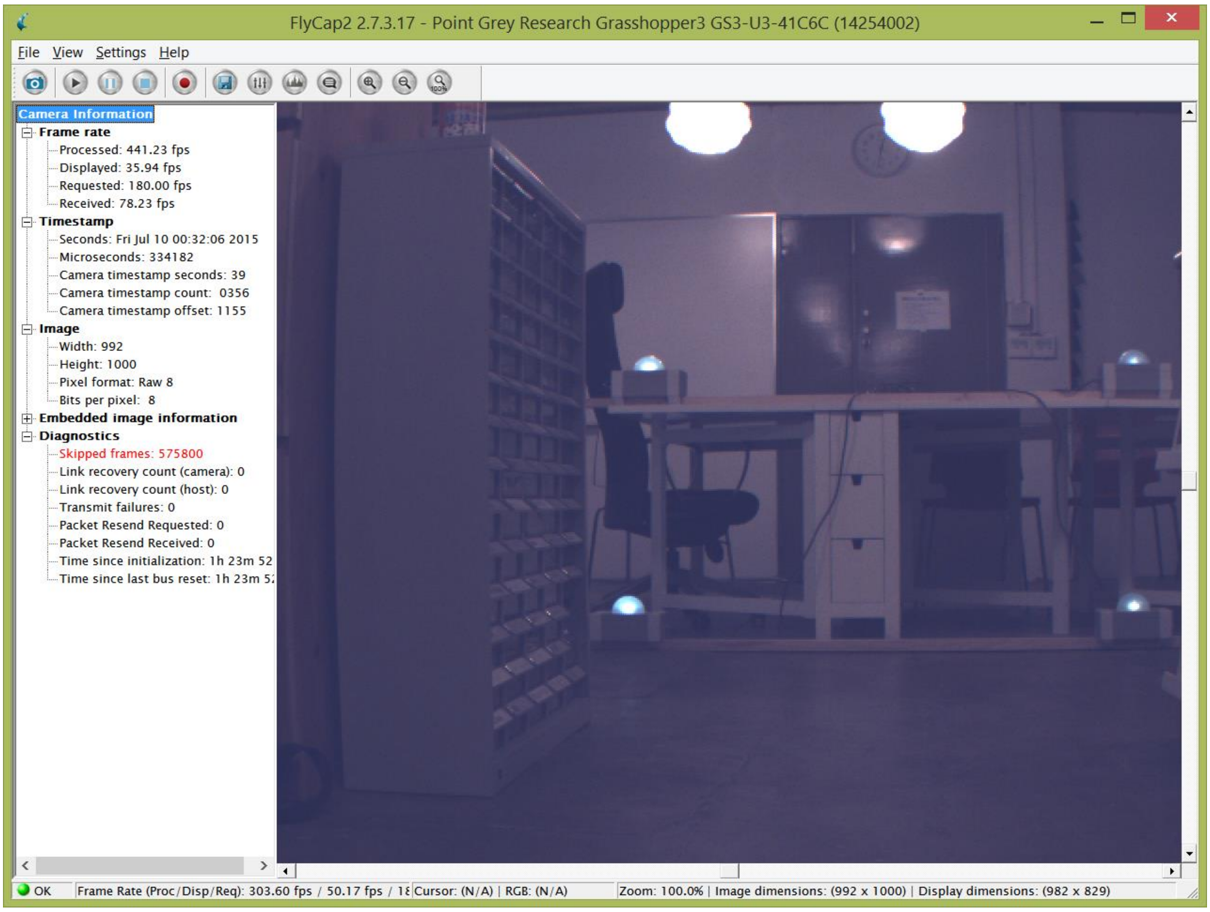


Figure 26: Physical System Design

6 Results

This section contains the actual screenshots from the experiment and their explanations. Till now AR system using visible light communication has been discussed as an algorithm, in terms of technical architecture and physical components. This section discusses the functioning of actual prototype AR system. Let us now analyze what happens at the image level. These screen shots were taken indoor with external florescent lights on. Which shows the robustness of the system in external bright light scenario. The chronological order of results shown below:

a. The camera is set on the target area. Four LED lights along with other lights and shadows (in this case noise) could be seen in camera frame.



```
Camera Information
⊟ Frame rate
    Processed: 441.23 fps
    Displayed: 35.94 fps
    Requested: 180.00 fps
    Received: 78.23 fps
⊟ Timestamp
    Seconds: Fri Jul 10 00:32:06 2015
    Microseconds: 334182
    Camera timestamp seconds: 39
    Camera timestamp count:  0356
    Camera timestamp offset: 1155
⊟ Image
    Width: 992
    Height: 1000
    Pixel format: Raw 8
    Bits per pixel:  8
⊞ Embedded image information
⊟ Diagnostics
    Skipped frames: 575800
    Link recovery count (camera): 0
    Link recovery count (host): 0
    Transmit failures: 0
    Packet Resend Requested: 0
    Packet Resend Received: 0
    Time since initialization: 1h 23m 52
    Time since last bus reset: 1h 23m 5

OK  Frame Rate (Proc/Disp/Req): 303.60 fps / 50.17 fps / 1  Cursor: (N/A) | RGB: (N/A)  Zoom: 100.0% | Image dimensions: (992 x 1000) | Display dimensions: (982 x 829)
```

Figure 27: Camera View

b. Camera quickly captures N frames and converts them into grayscale images. Then difference in pixel intensity is computed between successive frames. Screen shots below show two successive frames.

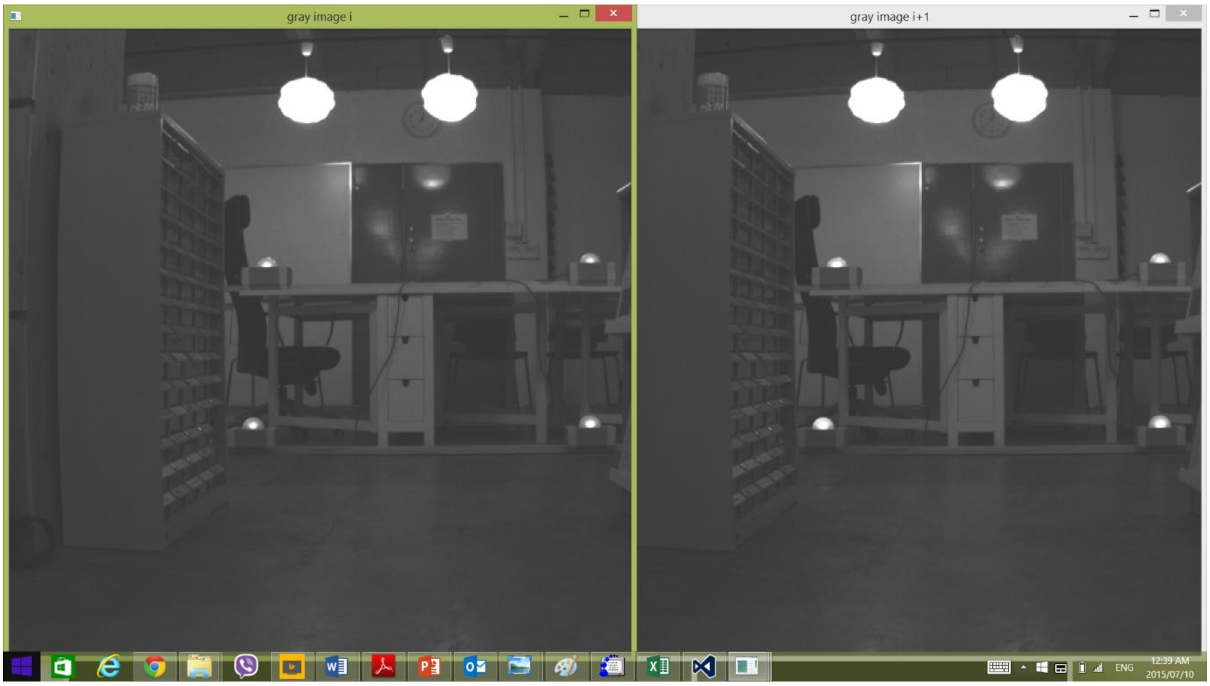


Figure 28: Two Successive Frames

c. Now difference between two successive frames is computed and a threshold is applied on computed difference image. As explained in Figure 16 the binary threshold changes values below threshold to 0 and above threshold to maximum value. As can be seen in the left image, light spots appear to be faint whereas after application of threshold spots are much visible and all other noise signals are gone after thresholding. Practically we observed that brighter the environment lower should be the threshold. A threshold of value 10 (this is pixel intensity value) works outside and a threshold of value greater than 50 works inside.

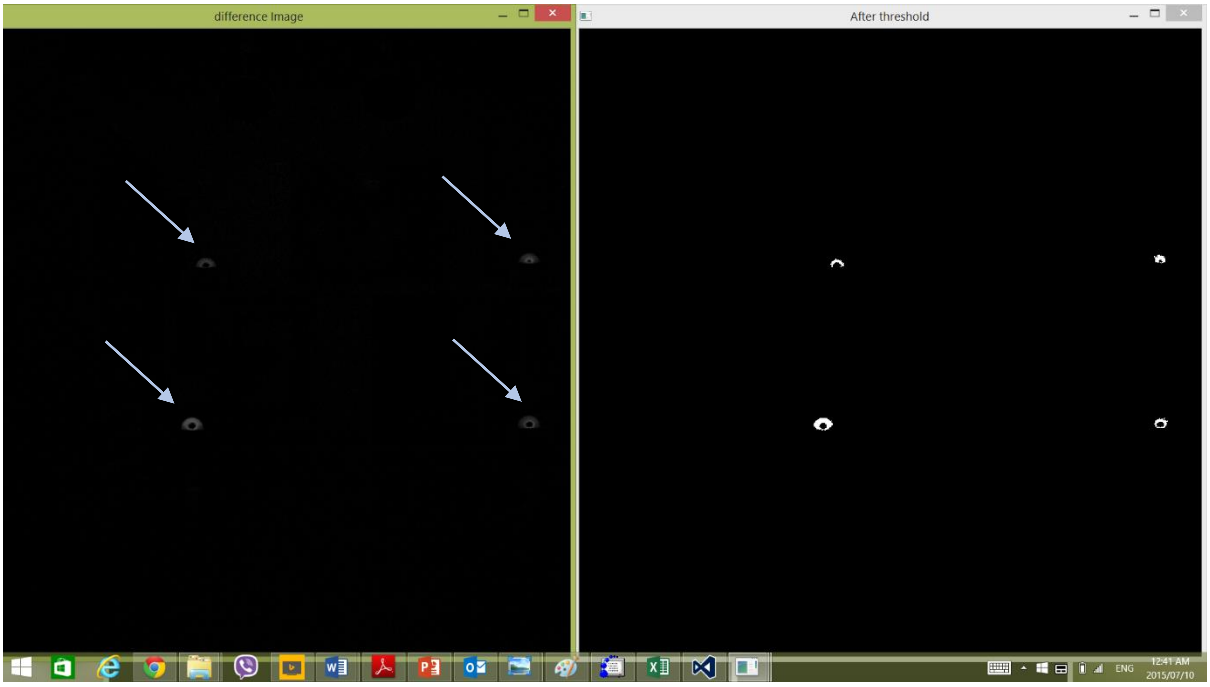


Figure 29: Result of difference between two frames (left) and Thresholding the resulting image (right)

(Arrows are for pointing purpose only)

d. After threshold is applied on the difference image the light spots are more prominent but they are not consistent in shape and are broken sometimes. Hence the identified spots are smoothen using a kernel. This will remove broken segments as shown in right picture below.

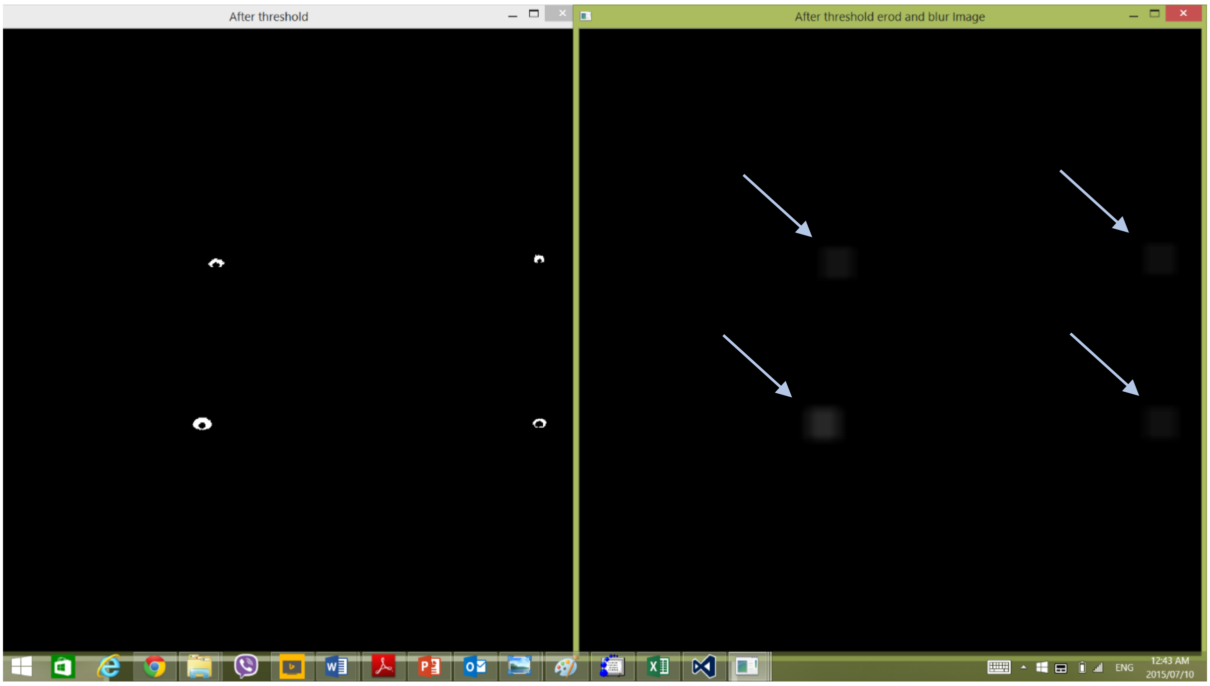


Figure 30: After applying Threshold (left), After Smoothing the Same Image (right) (arrows are for pointing purpose only)

e. Now the centers of the light spots are identified using a minimum bounding rectangle technique for the area.

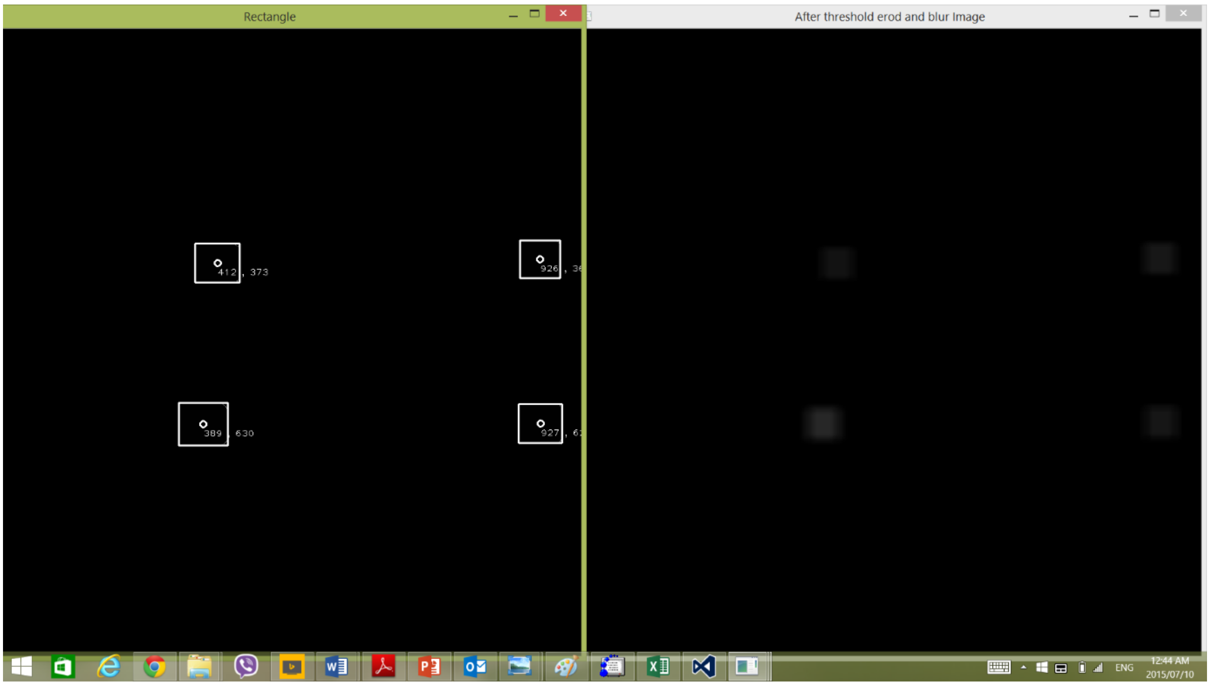


Figure 31: Minimum Bounding Rectangle and Its Center (left), the smoothen source image (right)

f. Similar computation is done for N number of frames and centers computed from each pair are stored in R*-tree data structure which is basically used to store and query spatial data. The center points may or may not coincide in each frame pair because of various reasons like changing light conditions, changing shadows and others. The main purpose of storing these points in R*-tree is to easily query neighboring points within a specified distance (here the size of Light in original image). The neighboring points are grouped and counted. The count of neighboring points should relate to the frequency of light with some tolerance. Here is an example how data is stored in R*tree. Let us take four coordinates shown in Figure 31, (413, 373), (926, 368), (389, 630) and (927, 629). In this particular test case we cannot see other noise coordinates but practically some of them appear, as lighting and shadow conditions change or some movement is seen within camera frame. If 30 frames are captured to locate LED lights in camera frame and camera frame rate has been set two times the LED blinking frequency. Ideally each point will appear 15 times and a simple array could be used to organize the data. But this will not happen. Each LED will be detected at slightly different point in each frame pair. Hence they are stored in R*-tree which would appear as follows.

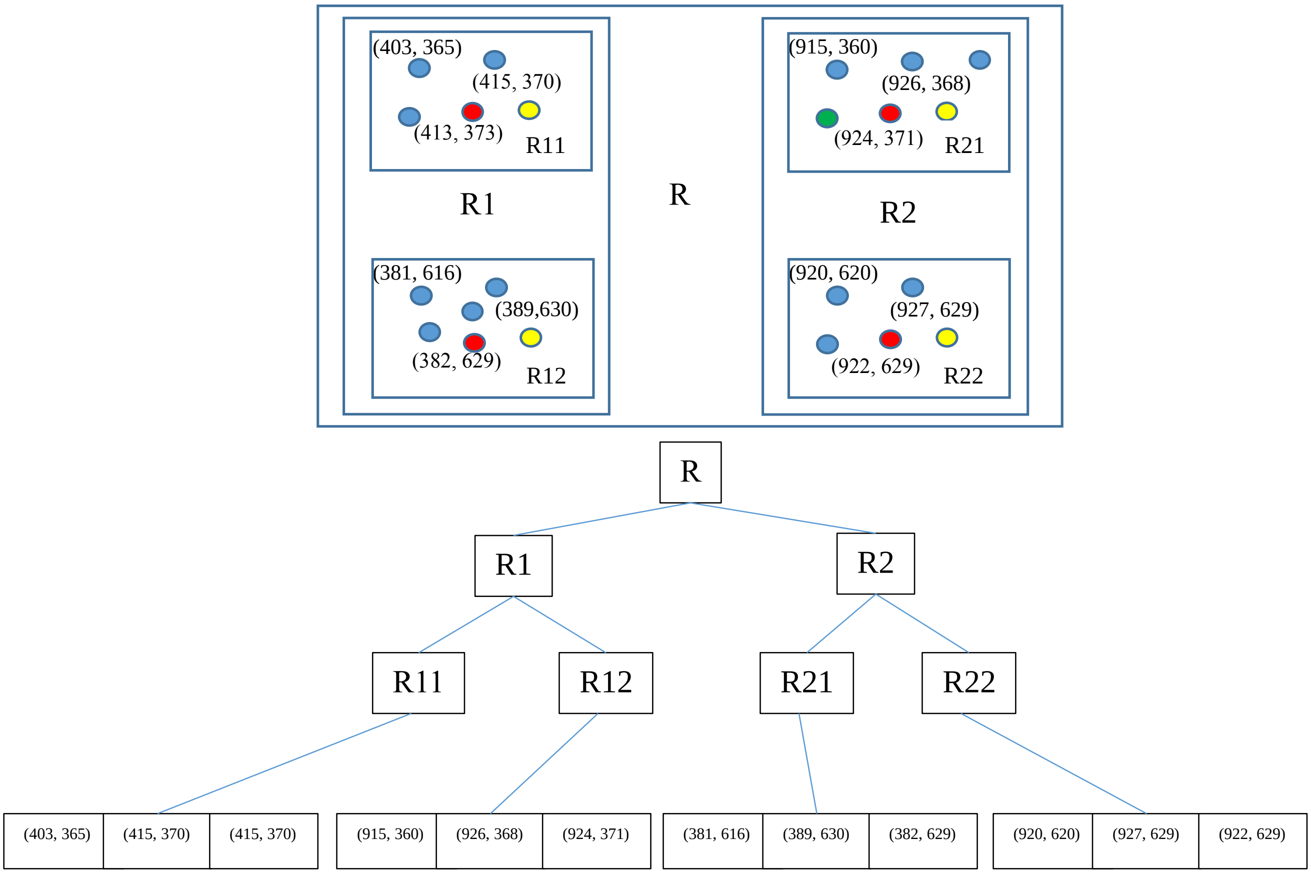


*Figure 32: R*Tree Data Structure of Cartesian Points Representing LED Positions in Each Frame Pair*

As can be seen in above tree diagram points are stored in a group defined by minimum bounding rectangles. This is exactly how we would like to query our data once all the frames are processed. Real benefit of R*tree could be realized (significantly fast query) when querying points within distance d (in this case 10).

g. Finally the final candidate points are selected and sorted if their count is four. Then the virtual information is overplayed over the real environment.

Figure 33: Finally virtual information is superimposed over real environment at the points computed earlier

6.1 Test Cases

6.1.1 Computing Machines

The prototype has been tested both on windows and ARM machines. When running identical code on windows laptop (Intel i5 processor and 8 GB primary memory) and GPU (NVIDIA Terga K1), GPU gives an efficiency of two times than windows. The code needs to be optimized to fully utilize OpenCV GPU computing features.

6.1.2 Location

The prototype of the system has been tested indoors as well as outdoors. Operating distance of 20 meters has been achieved indoors with 95% accuracy. A 93% accuracy was achieved at 12 meters of distance outdoors. Here the percentile corresponds to the number of times all lights could be located and virtual information could be overlaid accurately per hundred frames.

6.2 Lightning Condition

The system has been tested in the dark, under florescent lights and daylight. The accuracy in the dark increased up to 98% at a distance of twenty meters. An accuracy of 95% has been achieved under florescent lights at twenty meters.

All these testing were done in case where the LED lights were identified by their blinking frequency and a tolerance has been set to compensate the missed frames and synchronization error. Initially thirty frames were processed and the relative number of frames where light would appear have been parameterized. Once the location of lights were identified in the image plane, homography would be computed to overlay a virtual information. In all test cases an image has been considered as virtual information. It could be replaced with any other multimedia information without much overhead.

7 System Evaluation

The goal of evaluation process is to establish confidence that the product or service is fit for its purpose. This does not mean that the product or service is completely free of defects or bugs. Rather the product or service must be good enough for its intended use and type of use will determine the degree of confidence that is needed.

Basically system evaluation involves answering following two fundamental questions:

a. Have we done the right thing?
b. Have we done the thing right?

The product or service must be built right which basically means it meets the requirements and objectives that were defined/identified at the beginning of the system design process.

The right product or service must be built which basically means that it meets the user needs.

The distinction between above two criterion is obvious, even though the product/or service satisfies all the requirements and objectives it may not be useful for the end user or vice versa, it may be useful for end users but may not satisfy all the objectives and requirements. In order for the product or service to be successful both criterion should be met.

System Engineering Handbook version 2a published by International Council on System Engineering (INCOSE) defines verification and validation as follows.

“The objective of the verification process is to ensure conformance of the implemented product and processes to all source and derived requirements and that the planned development process has been followed.

The objective of the validation process is to ensure the implemented product functions as needed in its intended environment, including its operational behavior, maintenance, training, and user interface. “

Verification is about evaluating the process i.e. are we following the right process to achieve our objectives? And validation is about evaluating the product i.e. whether the product is useful to the user?

7.1 System Verification

System verification basically confers that the system meets or exceeds the user requirements and system objectives. Both functional and non-functional requirements must be verified along with user requirements in system verification phase. In order to verify our system each item in Table 5: Objectives, Requirements and Assumptions Mapping is matched with product feature of the prototype. System verification is not a one-time process rather it’s an iterative process;

means each phase of system development lifecycle must be verified. Some features or requirement which could not be met by the system in its original form, the modifications have been proposed.

Table 6: Requirement V/S Prototype Feature Mapping for System Verification

SN	Requirement	Prototype Feature
1	LED must be blinking at predefined frequency	LED light is equipped with the circuitry so that its frequency could be set programmatically
2	Data Structure of the information transmitted should be known in advance	The LED blinking pattern is known and has been parameterized in the prototype
3	Camera frame rate should be in multiples of LED frequency	Camera frame rate could be set programmatically. The parameters affecting camera frame rate for point gray camera are image width, image height and packet size.
4	Location of all four LEDs must be known	While testing prototype in test environment, the LEDs could be located accurately for 95% of the time (indoor)
5	Storage location of virtual information must be known.	In this prototype the virtual information has been stored locally and the storage location has been hardcoded
6	The virtual information would adapt to the depth and angular displacement of lights in real environment	The image used in prototype, changes its size and shape according to location of LED lights in image plane.
7	Adjustable size of information display area	There is no restriction set for the distance between LED lights or the size of rectangle that is formed by four LED lights
8	Information panel should not overlap	The panel depends on how LED lights are installed in target area. Hence at the time of Installation if LED lights do not form an overlapping area then it is not an issue
9	Information should be visible from less than 10 meters	The prototype has been tested both indoors and outdoors to work successfully at distances 20 meters and 12 meters
10	Should work in dark as well as daytime	The prototype has been tested in the dark as well as in the daytime. It works more accurately and at more distance in the dark than during daytime
11	Multimedia virtual information should be available	In this research work virtual information has been chosen to be an image and could easily be replaced with other multimedia information as well.
12	Information should be displayed in color	The prototype is capable of displaying information in any color mode
13	Option to control opacity of virtual information	This requirement couldn’t be realized in prototype

SN	Requirement	Prototype Feature
14	Should be easy to turn on and off the information	A flag has been provided in prototype which would simply turn on and off the virtual information.
15	Option to filter the specific information only	This depends on the data structure of the information transmitted from LEDs. One implementation mode might be each type of information is associated with a particular frequency range of LEDs. Users are given control to choose the frequency of LEDs and in turn the type of information displayed.
16	LEDs should not be flashing at some frequency range which is more likely to trigger photosensitive epilepsy	No restriction has been added on frequency in the solution. Hence at the time of installation of LEDs at a particular site, the frequency in the safe range must be chosen.
17	Virtual information should not override privacy information	The user can basically chose what to display. The algorithm doesn't do any privacy checking

The system has thus been verified against the user requirements. Any test may confirm the presence of some kind of error, but it's almost impossible to confirm absence of error.

7.2 System Validation

Any system is designed for specific purpose or goal. System is not an autonomous entity, it serves end users. Even if all the requirements are full filled, system is compliant to all the regulations if it is not useful to the end user the very purpose of its design is defeated. This is where system validation comes. System validation is basically assessing the usefulness of the system to the end users. End users evaluate the system for its usability in real life.

7.2.1 Concept Validation

The first step in validation is validating the concept. We came up with an idea of using visible light communication to realize an AR system. It seemed that this system is useful to tag exteriors of the buildings and display some relevant information at specified position. But before proceeding ahead for system design, the concept itself needs validation. Hence a survey was conducted among fifteen people living in eight different countries. The conceptual AR system was explained both in words and pictorially and then the people were asked to rate the usefulness of the system on a scale of five. One being the least useful and 5 being the most useful. The results have been tabulated below:

Table 7: Survey Results for Concept Validation

SN	Usefulness of AR Display (on a scale of 5, 1 being the least useful)	AR Solutions Ever Used
1	4	Oculus rift
2	5	No
3	4	No
4	5	Google Glasses
5	3	No
6	5	No
7	1	No
8	4	Oculus Gear of Samsung
9	4	No
10	5	No
11	2	No
12	4	No
13	4	Yes, App to map sky and show names of the stars

As evident from the above table about 70% of the people never knowingly used some AR system. The average utility ranking of the system on a scale of 1 to 5 is 3.85. It is interesting to observe that the average usefulness rating of system by people who never used AR system was lower (3.67) than people who have used AR systems (4.25). This survey validates the concept by a score of 77% (3.85 / 5) on usefulness scale. Which means people would find this system useful in their daily life.

7.2.2 Product Validation

Once the concept has been validated, the product was developed based on user requirements and functional objectives. Before asking the users how do they find the system? The system itself had to be validated against the definition of Augmented Reality. Means whether the prototype developed could be termed as Augmented Reality system or not? An interview with an expert in the field computer vision was conducted for confirmation. Assistant Professor Hideaki Uchiyama from Kyushu University is an expert in computer vision and has been doing research in the field since last ten years. The prototype was demonstrated to him and he was asked if it could be termed as AR system. In reply he stated that any system that uses past information to locate the markers and overlays virtual information on real environment in current time or future could be termed as AR system. Since the prototype functionality was consistent with this definition it could be termed as AR system.

Finally users were asked to use the system and evaluate its usefulness and what do they like about the system and what they don't like about the system. User responses have been tabulated below:

Table 8: Product Demo and User Feedback

SN	Usefulness rating	Positives	Negatives
1	5	It looks cool	
2	5	Light could be used for both lighting purpose and marking purpose	This is noise prone as the movement between camera and scene impairs the quality of output.
3	5	Useful, if it works for cell phone	Continuous power supply is needed to light up LEDs, accuracy depends on the lightning condition
4	4	As compared to sign boards and other digital displays cost is the biggest advantage	The information is visible to only those people who see through PDAs and cameras, which is a big disadvantage
5	3	Looks interesting but impractical	Users may miss information, they need to know where the information is being displayed using lights
6	4	I would like to use this system	Companies may still be interested in digital display
7	5	If it becomes popular, it can bring in significant savings, can be used to prevent accidents, will be quite useful if installed in monuments	If there are more lights with complex structures the system should be transparent to the user

Seven users (SDM students) were asked to use the prototype and provide feedback. In addition the users were asked to rank the system at usefulness scale (1 being the least useful and 5 being the most useful) as earlier. The average usefulness rating of the system is 4.42. Some users found the system to be interesting but impractical. Some pointed out that it is useful only if it works for cell phones, it doesn't look useful in its current form. Some users pointed out that the lights could be used for dual purpose i.e. for lightning and as markers. Almost all users have used some kind of AR system in the past and had basic idea about how AR works. The usefulness scale of the prototype confirms that users find the system useful in everyday life.

8 Conclusion

An AR system has been conceptualized using visible light communication. The concept was validated from a survey. Past research works and existing products have been studied for the inherent problems and an objective has been set. User needs of conceptual system were collected and a system boundary has been set. Henceforth requirements and assumptions have been formalized within system boundary. The subsystems within that system were identified and their interface was visualized. Then the functional flow of the system was constructed. An algorithm was developed to realize the conceptual system. In other words an algorithm was developed to identify the positions of LEDs in camera feed. Then the algorithm further developed to include the logic to overlay virtual information at exactly the same position marked by LED lights. This way the algorithm has been formulated on logical grounds. We started from a conceptual system and now we have developed a logical flow or logical system to realize the concept. Then technology tradeoffs were studied carefully. What technologies were needed to realize the system and which among many alternatives should be chosen has been evaluated. Finally A prototype was developed in its preliminary form using the physical components identified during technology trade off step. The prototype included four LED lights installed at target site. LEDs transmit a simple frequency based information. This information is read by image sensor in a high speed camera. The intensive computing power needed to read light information has been aided by specialized GPU processor. Finally some predefined information is augmented to live camera feed. To elaborate more the predefined information could be an image or text also known as virtual information in AR terminology. AR terminology real environment refers to live camera feed in this context. The output of the prototype were discussed in detail and different test cases were formulated. The system was then verified against the requirements collected earlier. Both plan and product has been validated. The plan was validated through a survey where users were asked to rate the usefulness of the conceptual system. Then the product was evaluated for the fitness to be used.

In this research work AR system has been conceptualized and realized using visible light communication. It is believed that this system overcomes the limitation of past research and existing products using visible light communication. Unlike existing AR systems using visible light which do not align virtual information over real environment accurately or which do not work in all lightning conditions this system aligns virtual information accurately and it has been tested to work under different lightning conditions. In addition it doesn’t limit the size of display area. The display area is defined by the LEDs installed in target area which could be changed without affecting the system functionality.

The system was developed to mark the exteriors of a building and display useful information but this is only one use case of the system. It could be used in exhibitions to display information at some specific place describing about the artifacts. It could also be used to describe each section of some historic monument and many more.

9 Future Research:

This research used a high speed camera and blinking lights at a frequency which is not apparent to human eyes. Most commercially available tablets and cell phones have a camera of specification 30 frames per second. The concepts developed in this research work could only be used in this scenario if the frequency of LED is less than 15Hz. At this frequency the blinking pattern of light is visible to naked eyes. Hence this concept is not applicable to mobile devices in its original form. Further research is needed to make similar solutions work on mobile platforms.

Mobile platforms are not only limited by the capacity of camera but the computing capacity is also limited. The algorithm needs to be modified to use some other sensors like gyroscope or pedometer to track the position of LEDs once they are located (using the logic in this algorithm).

If data transmitted using visible lights needs to be read with high accuracy, some wired or software based synchronization technique between LED blinking pattern and camera frame rate is recommended. The other way is to increase the number of frames captured initially but this will significantly impact the system performance.

Signal to noise ratio of this algorithm could be significantly improved by increasing the number of frames captured initially. On the other hand it will impair the system performance. Hence the algorithm needs to be optimized to minimize the impact.

10 Bibliography

- AR, A. (2015, 06 10). *Computer Graphics and Scientific Visualization*. Retrieved from http://www.cs.unc.edu/~azuma/azuma_AR.html
- Azuma, R. T. (1997, August). A Survey of Augmented Reality. *Presence*, pp. 355-385.
- Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent Advances in Augmented Reality. *IEEE*.
- Boost. (2015, 07 07). *Rtree Introduction*. Retrieved from Boost.org: http://www.boost.org/doc/libs/1_58_0/libs/geometry/doc/html/geometry/spatial_indexes/introduction.html#ftn.geometry.spatial_indexes.introduction.f2
- Cambridge, D. (2015, 07 11). *Meaning of Real-time*. Retrieved from Cambridge Dictionaries Online: http://dictionary.cambridge.org/dictionary/british/real-time
- Casio. (2015, 07 03). *Picalico*. Retrieved from Casio Picalico: http://picalico.casio.com/en/
- Company, M. &. (2012). Lightning the way: Prespectives on the global ligtning market.
- CrunchBase. (2015, 07 04). *Sekai (World) Camera*. Retrieved from CrunchBase: https://www.crunchbase.com/product/sekai-world-camera
- CV, O. (2015, 07 07). *Basic Thresholding Operations*. Retrieved from Open CV documentation: http://docs.opencv.org/doc/tutorials/imgproc/threshold/threshold.html
- Guttman, A. (1984). R Trees A dynamic index structure for spatial searching. *ACM*, 47-57.
- Haruyama, S. (2013). Visible Light Communication. *ITU Kaleidoscope Academic Conference.*
- http://ilovenewwork.com/2009/10/02/lego-ar/. (n.d.).
- Komine, T., & Nakagawa, M. (2003). Integrated System of White LED Visible Light Communication and Power Line Communication. *IEEE*, 71-79.
- Matsushita, N., Hhara, D., Ushiro, T., Yoshimura, S., Rekimoto, J., & Yamamoto, Y. (2003). ID Cam: A smart Camera for Scene Capturing and ID Recognition. *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 03).* IEEE.
- Milgram, P., & Kishino, F. (1994, December 12). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems,*.
- Naimark, L., & Foxlin, E. (2005). Encoded LED System for Optical Trackers. *Proceedings of International Symposium on Mixed and Augmented Reality* (pp. 1-5). IEEE.
- opencv.org. (2015, 07 09). *About*. Retrieved from Opencv.org: http://opencv.org/about.html
- Prince, S. J. (2012). Computer Vision: models, learning and interface. In S. J. Prince, *Computer Vision: models, learning and interface* (p. 395). Cambridge University Press.
- Robler, T., Rogge, S., & Hentschel, C. (2011). A Case Study: Mobile Augmented Reality System for Visualization of Large Buildings. *IEEE International Conference on Consumer Electronics* (pp. 311-314). Berlin: IEEE.

- Sutherland, I. (1968). Head Mounted Three-Dimensional Display. *Fall Joint Computer Conf. Am. Federation of Information processing Soc. (AFIPS)* (pp. 757-764). Thompson Books, Washington D.C.
- TechChurch. (2010, 07 14). *Augmented Reality App Sekai Camera Goes Multi-Platform. Adds API And Social Gaming*. Retrieved from TechChurch: http://techcrunch.com/2010/07/14/augmented-reality-app-sekai-camera-goes-multi-platform-adds-api-and-social-gaming/
- Wikipedia. (2015, 07 07). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/R-tree
- Wikipedia. (2015, 07 12). *Photosensitive Epilipsy*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Photosensitive_epilepsy
- Wikipedia. (2015, 06 04). *Visible Light Communication*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Visible_light_communication
- Yamazato, T., & Haruyama, S. (2014, 09). Image Sensor Based Visible Light Communications and Its Application to Pose, Position, and Range Estimations. *IEICE Transaction Communication VOL. E97-B*, pp. 1759-1765.
- You, S., Neumann, U., & Azuma, R. (n.d.). Hybrid Inertial and Vision Tracking for Augmented Reality Registration.