

LAPPEENRANTA UNIVERSITY OF TECHNOLOGY

LUT School of Energy Systems

LUT Mechanical Engineering

*Igor Soroka*

**DESIGN AND IMPLEMENTATION OF MACHINE VISION SYSTEM FOR THE MOBILE  
ASSEMBLY ROBOT**

Examiners: Professor Heikki Handroos

D.Sc. (Tech) Hamid Roozbahani

## **ABSTRACT**

Lappeenranta University of Technology  
LUT School of Energy Systems  
LUT Mechanical Engineering

Igor Soroka

### **Design and Implementation of Machine Vision System for Mobile Assembly Robot**

Master's thesis

2016

60 pages, 46 figures, 3 tables and 1 appendix

Examiners: Prof. Heikki Handroos  
D.Sc. (Tech) Hamid Roozbahani

Keywords: machine vision, digital camera, wireless transmission, pattern recognition, pattern matching, LabVIEW, video switcher, 3D video, mobile robot

The review of intelligent machines shows that the demand for new ways of helping people in perception of the real world is becoming higher and higher every year. This thesis provides information about design and implementation of machine vision for mobile assembly robot. The work has been done as a part of LUT project in Laboratory of Intelligent Machines. The aim of this work is to create a working vision system. The qualitative and quantitative research were done to complete this task.

In the first part, the author presents the theoretical background of such things as digital camera work principles, wireless transmission basics, creation of live stream, methods used for pattern recognition. Formulas, dependencies and previous research related to the topic are shown.

In the second part, the equipment used for the project is described. There is information about the brands, models, capabilities and also requirements needed for implementation. Although, the author gives a description of LabVIEW software, its add-ons and OpenCV which are used in the project.

Furthermore, one can find results in further section of considered thesis. They mainly represented by screenshots from cameras, working station and photos of the system. The key result of this thesis is vision system created for the needs of mobile assembly robot. Therefore, it is possible to see graphically what was done on examples.

Future research in this field includes optimization of the pattern recognition algorithm. This will give less response time for recognizing objects. Presented by author system can be used also for further activities which include artificial intelligence usage.

## **ACKNOWLEDGEMENTS**

It is no doubt, that I would like to thank Professor Heikki Handroos, Doctor Hamid Roozbahani, Juha Koivisto for their active support and expert advice. Although, I appreciate a lot other members of our project, they are: Andres Belzunce, Weiting Lee, Samrat Gautam, and Sameer Kunwor. I was happy to work with you, colleagues. By the way, I want to express my gratitude to Asia Khairetdinova and Kseniia Frumkina. Moreover, I would like to thank my parents for their mental help and faith in me on every stage of my writing and researching processes.

Igor Soroka

Lappeenranta 05.02.2016

## TABLE OF CONTENTS

### ABSTRACT

### ACKNOWLEDGEMENTS

### TABLE OF CONTENTS

### LIST OF SYMBOLS AND ABBREVIATIONS

<b>1</b>	<b>INTRODUCTION.....</b>	<b>6</b>
<b>2</b>	<b>THEORETICAL BACKGROUND.....</b>	<b>10</b>
	2.1 Digital video camera work principles .....	10
	2.2 Illumination.....	13
	2.3 Interfaces and connectors .....	15
	2.4 Wireless transmission and modulation .....	17
	2.5 Video switching .....	24
	2.6 Three-Dimensional view.....	25
	2.7 Pattern Recognition methods and augmented reality .....	27
	2.7.1 Template matching .....	28
	2.7.2 Color detection.....	31
	2.7.3 People detection .....	35
<b>3</b>	<b>PRACTICAL IMPLEMENTATION.....</b>	<b>38</b>
	3.1 Requirements.....	38
	3.2 Cameras .....	38
	3.3 Transmitter-Receiver Sets.....	42
	3.4 3D Generation and platform .....	43
	3.5 Video Switcher .....	46
	3.6 Frame grabber .....	47
	3.7 Program with use of LabVIEW and OpenCV libraries.....	48
<b>4</b>	<b>RESULTS AND ANALYSIS.....</b>	<b>51</b>
<b>5</b>	<b>CONCLUSION AND FUTURE WORK.....</b>	<b>53</b>
	<b>LIST OF REFERENCES .....</b>	<b>54</b>

### APPENDICES

APPENDIX 1: Machine vision algorithm

**LIST OF ABBREVIATIONS**

2D	Two-dimensional
3D	Three-dimensional
3G	3 <sup>rd</sup> Generation
4G LTE	4 <sup>th</sup> Generation Long-Term Evolution
B&W	Black and White
BGR	Blue Green Red
BNC	Bayonet Neill–Concelman
CCD	Charge-coupled device
CMOS	Complementary Metal Oxide Semiconductor
CP	Cyclic prefix
DLL	Dynamic Link Library
DVI	Digital Visual Interface
FFT	Fast Fourier Transformation
FPS	Frames per Second
HD	High-Definition
HDMI	High-Definition Multimedia Interface
HDTV	High Definition Television
HOG	Histogram of Oriented Gradients
HSV	Hue Saturation Value
IFFT	Inversed Fast Fourier Transformation
LabVIEW	Laboratory Virtual Instrumentation Engineering Workbench
LED	Light-Emitting Diode
MIMO	Multi in Multi Output
OFDM	Orthogonal Frequency-Division Multiplexing
PTZ	Pan-Tilt-Zoom
RGB	Red Green Blue
SDI	Serial Digital Interface
SVM	State Vector Machine
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Architecture Network

## 1 INTRODUCTION

The robotics creates the great mixture of science and technology. It uses findings from electronics, mechanics and computer sciences. As a part of technology, robots can be used in practically all kinds of human activities: from the military robots for securing South Korean border to domestic robotic vacuum cleaners. Moreover, it means that in industry automatic machines are widely used nowadays for purposes connected with improvements in efficiency and productivity. (Nava Rodríguez, 2011, p. 1.)

To be more precise, there are several levels of robot autonomy, which vary on the human participation in the human-robot interaction. According to proposed taxonomy by Beer et al, the lowest is 'Manual' level where the man acts without any help in sensing or planning from the central computer of machine. The highest level is 'Full autonomy', which means that robot can do all tasks without human at all. Although, 'Assisted Tele-operation' goes 'after 'Manual' stage in this taxonomy. This type gives the opportunity to robot to help man in performing tasks. Machine can advise to operator the trajectory or avoid obstacles if the distance is too low. (Beer, Fisk, & Rogers, 2014, p. 74.)

Word 'Teleoperation' can be interpreted as 'control from distance'. In this way 'distance' can be changed from 10 meters to millions of kilometers. In Figure 1.1 the scheme of typical tele-operated system is presented. The human operator on the station controls the Slave Robot through communication channel with the use of Master Robot. (Li & Su, 2015, pp 1–2.)



**Figure 1.1.** Scheme of tele-operated robot (Li & Su, 2015).

This thesis work is a part of LUT (Lappeenranta University of Technology) Mobile Assembly Robot project. The robot itself will use the 'Assisted Tele-operation' with haptic device (for example, joystick). The main channel of information for the human operator is vision. With the similarity of human body, the cameras for machine vision are situated in the head. The

operator will seat in the control room and control the robot, presented in Figure 1.2, by the means of joystick with haptics. The principle is very similar to closed loop system which means that operator sends commands through wireless network on site. The feedback for operator will be gathered through different sensors, like cameras, force sensors and many others. The camera system will generate recommendations for operator after pattern recognition on site. Robot is planned to perform actions such, as opening/closing valves with both hands, replacing defect electronics or even parts of some assembly, drilling and sawing by working tools.

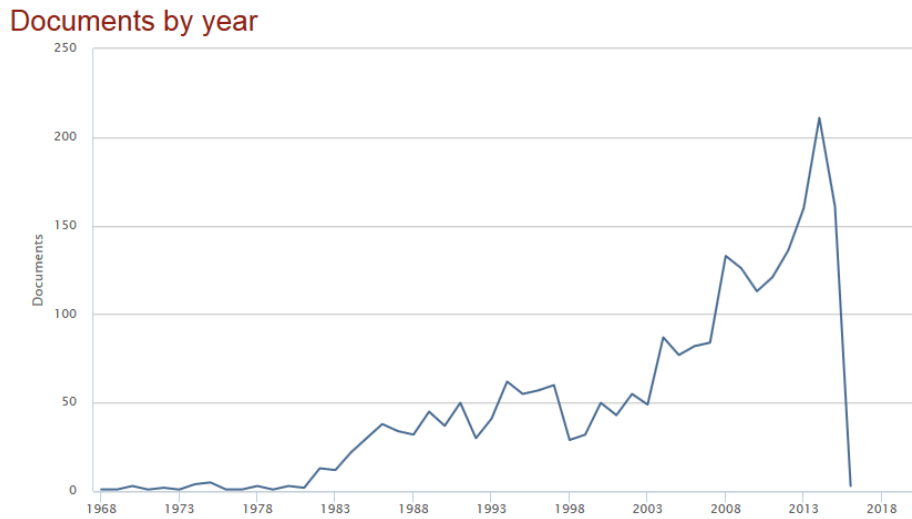


**Figure 1.2.** The appearance of Mobile Assembly Robot.

Machine vision is a topic of computer science, which considers software, hardware and image acquisition for real-time purposes. It simulates the functions of human eye like pattern matching, defining colors or shape detection. The idea is to learn computer to do the same functions in recognizing world around as human vision. (Davies, 2012, pp. 12–13.)

The attention to the topic of image acquisition, pattern recognition and machine vision has increased dramatically, according to the materials from Scopus database. The graph about the quantity of publications per year is presented in Figure 1.3 From the late 60s the awareness about this topic started to grow. During these years, the mathematical algorithms were created. It should be noticed, that the strong increase of materials on the topic has

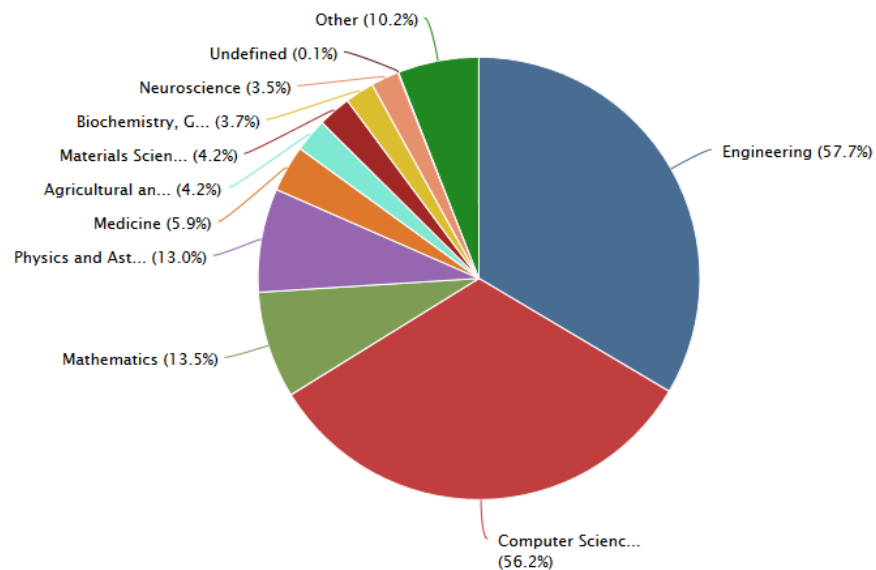
been started since 2000s. Moreover, last year was the most productive in this sphere of knowledge.



**Figure 1.3.** Number of materials by the request “pattern recognition & machine vision” (Scopus, 2016a).

In Figure 1.4 there is a pie chart which shows the distribution of materials by the field of studies. It is obvious, that Engineering (57.7 %) and Computer Science (56.2 %) are on the leading positions. Other relatively small percentages are showing main fields of the usage. For instance, ‘Mathematics’ is working on the improvement of the pattern recognition methods. To give another example, ‘Physics and Astronautics’ is the field where machine vision can be used like eyes of a planetary rover. It is proven, that the utilizing machine vision today is increasing in microsurgery (see ‘Medicine’ in Figure 1.4).

#### Documents by subject area



**Figure 1.4.** Subject areas of the considered topic (Scopus, 2016b).



The aim of this work is to design and implement the machine vision system with wireless transmission for tele-operated mobile assembly robot. On the way of doing this research work, there are practical and scientific problems. Practical problems - poor light conditions, distance between receiver and transmitter, which causes delays or interrupts, optimal in cost devices for the implementation. Scientific problems - veracity of the chosen methods, the optimal algorithm in implementing pattern recognition, correlation between expectations and real infield results. These problems cause research questions:

1. Can the problem of light be solved with the LED (Light-Emitting Diode) torch?
2. How to choose the optimal distance between transmitters and receivers to avoid delays and interrupts.
3. How to find the best relation between spent money and the quality of the equipment, which is needed for implementation?
4. How to check the veracity of the results and its correlation to expectations?
5. How to find the optimal algorithm in recognizing objects?

In this thesis there will be three chapters to find solutions of these problems and answer the research questions. In the first part all the theoretical background with literature survey is presented. It is related to camera operating principles, wireless transmission and signal modulation, generation of 3D video (three-dimensional), image acquisition methods and pattern recognition, also programming techniques and used integrated development environment - LabVIEW (Laboratory Virtual Instrumentation Engineering Workbench). Second part provides information about practical implementation of the work. It answers the question: what hardware is used for creating the stable system. Although, it will tell the reader about technical requirements of the system and specifications of the used devices. In the third part, one can find the practical results of this thesis with figures and stating what functions are implemented. Last part deals with generalization of results and future work.

## 2 THEORETICAL BACKGROUND

This chapter will provide the reader information about theoretical basis of the work. As it is mentioned in 'Introduction' chapter, vision system consists of both software and hardware. At the beginning, the concentration will be held on digital video sources, radio transmission and signal modulation. Furthermore, pattern recognition methods will be presented in the end of the chapter.

### 2.1 Digital video camera work principles

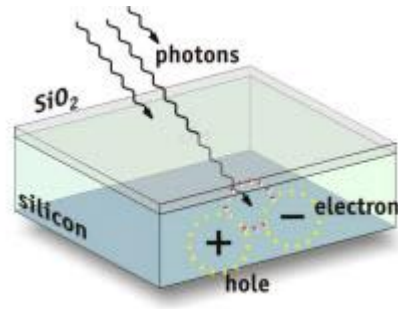
The basic inventions of recent twenty years, such as CDs, DVDs, HDTV (High Definition Television), MP3s, created the new era in the world. They based on the same principle. These inventions convert the analog world into a digital form on the screen of the computer. It means that all the information, no matter audio or video presents like a set of 1 and 0 bits. (Gurevich, Nice, Wilson, 2014.)

Optical signals came from the real physical world for conversion, which is called digitization. There are two processes in this transformation: sampling and quantization. Sampling is the process of changing continuous 2D (two-dimensional) plane into discrete regions. Samples lie on the rectangular grid which is the distribution function of intensity over a tiny region of space. In every interval or region, quantization gives an integer to the amplitude of the signal, so it becomes a determined level. (Poynton, 1996.)

It should be mentioned, that there are plenty of cameras on the market. They can use different technologies for capturing image - CMOS (Complementary Metal Oxide Semiconductor) and CCD (Charge-coupled device). The purposes of usage are also different: industrial cameras (with very high FPS (Frames per Second)), professional TV cameras for recordings (celebrations, talk shows, CCTV cameras with built-in encoding, high IP and sensors like movement, broadcast cameras for streaming video to the server or vision mixer remotely. The sizes are varying dramatically from 30 cm in length to 2 cm. All of these factors depends on how one will use a camera. It can be attached on some PTZ (Pan-Tilt-Zoom) platform, kept in hands or even fly with a quadcopter.

The video camera is required. The main difference of video camera is that it can immediately transmit the video stream through some digital interface instead of storing the image in some storage. It creates the broadcast video images in real time. (Woodford, 2007.)

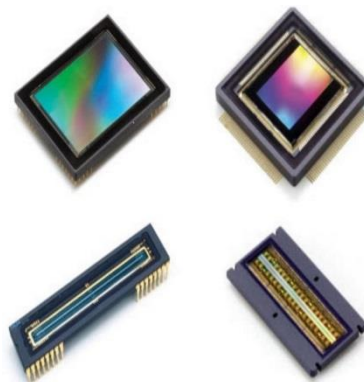
Digital video camera is capturing the light with a small lens at a front using CCD or CMOS imager sensors. Both types of image sensors they transform an optical image into an electronic signal. Light (photons) falls on a silicon wafer with electrons and holes, after that it is transforming into the electrical signal. This is called photoelectric effect, which is presented in Figure 2.1. (Teledyne Dalsa, 2014.)



**Figure 2.1.** The scheme of the photoelectric effect (Teledyne Dalsa, 2014).

CCD and CMOS are different types, which have its own features and drawbacks. Dr. Savvas Chamberlain created them in the late 1960s and 1970s. He is the 'father' of silicon image sensors. (Teledyne Dalsa, 2014.)

To begin the discussion about image sensors types, it is important to state the smallest element in digital graphics – pixel (originally 'pix' + element). This is a discrete element of a charge-coupled device in 2D picture. All together pixels form the image (Merriam-Webster, 2016). Figure 2.2 shows the two types of sensors: CCD in the left side and CMOS in the right part.

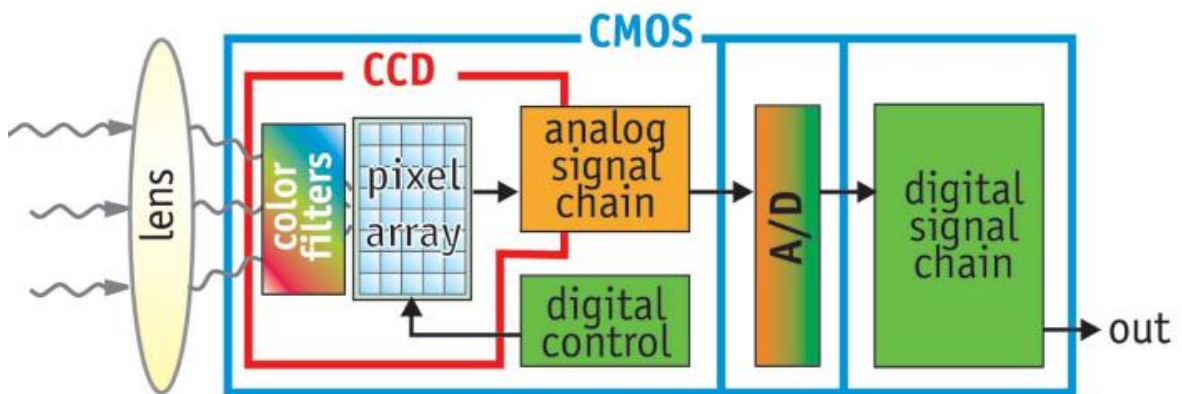


**Figure 2.2.** CCD and CMOS examples (Teledyne Dalsa, 2014).

In CCD image, sensor charge of each pixel is moving through mainly only one output node. Than it can be converted to the voltage, buffered and sent off-chip just as an analog signal. Therefore, all of the pixels assign to the mechanism of capturing light. Thus, the image produced by this type of sensor is in high quality, which means that output picture is uniform.

On the starting point, CCD was more in use than CMOS sensors by cause of superior image quality with the available manufacturing capabilities. (Teledyne Dalsa, 2014.)

On the other hand, there are CMOS sensors with some differences in the algorithm of digitizing the analog signal. In that image sensor, every pixel has its own charge-to-voltage switching. The general scheme is presented in Figure 2.3. One can notice that CMOS imager contains CCD inside to capture the photons. Although, it includes analog signal chain, digital control, analog-to-digital converter and digital signal chain (Litwiller, 2005, p. 55). The design is more complicated in comparison with CCD. Moreover, the effective area of light capture is becoming lower. While every pixel has its own conversion, the uniformity of the picture is becoming lower. However, this drawback creates the advantage because the speed of image capturing is dramatically higher with the parallel conversions. (Teledyne Dalsa, 2014.)



**Figure 2.3.** CMOS imager sensor general scheme (Litwiller, 2005, p. 3).

The choice between above-mentioned technologies should be done in accordance with an application of the camera. CCDs produce the highest possible quality of the image, which gives the opportunity to utilize them in spheres, like digital photography, broadcast television, and medicine. CMOS imagers can be used in places where high quality does not need and the size matters. For example, security cameras, video conferences, consumer scanners, biometrics and automotive industry, including machine vision and pattern recognition applications. (Litwiller, 2001, p. 3.)

For the application of machine vision, the two parameters are significant – speed of digitizing the analog image and low noise. A camera can achieve this with CMOS image sensors, which have higher speed performance, lower electrical consumption in comparison with

CCD. About the quality, it should be mentioned, that nowadays modern CMOS sensors are practically equal in high uniformity with a small amount of noise, like grains.

## 2.2 Illumination

When it comes to the development of the machine vision system, after choosing the right type of camera the designer should think about light conditions in the working area of the machine. Although, how the camera will interact with a light source and considered objects. In standards and in the science of life safety for estimating light conditions term 'illuminance' is used. It has a great influence on the way of machine vision will be performed. In some areas for the machine, it can be hard to have a smooth picture without any noise or inaccuracies of contours. To be more precise, often ambient light is not enough for pattern recognition tasks because working area can be situated in different locations with long dark time or, vice versa, in practically endless daytime (for example, Norway and other parts situated to the north of circle pole).

It is no doubt, that the source of light is always lamps. According to Daryl Martin (2014, p. 2), there are several types of lighting sources are used these days in machine vision area:

- Fluorescent;
- Quartz Halogen – Fiber Optics;
- LED;
- Mercury;
- Xenon;

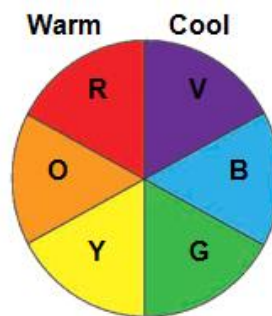
High-pressure Sodium.

First three types from above list are used mainly in the working areas of small and medium scale industry. Others are most frequent in use where the bright source is of the great importance. In microscopic applications, typically mercury lamps are widely spread because of having many discrete wavelength peaks. For usage with the requirements for bright strobe light, a xenon light source is irreplaceable. (Martin, 2014, p. 2.)

There are four major points about vision illumination by Daryl Martin (2014, p. 3):

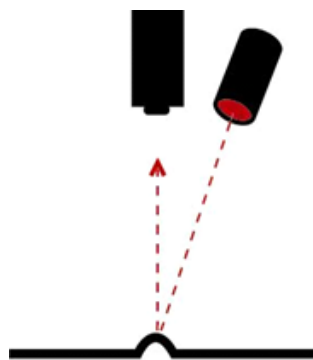
- Geometry – how camera, considered object and light source are situated to each other;
- Pattern – the shadow from object;
- Color – how the light from source reflected or absorbed by object;
- Filters – blocking or passing some wavelengths.

When a question comes to materials' reflection, it is important that they do it in various ways. It does not mean that it is happening in only B&W (Black and White) or color cameras because this is true for both types of images. With the usage of color wheel presented in Figure 2.4, the designer of machine vision system can clearly understand how to determine a contrast between a sample and its background. This happens on the grounds of a well-known rule that color reflects and the sample's surface is brightened, whereas, in contrary, color's absorption causes darkening of the surface. Moreover, if needed, to find colors on B&W with the use of filters or special colored light. (Martin, 2015a, p. 2.)



**Figure 2.4.** Color wheel with warm and cold colors (Martin, 2015a, p. 2).

The method of lightning presented on Figure 2.5. The everyday example of this type is how objects are lightened by sunlight. A light source directionally illuminates the object, and a surface of an object reflect the amount of light. A camera catches this by lens, therefore, the image will be illuminated (means that the light level of object becomes higher than in background area) and suitable for image acquisition. (Martin, 2015b, p. 2).



**Figure 2.5.** Directional lightning (Martin, 2015b, p. 2).

To summarize above said, it is significant to have clear understanding in what light conditions the machine vision system will work. The quantity of ambient light determines a necessity and position of additional light source or several sources. A nature of surface also contributes to the machine vision illumination. These factors are: color, roughness and even geometry form.

### 2.3 Interfaces and connectors

Camera creates the video stream during the operation time. There are two ways of processing this information. One is to store it on some flash disk. Another is to send it by video interface to the receiver directly. In this work, second choice is optimal because there is a need of live streaming. Thus, there will be a search of optimal solution in high definition video quality and machine vision applicability.

This chapter gives insight on the topic of video interfaces and connectors. Table 2.1 presents the comparison between main video signals by signal standards. The story of video interfaces started in the beginning of a digital era in photography and videography. In the table, there are only those that are used in modern television and machine vision industries.

*Table 2.1 Different signal standards and connectors (Wikipedia, 2015).*

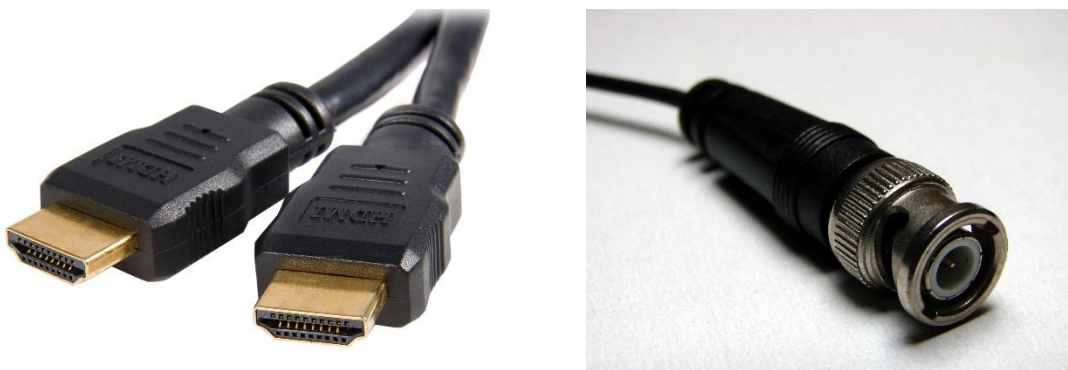
Signal name	Connector	Type	Max resolution	Used for
DVI (Digital Visual Interface)	DVI, Mini-DVI, Micro-DVI	Both	2560×1600@60, 3840×2400@33	Recent video cards
HDMI (High Definition Multimedia Interface)	19 pin HDMI Type A/C	Digital	2560×1600@75, 4096×2160@60	Many A/V systems and video cards (including motherboards with IGP)
GigE	Ethernet	Digital	1280x1024	Computer vision, industrial cameras
CameraLink	MDR26	Digital	1280x1024	Computer vision, industrial cameras
SDI (Serial Digital Interface)	BNC (Bayonet Neill-Concelman)	Digital	From 143 Mbit/s to 2.970 Gbit/s, depending on variant. 480i, 576i, 480p, 576p, 720p, 1080i, 1080p.	Broadcast video. Variants include SD-SDI, HD-SDI, Dual Link HD-SDI, 3G-SDI

After gathering data on these video standards and connectors, analog ones (composite video, SCART, S-video, VGA, DVI) are not good because for these type of data one needs converters to digital format which enlarges the calculations and processing time. This is the reason, why these standards were excluded from the Table 1. DVI is suitable for connection to monitor, TV set or display. Again, there will be a need of conversion because it is rarely used. GigaE and CameraLink are good for pattern recognition, but the aim of this work is to create tele operated robot which means that the image should be smooth and high quality for an operator's eye. HDMI and HD-SDI will be considered in details in further paragraphs.

First, HD (High Definition) is a name for a family of different screen resolutions. A number of pixels defines it. There are three resolutions 720p, 1080i and 1080p. For instance, 1080p means that 1920 (tall) x 1080 (width) and 'p' means 'progressive scan'.

Several interfaces transmit HD video. The well-known and widespread ones today are HDMI and HD-SDI. To understand what is better and why, it is important to compare these types of connection in details. Figure 2.6 shows the HD-SDI and HDMI connectors.

Secondly, it is necessary to give general information on the topic. HD-SDI was created to stream uncompressed data on 4:2:2 HD. The bitrate in this interface is 2.97 Gb/s (EBU – TECH 3299, 2010, p. 9). HDMI was designed especially for HDTV. It transmits video and audio simultaneously. The speed of connection is up to 5 Gbit/s. In next paragraphs, advantages and disadvantages of both standards will be discussed.



**Figure 2.6.** HDMI (left) and BNC Connectors (right) (Satinet, 2016; Wikipedia, 2013).

Consumer electronics companies, cable producers and movie industry are developing HDMI. Their main aim was not only transmission of HD video and audio but also to protect



their content. This security protocol forces the source device to create a security key for TV to display the video signal. Its name is HDCP. This protection type requires a synchronization between devices in order to transmit the key. That is why, when HDMI is plugged in the TV set there are nothing on the HDMI inputs list. In contrast, SDI is not required such handshakes which explains why SDI is not produce latency in switching between various devices and image lags. (Scully, 2013.)

Another point is a problem with connectors. HDMI is very similar to USB connectors that do not have mechanism of protecting the interface from plugging off in emergencies. HD-SDI uses BNC connectors, which are saved with specially designed lock system. (Scully, 2013.)

Although, HD-SDI and HDMI varies in cables. SDI is a standard coaxial cable (RG-6) which has relatively small production costs. In contrast, HDMI uses own type of cable with complex structure in addition, things like gold sputter coating. These reasons create the higher price for HDMI cables. Nevertheless, the maximum length of the HDMI cable without amplifiers is only 9 meter whereas SDI cable can have the range up to 300 meters without any delay and signal distortion. (Scully, 2013.)

As a consequence of above said, HDMI is a perfect solution for home use. Almost every video device supports today this interface, which has become the standard in HD video connection. Compare to HDMI, application of SDI is the most reliable and convenient decision for professional broadcasting field where quality, reliability, speed and range make the choice done.

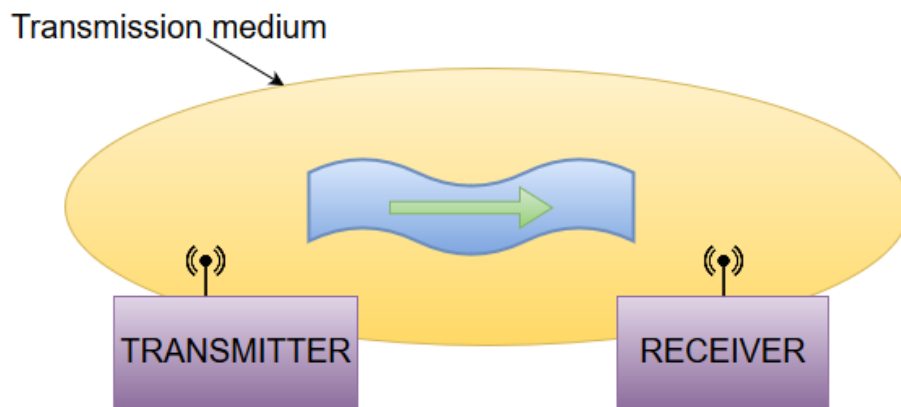
#### 2.4 Wireless transmission and modulation

As it is stated in introductory part, operator will use teleoperation or telepresence for controlling the mobile assembly robot. It means that it is obligatory to have smooth high definition image, which gives an opportunity to work efficiently with objects even quite small. To perform these actions, it is possible to use cameras of high quality. However, that is not the last problem to solve. Next question comes logically: what is the fastest way to transmit the continuous stream of video data?

There are two main ways of transmitting amount of data: by using cables or wirelessly. A choice depends on the purposes, distances, exact number of data and a working area. Therefore, there is no universal solution. In the case of particular mobile assembly robot, there is a need for wireless transmission. The reasons are high mobility of robot, the future

opportunity to work in areas in the distance more than line of sight, places with hazardous environment where cabling can be damaged.

It is doubtless that basics should be considered for starting the discussion about wireless transmission. Figure 2.7 shows a principal scheme of wireless communication link. It shows basic components in a simplified form. Transmitter with the help of antenna sends amount of data through a transmission medium, which in our case will be air. Transmission medium in the view of data transfer is the way through which, one object sends information to another. (Studytonight, 2016). In the figure it depicted by yellow ellipse to show that it effects on antennas and other devices. Usually sent data amount is going by radio waves (blue curved line with green arrow showed the direction). Receiver after some period takes this data by means of own antenna.



**Figure 2.7.** Basic parts of wireless communication link.

Today wireless connection is very closely connected to the term “WLAN” ( Wireless Local Architecture Network). It means computer network operates on some frequency band with supported maximum data rate (amount of information in bits transferred per second). The most used WLAN standard is Wi-Fi, which can be found practically in every place people live. Figure 2.8 approximately illustrates this statement. There is also practically endless number of home networks, closed working area networks and peer-to-peer networks. It is predicted to be 10 Billion public networks by 2018 year. Moreover, since 2014 the number of devices connected to wireless networks exceeded the Earth’s human population. (Visualistan, 2015.) That is why, the Internet of Things started. That means physical objects with controller, sensors and actuators connected to Internet. (McEwen & Cassimally, 2014.)



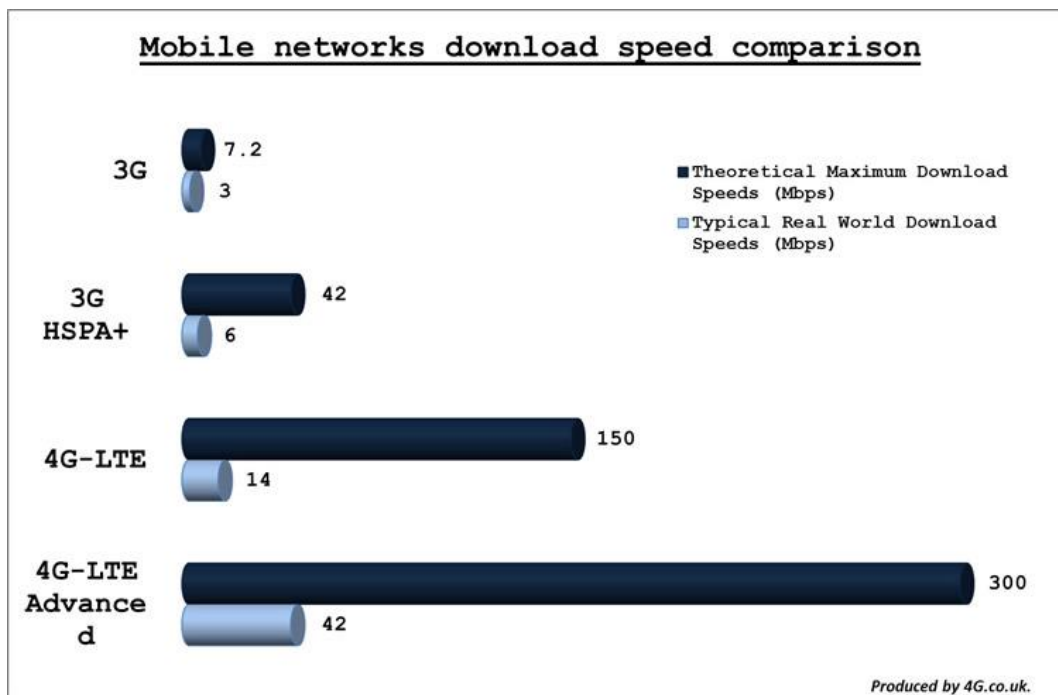
**Figure 2.8.** Growth of public Wi-Fi networks (Visualistan, 2015).

The story of wireless networks started in 1990. The first WLAN worked at 900 MHz with speed up to 1 Mb/s. It was slower in 10 times than wired connections of those times. At the same times, the IEEE 802.11 project started to create standards and working algorithms. This is the original standard but there are several modifications, which show the development of wireless networking. For instance, in 1999 after issuing 802.11a and 802.11b WLANs the fame of this technology started to increase. Although, the data rates became higher – up to 11 Mb/s and frequency 2.4 GHz. Later such technologies like MIMO (Multi In Multi Output) and OFDM (Orthogonal Frequency-Division Multiplexing) were introduced, the operational frequency 5 GHz, which will be discussed in following paragraphs. Changes to standard were done by amendments. After several amendments summarized, the IEEE association creates the full standard. Next standard will be published in 2016 (IEEE 802.11, 2016).

In ideal case, data streaming gives the opportunity for a user not to wait until some part of content is downloaded. Playback happens with no delay or even no delay at all, if it compared with real world event, which is captured. The challenge is to find the best way of sending streaming data from remote place to control room. In general, it can be done with creating own WLAN locally or using existing wireless mobile network. In next paragraphs, advantages and disadvantages of both methods will be considered.

With cellular networks, one can use 3G (3<sup>rd</sup> Generation), 4G LTE (4<sup>th</sup> Generation Long-Term Evolution) and WiMAX (Worldwide Interoperability for Microwave Access). The speeds of the most frequently used technologies are presented on Figure 2.9. A one can see from the picture – the best available option now is 4G-LTE Advanced. One of the

advantages of using 4G is a fast reliable connection on high speed in different areas. Furthermore, there is a need only of modem or router with 4G capable sim card, which will send the data to remote server. The operator station will be as a listener of this internet stream. Disadvantages are not stable cellular networks' coverage and the fast possible speed will not be achieved. For example, in rural areas or even inside some buildings signal of mobile network is not the same as in cities or places where many antennas placed. Thus, for the purposes of mobile assembly robot is not stable and efficient solution.

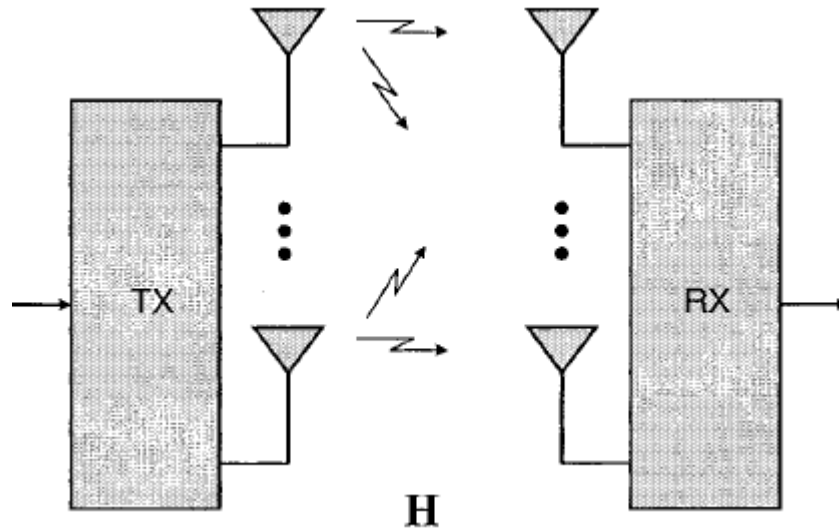


**Figure 2.9.** Download speeds of cellular networks (ITechnospot, 2015).

Next option is to establish local wireless connection on the same as pictured on Figure 10. So the idea is to send data from the operation area to workstation with operator. It will be local, which means that there will be no connection to internet. Video stream will be directed only to one direction. To ensure stream video playback without any delays and stops there is a possibility to use MIMO-OFDM technology. They were both created to increase the data transfer speed and signal stability.

The idea of using MIMO technology is as follows: to avoid interference, to realize diversity and array gain through coherent combining. It is for situation with utilization of MIMO only on one side. Nevertheless, when multiple antennas are used on two sides it gives another gain – spatial multiplexing gain. (Bölcskei, 2004, p. 31.) This type of gain concludes to incremented spectral efficiency which is the amount of data bandwidth extracted by considered technology from a specific amount of radio spectrum. (Rysavy, 2014, p. 1.)

In Figure 2.10 there is a schematic representation of MIMO system at operation. TX means transmitter, which creates the signal as output. RX is a receiver, which takes signal as input.



**Figure 2.10.** Scheme of transmitting and receiving radio signal with multiple antennas (Biglieri, 2005, p. 302).

Thus, the quantity of transmitting antennas is  $t$  and receiving antennas  $r$ . The channel gains can be described as matrix:

$$H = r \times t \quad (1)$$

With the use of (2), it is possible to assume that:

$$y = Hx + z \quad (2)$$

where, according to Biglieri (2005, p. 302): "H is an  $r \times t$  complex, possibly random, matrix, whose entries  $h_{ij}$  describe the gains of each transmission path from a transmit to a receive antenna,  $z$  - is a circularly symmetric, complex Gaussian noise vector. The component  $x_i$ ,  $i = 1, \dots, t$ , of vector  $x$  is the elementary signal transmitted from antenna  $i$ ; the component  $y_j$ ,  $j = 1, \dots, r$ , of vector  $y$  is the signal received by antenna  $j$ ."

After mathematical assumptions and transformations, it is possible to get:

$$y_j = \sum_{i=1}^t h_{ji} x_i + z_j, \quad j = 1, \dots, r \quad (3)$$

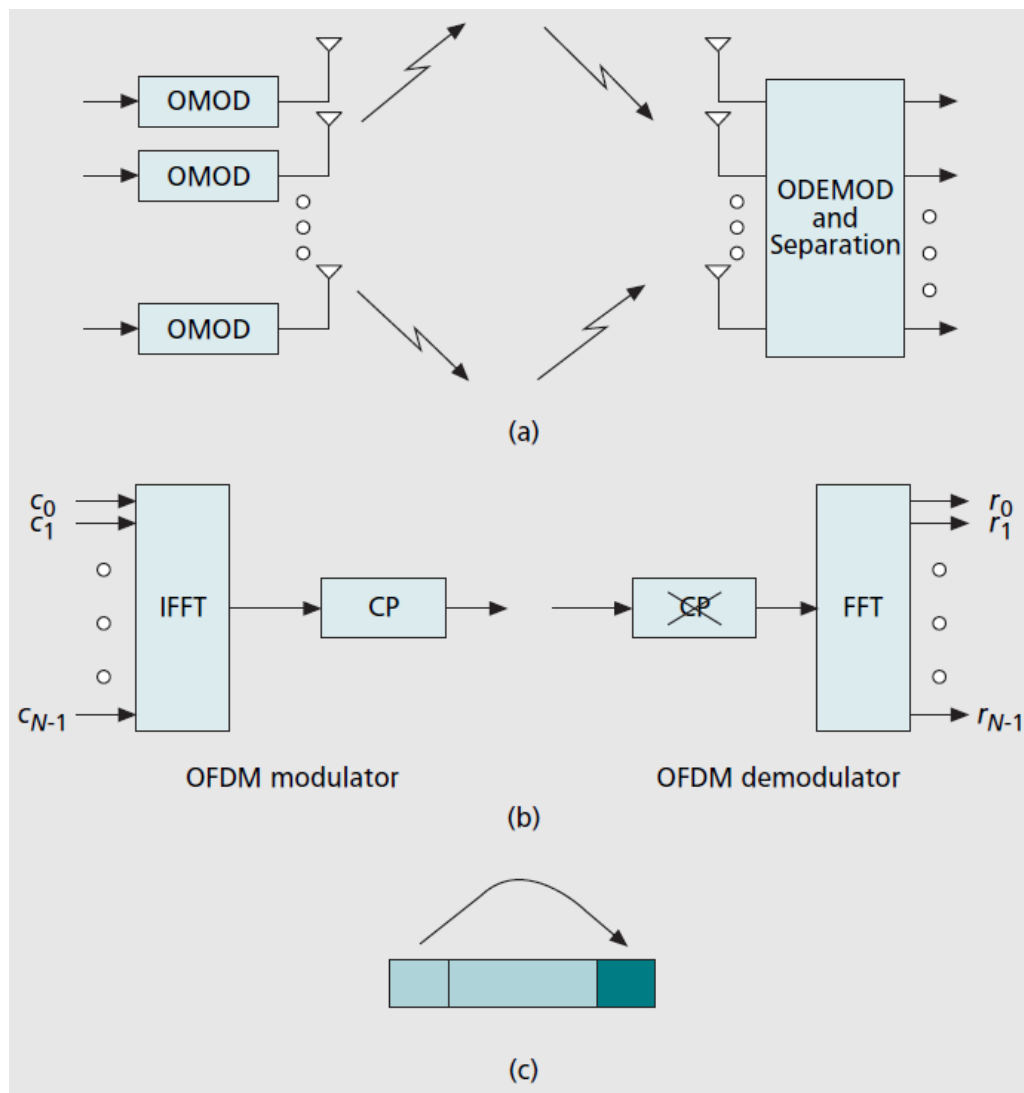
Formula (3) shows how each part of the received signal consists of a linear combination of the signal emitted by every single antenna, both from receiver and transmitter. It is possible to state that  $y$  influenced by spatial interference, generated by the signals from different

antennas. To avoid or manage with such interference, the single transmitted signals should be divided. (Biglieri, 2005, p. 303.)

The idea of using OFDM during wireless transmission is not new. It was firstly used in the format called DVB-T. It is primarily used for live broadcasting in television industry. This proves that OFDM is a concept with practically 50 years old history. However, it becomes truly widespread around the globe only in last 10 years. One of the most significant benefits in utilization of OFDM principle in wireless transmission is that it has all means to transform dispersive broadband channels into parallel narrowband sub channels; hence, it helps to reduce calculations at the receiver part of transmission. (Hui Liu & Guoqing Li, 2005, pp. 13–14.)

No one can deny that reality amends the ability of the radio signal to propagate in the correct and controlled ways. It creates some severities while a signal goes over wireless environment. There are envelope fading, which weaken the wave strength randomly, and dispersion, that changes the authentic signal waveforms in either time or frequency domains. (Hui Liu & Guoqing Li, 2005, pp. 13–14.)

Figure 2.11 shows the principle scheme for MIMO-OFDM system. OFDM's main principle is the addition of a guard interval, called CP (Cyclic Prefix), which is a copy of last part of the OFDM symbol shown in Figure 2.11c. The application of CP changes influence on the channel on the transmitted wave from a linear convolution into a cyclic convolution. This means that on the OFDM modulator-transmitter the IFFT (Inversed Fast Fourier Transformation) will be used, while on OFDM demodulator-receiver FFT (Fast Fourier Transformation) respectively. Therefor the overall frequency-selective channel is transformed into  $N-1$  number of parallel flat adding channels. As above said, this dramatically shorten a time interval for equalization which is the process of eliminating signal distortion during transmission. (Bölcskei, 2004, p. 32.)



**Figure 2.11.** a) Basic principle of MIMO-OFDM system (OMOD is a OFDM modulator, while ODEMOD means demodulator); b) single-antenna OFDM modulator and demodulator; c) adding the CP (Bölcskei, 2004, p. 32).

After discussing basic theoretical background of the modern wireless communication, it is time to talk about real-life products and its implementation of these principles, technologies and techniques. One of them is Wireless HD, which is a wireless standard of 60 GHz created for transmitting high definition video up to 1920x1080p. For development of the standard, distribution and promotion, WirelessHD Consortium is existed. After publishing the specification of WiHD version 1.0, the characteristics are:

- High definition uncompressed video streaming;
- High reliability;
- Low cost;
- Low power solutions for portable devices;

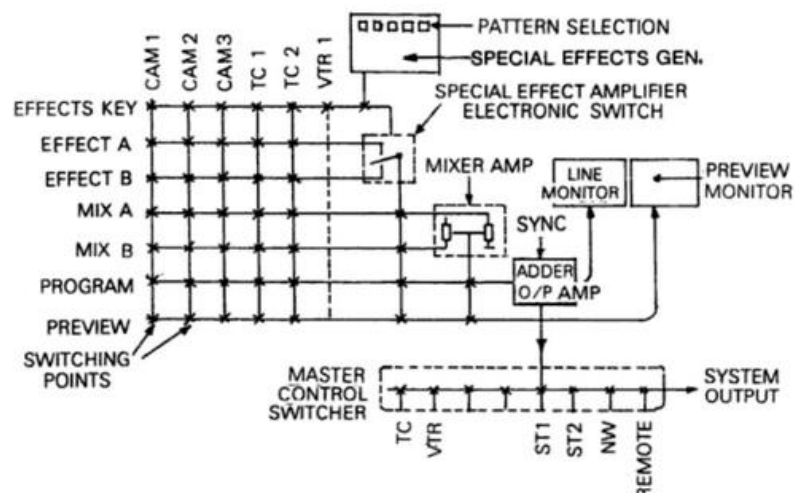
- Antenna technology which gives the opportunity to avoid line of sight constraints;
- Safe communication link.

All of the above mentioned is the requirements, which were successfully implemented, in this new standard. It is possible to add that video streaming goes practically without any delay or delay around 1 millisecond which cannot be spotted by human's eye. Low power consumption is one of the key advantages, which gives the opportunity to significantly use these systems in mobile robotics applications. MIMO-OFDM technology enables the way of transmitting signal even not on the line of sight. The manufacturers promise that link between transmitter and receiver is safe enough.

Overall, in the assembly mobile robot a wireless system with multiple antennas and under Wireless-HD standard will be used. The key factors for this choice are very high mobility, reliable connecting link, extremely high video quality. Mobility is achieved by total absence of wires for connection. Reliable data transfer channel is formed by MIMO-OFDM. Although, with the adoption of Wireless-HD the video stream follows without any encoding or other format conversions.

## 2.5 Video switching

A video switcher or vision mixer provides an opportunity to select between video signals from cameras, as they need on the screen of the display (Dhake, 1995, p. 121). Moreover, this device can show the combination of cameras, such as 4 in 1, split screen or even image in image. These techniques are widely used in applications where the quantity of cameras more than two. The most common applications are CCTV systems and television studios. Real-time Vision mixer is an essential part of live production. The use of video switcher in the case of mobile robot is a wise solution. (Xiao & Zhou, 2008, p. 180.)



**Figure 2.12.** Principle scheme of video switcher (Dhake, 1995, p. 121).



In Figure 2.12 the scheme of video switcher is presented. This structure is made by a bank with switches, mixing amplifier, special effects apparatus, sync adder-stabilizing amplifier and displays. Every vertical row has own single picture source, whereas every horizontal row correlates to an output bus. (Dhake, 1995, p. 122.)

## 2.6 Three-Dimensional view

The main drawback of every vision system is that one cannot estimate third dimension – depth of the image. That is why, it was decided that the mobile robot would have stereoscopic or three-dimensional video stream for an operator's convenience. The efficiency will rise because the operator can watch through robot's eyes like his own.

The idea of every 3D video generator is the same. It simulates how human's eyes watch at real world. It is obvious, that the image from the eye is different because of the location and viewing angle. Therefore, according to biology, the simplest method of 3D video generation is to synchronize two cameras. It is called binocular vision. (Ayache, 1991.)

In the beginning of 3D creation, there are two options. They are use of two cameras to shoot the video stream or one camera with simulation. Second choice is not as fast as first because there is a need of additional calculations for simulation of second camera. In the situation of mobile robotics, the best choice is to use two cameras for creation of stereoscopic image. In the paragraphs under, several types of creating stereoscopic images will be considered.

The oldest way of creating 3D video is to use anaglyph technique. For watching such image viewer must wear special red-blue, red-green or magenta-cyan glasses. The example is shown on Figure 2.13. Those glasses filters the incoming light information for a person's eye. According to Bing (2015, p. 15): "These glasses allow vertical or horizontal polarized light to enter the left or right eyes." It is although the cheapest way to produce 3D image on a simple screen of every device. The major disadvantage of this type of glasses is that some color information can be mixed and even lost. After filtering, image will be darker than in reality. It means that the image taken by brain will be poor in color. Furthermore, watching this image can be challenging after an hour of work. (Fernando & Worrall et al., 2013; Bing, 2015, p. 15.)



**Figure 2.13.** Various examples of anaglyph glasses (Fernando & Worrall & et al., 2013).

For watching movies in modern cinema theatres the display needs to be passive stereoscopic, whereas for anaglyph movies the viewer can use a normal screen. Such displays produce image with depth effect and special 'polarized' glasses divide each frame of the movie for both eyes. 'Passive' in this context means that there is no any special electric circuit inside a pair of glasses. Use of polarized glasses is an advanced version of anaglyph principle. The problem of this method covers in quality. The resolution for every eye can be maximum 1920x540, which is less than half of normal HD. (Fernando & Worrall et al., 2013.)

The name for another method of producing stereoscopy is 'Active stereoscopic display'. The most important there is a pair of glasses. In Figure 2.14, two pairs are shown. Left is for passive and right is for active. Inside this kind of glasses, there is a special liquid crystal layer. The idea is to blacken each eye in a very short period in alternate manner. The frequency is 60 Hz, which means that each eye is closed 60 times in a second. 3D display shows only corresponding eye's image. Glasses and display are usually synchronizing wirelessly through Bluetooth. In this type of producing stereoscopic image all previous drawbacks are eliminated. However, active technology can be tiring because of so often blinking of screen. (Fernando & Worrall et al., 2013.)



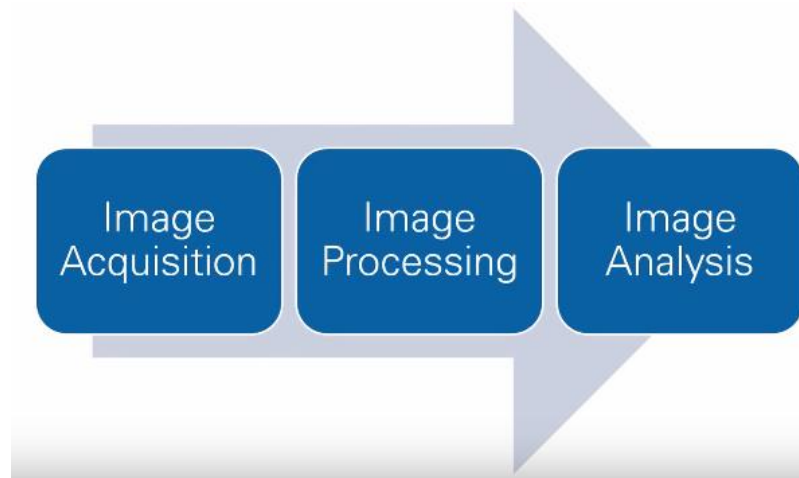
**Figure 2.14.** Passive (left) and active (right) shutter glasses (Ballegoie, 2012; Campbell, 2011).

For this project, it was decided to use an active stereoscopic display with special 2D to 3D converter, which synchronizes and produces input image for a 3D display via HDMI interface. Separate device dedicated only for one function gives the reliability and high performance.

## 2.7 Pattern Recognition methods and augmented reality

As above said, the idea of pattern recognition, as a division of IT and engineering is to reconstruct humans' identification methods to produce specific values. Pattern recognition is closely connected to artificial intelligence and augmented reality. In the project of mobile assembly robot, it will be used as a foundation for augmented reality. In practice, the computer will give advice to an operator for highlighting objects, which can be interested during an operation. As one can understand, it is a vast and modern topic needed to be narrowed for only several functions:

- Template matching for recognizing signs and some real life objects
- Color detection
- People detection



**Figure 2.15.** Basic algorithm (National Instruments India, 2014).

At first, there is a basic algorithm for creating an application, which do tasks related to vision systems depicted on Figure 2.15. In the first look, it can be divided into three main stages. For the first stage (Image Acquisition), the application needs to have camera and a frame grabber. Frame grabber feeds video stream to the program after initializing. Image processing is about doing mathematical actions for converting image into set of data suitable for further image analysis. Last stage is about extraction meaningful data from this data set. These stages consume many calculations, because the frames are of high definition and resolution, which means the enormous amount of pixels to proceed with.

### 2.7.1 Template matching

There are many different algorithms for using in template matching. In this work, I am limited by the IDE, which is used (LabVIEW). In the NI Vision module, there are two main template-matching algorithms - pyramidal matching and low discrepancy sampling. Both methods are using normalized cross-correlation method in the basis.

Both methods have two main stages learning and matching. In the process of learning, the algorithm takes gray value or edge gradient information, which can be found after this in the template image. In the matching stage, this same kind of information is finding in the video stream. It is compared with the gray value or edge gradient information from template to find particular region of interest to produce points of rectangle drawn around the object as an output. (Zone NI, 2013a.)

Normalized cross-correlation is often using when one image contained in another. So, it is a classic case of template matching. The aim is to determine the position of the template image in a 2D frame ( $f$ ). According to Briechle and Hanebeck (2011), the formula for normalized cross-correlation coefficient is as follows:

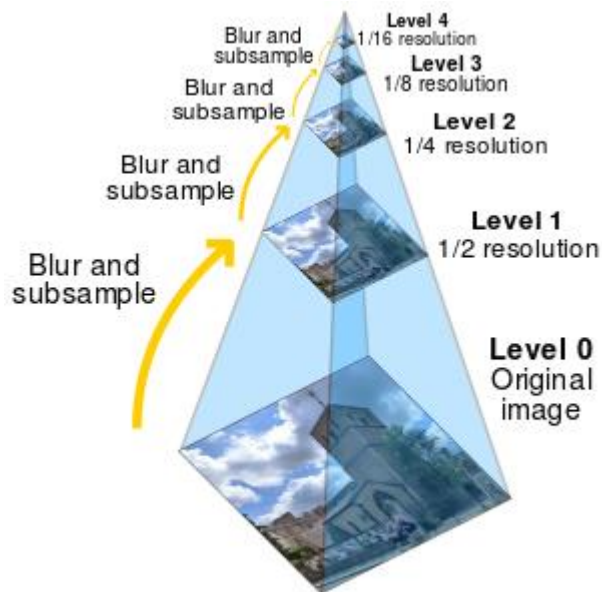
$$\gamma = \frac{\sum_{x,y} (f(x,y) - \bar{f}_{u,v})(t(x-u, y-v) - \bar{t})}{\sqrt{\sum_{x,y} (f(x,y) - \bar{f}_{u,v})^2 \sum_{x,y} (t(x-u, y-v) - \bar{t})^2}} \quad (4)$$

In formula (4)  $f(x, y)$  means the intensity value of the image  $f$  of the size  $M_x \times M_y$  at the point  $(x, y)$ ,  $x \in \{0, \dots, M_x - 1\}$ ,  $y \in \{0, \dots, M_y - 1\}$ . The template is presented by  $t$  of the  $N_x \times N_y$  size. The idea behind the formula is calculating position  $(u_{pos}, v_{pos})$  of the given template image with the help of value  $\gamma$  at each point  $(u, v)$  for  $f$  and the template  $t$ , which is by  $u$  on the  $x$  axis and by  $v$  steps on the  $y$  axis. The utilization of (4) in the pattern matching is more robust than other comparable techniques like simple covariance and the sum of the absolute differences (SAD). However, the main disadvantage is that the amount of calculations for this method is enormous. In the case of this project, this problem can be solved by using a powerful computer with the advanced video card inside. Hardware will be discussed in the next chapter. (Briechle & Hanebeck, 2001.)

The above-mentioned technique is good only when the image inside frame is not rotated and the size is not changed. Normally, normalized cross-correlation can recognize the same proportions image with the angle of turning in the range of  $-10^\circ$  to  $10^\circ$ . When the scale is unknown there will be a need to resize the template image and repeating of correlation

procedure from the beginning every change of the size. This will add more calculations for comparison. With the rotation, the situation is worse for the reason that the unknown nature of turning can cause lots of excess computation. (Zone NI, 2013a.)

The solution of this issue is to use pyramidal matching. It is a method of reducing sizes of the image and the pattern. With the use of Gaussian, pyramids both of them are tested with smaller spatial resolutions. In the Figure 2.16 you can find a series of images illustrating the principle of a Gaussian pyramid. (Zone NI, 2013a.)

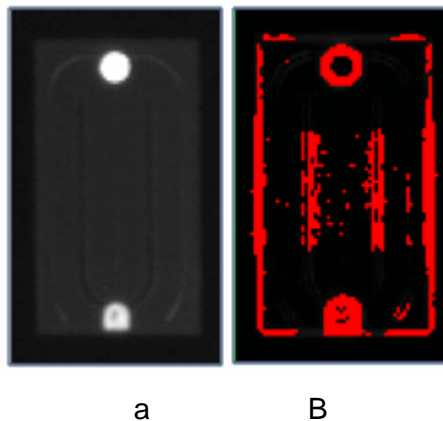


**Figure 2.16.** Gaussian pyramid (Wikipedia, 2016).

Maximum pyramid level for the considered pattern, in our case is 4<sup>th</sup> level, is determined during learning stage. The algorithm learns the information which is needed for depicting template extracting features possessed by this template and its turned types through all the pyramid values. Although, the method gives the 'optimal' pyramid level on which there is the fastest way of finding match. As it said before, the information can be of two types gray value (the foundation is each pixel intensity) and gradients (the foundation is edge data). (Zone NI, 2013a.)

In the gray value method, the features are normalized pixel gray values. It means that the problem of losing data can be avoided. It is working for the template with unstructured information with solid edges and complex textures. This technique is not suitable for poor lightning conditions. However, one can use in most cases. (Zone NI, 2013a.)

Another method for extracting information is gradient technique. After the learning phase during which the gradient intensity threshold is found, the image with edges calculated from the original image (see Figure 2.17). According to Gelsema & Kanal (2011, p. 347): This threshold divides the image pixels into two classes, the object pixels, which have a gray value of one side of the threshold, and the background pixels, which have a gray value at the other side of the threshold. The issue with lightning does not create inaccuracies, but a need of high resolution in this technique can be crucial. (Zone NI, 2013a.)



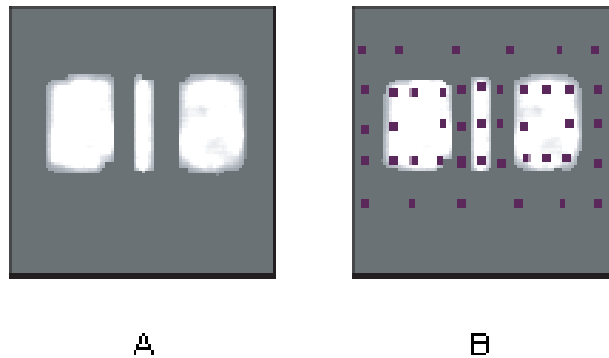
**Figure 2.17.** a – source grayscale image; b – image with detected edges (Zone NI, 2013a).

During the second stage (matching) of pattern recognition the starting point is the highest level of the pyramid (consider Figure 2.16). The resolution on this level is the lowest, which means that the sizes of template and source images is also the smallest. The correlation-based search can be started. During the sub-sampling, some of the details can be lost, that is why match locations are not fully trustworthy. The problem can be solved to find candidate matches with the best similarity degree instead of exact number of matches. (Zone NI, 2013a.)

After that, it is possible to go through other pyramidal levels with the changes in calculation of correlation scores. This method gives the simplifying because the regions around the best match candidates becoming smaller. (Zone NI, 2013a.)

The solution of the situation with rotated sample search will be based on the principles listed above. The best locations can be founded on the coarse angle step, but after the optimization takes place. It means that the refinement is on the process in lower pyramid levels for exact pattern recognition. (Zone NI, 2013a.)

The last technique is low discrepancy sampling. It is obvious that on the frame not all the information is necessary. When it comes to HD, video calculations of full image can play a crucial part in the speed of program work. In Figure 2.18, shows how intelligent sampling operates. It divides pixels into the region and edge ones. This plays a key role in an image understanding and pattern matching. (Zone NI, 2013a.)



**Figure 2.18.** Defining region and edge pixels (Zone NI, 2013a).

When it is necessary, NI Vision can recognize the rotation of the template. Intelligent sampling deletes unnecessary information and finds features that can implement stable and fast cross-correlation. Matching works as well with scaled patterns in the range  $-5\% \dots +5\%$  and rotated ones ( $0^\circ \dots 360^\circ$ ). (Zone NI, 2013a.)

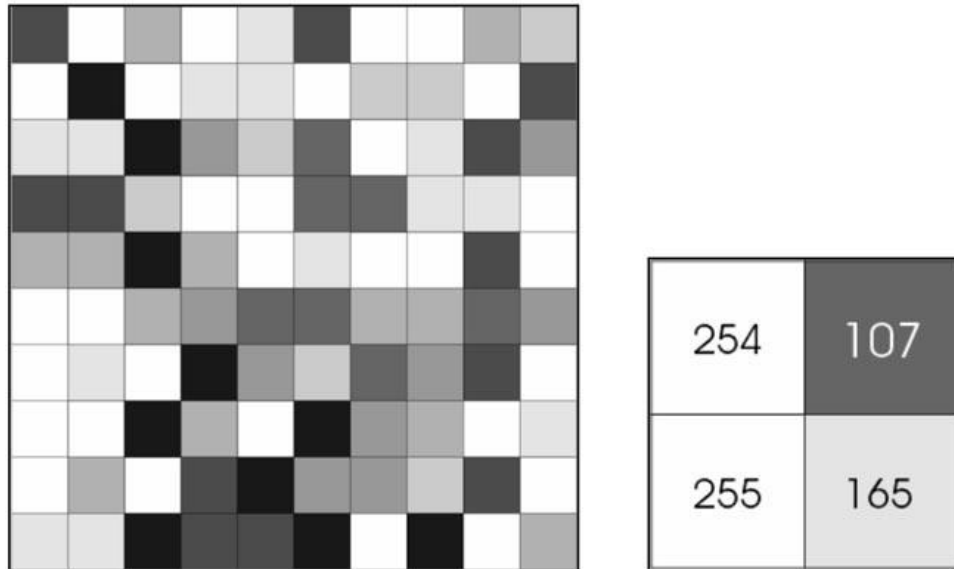
The choice of specific algorithm depends on the ambient conditions in the working area of camera. Sometimes it is better to use only one technique. However, developers of NI Vision give the opportunity to choose 'all' of the methods. The program decides itself what algorithm is better in speed.

### 2.7.2 Color detection

This is another task, which can be useful in the working environment. Color detection is a searching of the specified color in the image frame. In this section, the most reliable approach will be discussed with a short description of color modes, models and spaces.

In the beginning of this section, the information about color depth and color modes will be provided. Color depth (bit depth) is a maximum quantity of colors contained in a file or video image. For instance, 1-bit file is a black-white image. It can perform only two mentioned colors. The 8-bit image used in a section about pattern matching has 256 shades of gray,

black and white colors. In Figure 2.19, the example of this type is presented. Whereas 24-bit image file can store 16 million colors. (Pender, 2012, p. 25.)



**Figure 2.19.** Grayscale (8-bit depth) image representation (Hubble, 2016).

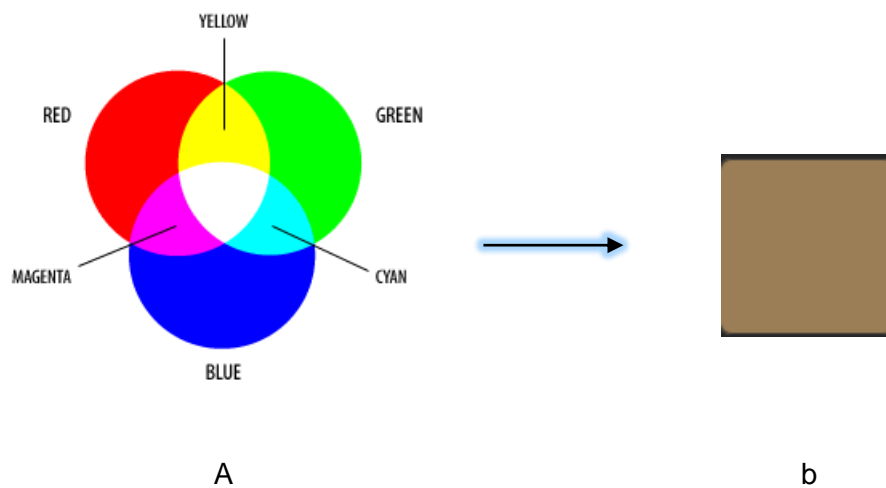
The color model is determined by color mode. Moreover, it is utilized for display and printing purposes. In this work devices will use grayscale images, BGR (Blue Green Red) and HSV (Hue Saturation Value). CMYK and L\*a\*b will not be considered because they are utilized mainly for printing while this thesis is about video acquisition and its playing. (Pender, 2012, p. 25).

One of the major points about the image representation is the conversion between standards and color modes. For example, RGB (Red Green Blue) can be converted to grayscale and not vice versa. The rule is only to downgrade the quality because the colors cannot be invented from nowhere. It is needed only to be limited. We are decreasing 16 million color to 256 only. Therefore, colors become shades while black and white are not changing at all.

The vision system starts with cameras. Every camera has attached specific color space. It is a virtual 3D space with coordinates showing all the palette of every color the device has an opportunity to perform. A number of color models in digital form describing those color spaces. (Pender, 2012, p. 26.)

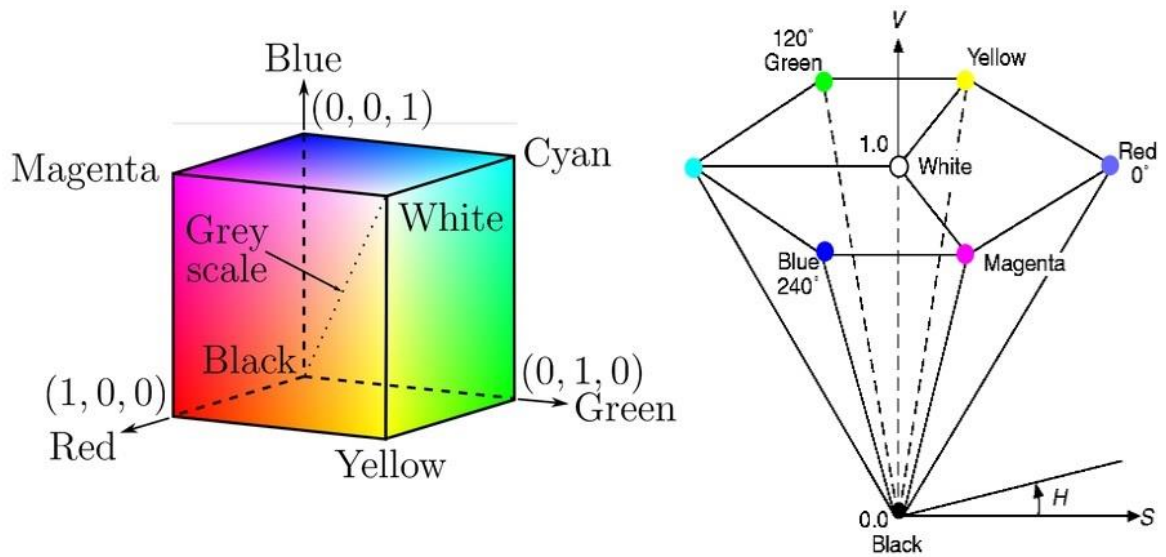


RGB color model is the most used in the sphere of video- and photography. It divides an image into three colors of light– Red, Green, Blue. These main components are mixing in various proportions to create a color. The demonstration of this principle can be found in Figure 2.20a. While in Figure 2.20b, it is possible to find the color got by values: R=156, G=126, B=86. The values of red (0–255 bits), green (0–255 bit), blue (0–255). Each value shows the influence of the particular color part. It although can be presented as percentage. BGR is a type of RGB model but in inversed form. Each pixel on an image or a frame has three-element array. In most of pattern recognition libraries, like OpenCV this type is used for feeding a program with video frames.



**Figure 2.20.** RGB model (a – color coordinate system (Pender, 2012, p. 26); b – color, which is got by values).

Another often used in machine vision appliances color space model is HSV. It consists of three constituents: hue (H), saturation (S), value (V). It is also utilized in drawing programs because of very deep description of colors. What is more, HSV is a cylindrical coordinate system showing RGB points. Thus, for better understanding the difference, Figure 2.21 shows RGB and HSV models as a cube and cylinder respectively. Hue is expressed in degrees as an angle. From the picture it is possible to obtain that 0 degrees means red color. (Ebner, 2007, p. 96.)



**Figure 2.21.** RGB model (left part) and HSV model (right part) (Black Ice Software, 2016a; Black Ice Software, 2016b).

The representation of the single color is different in HSV, but there are formulas for transformation between RGB and HSV. According to Ebner (2007, p. 98) they are, component by component:

$$V = \max\{R, G, B\} \quad (5)$$

$$S = \frac{\max - \min}{\max} \quad (6)$$

$$H = \begin{cases} \frac{1}{6} \frac{G - B}{\max - \min} & \text{if } R = \max \\ \frac{1}{6} \left( 2 + \frac{B - R}{\max - \min} \right) & \text{if } G = \max \\ \frac{1}{6} \left( \frac{R - G}{\max - \min} \right) & \text{if } B = \max \end{cases} \quad (7)$$

Nevertheless, all systems have its own drawbacks. In the situation of HSV, colors with low intensity can be fully saturated. It means that dark regions can have a noisy saturation. (Ebner, 2007, p. 98.)

The whole concept of finding the exact color on the image is about finding special values in HSV plane. RGB is not suitable for the color detection because it can cause a lot of calculations. Every component as R, G and B assume the amount of light reflected by the object and, consequently, with every other part of palette.

In real world, machine works in some place with the artificial light. Therefore, it is not effective to use only exact numbers of H, S or V because depending on the light conditions it is possible to find the range, in which the exact color is situated for finding it in the frame. Experimentally it is possible to find the ranges in which the color of the object is situated. The task on this stage is to create image like the one it is shown in Figure 2.22 The object will be white, while other image is looking like very black image.



**Figure 2.22.** Image after finding the exact color range in H, S, V (Fernando, 2013).

If the concept is to track specific object it is possible to use morphological transformations. Edge detection is also can help together with the pattern matching. After above-mentioned actions, there is a possibility even to track the specified object of particular color.

### 2.7.3 People detection

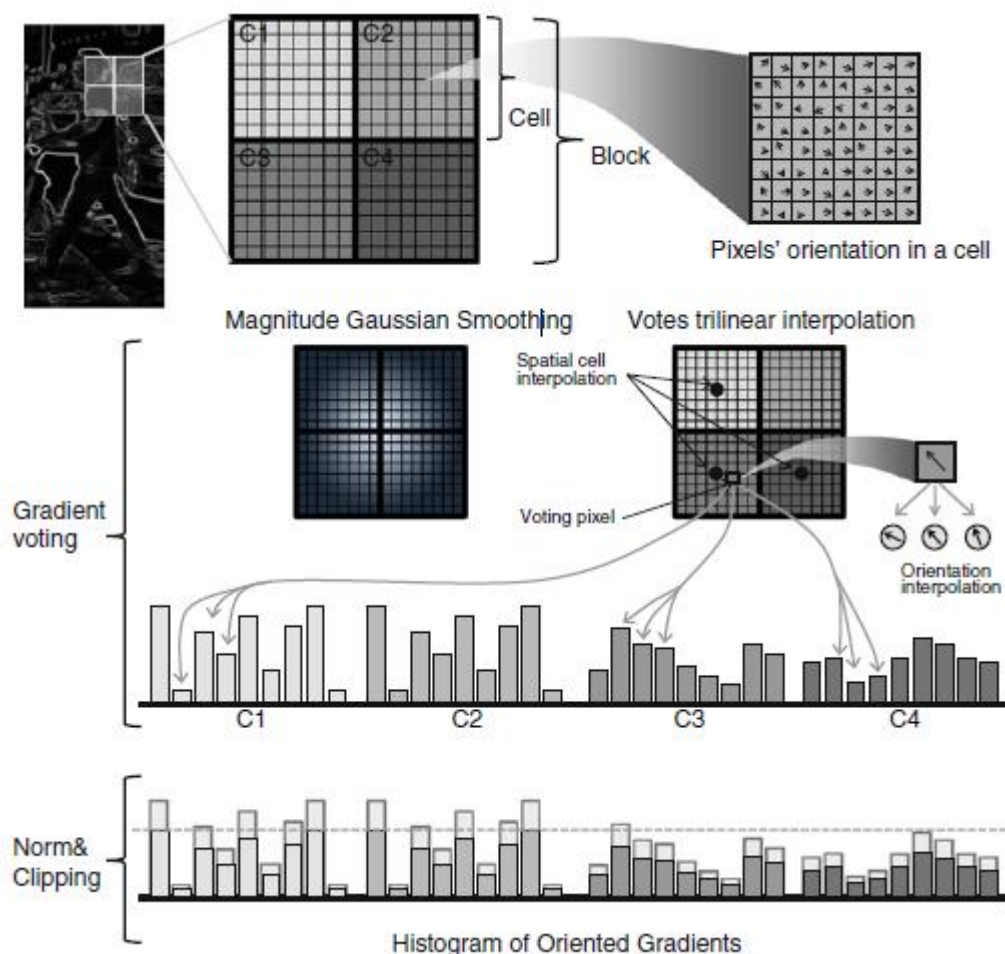
The next task of machine learning is harder because it requires more computational power and other resources. It is a very common task in the security cameras, which installed at some public places like squares, cafes and governmental buildings. Although, counting of automobile traffic is a very similar task.

In the mobile assembly robot, there is an urgent need of this function because for operator it is obligatory to know about person on site. This information can save people's life. The awareness of the operator is increasing. From time to time, it is can be heard from public sources that accidents are happening in different locations where human-robot interaction is involved. Advisory system in the case of person or people's detection will give a string of message with warning.

The idea of finding some similar objects is to train a classifier, which will understand presence or absence of certain object. In this case, silhouette or projection of a person. To achieve this result there is a need of training set with positive and negative samples. Positive samples will show the classifier the example of a person while negative shows an environment without a man.

The most successful way to find people in the frame is to use HOG (Histogram of Oriented Gradients) descriptor with SVM (State Vector Machine). It was proved in the article by Dalal and Triggs (2005). To start discussion, it is necessary to give understanding how HOG works with SVM.

HOG – Histograms of Oriented Gradients. It used these days not only for people detection but although for general object detection. HOG extracts gradient information from the image. Figure 2.23 shows the conceptual summarization of this method. As one can see, the method works with 8-bit image. This means that for starting the people detection there is a need of conversion from RGB to 8-bit image.



**Figure 2.23.** Main steps of HOG (Geronimo & Lopez, 2014, p. 34).

At the beginning, it is obligatory to have gradient information from the image. In each component of RGB, centered discrete differences are calculated. Onwards, in every pixel the max operation of the color component channels of gradient value resolves what channel color increasing ramp is utilized at every pixel. After that operation, each applicant frame is expected to be assembled of a rectangular cells' grid, consisted of rectangular cells, 8 x 8 pixels' cubicles are applied for people detection (Dalal & Triggs, 2005). Inside every cell, a HOG is calculated in a way like fixing k bins and accumulating the gradient magnitude. With the aim of shortening cell-induced aliasing, interpolation is utilized for allocating gradient values to nearest orientation bins and cells inside element called block, which is a grid of cells, their size is 2 x 2. (Dalal & Triggs, 2005.) Inside the block all cell histograms are connected as a block histogram. Further, the normalization of these histograms is happening. Actually, in advance there are calculations of the block orientation histograms of the cells, the gradient values of the pixels inside the block are down scaled in a way that there is a Gaussian filter in the middle of the block. The vector of features can be found with the help of linkage at all block histograms of an applicant frame. According to Dalal & Triggs (2005), the blocks inside an applicant frame intersect in one cell both in vertical and horizontal ways. Therefore, every cubicle adds four times, with various normalization, to the final window descriptor. These gives the resistance to illumination variations. When the person's pose changes HOG can handle this. After that SVM (state vector machine) can be used to train the classifier. The set with positive and negative samples is used as described before. (Geronimo & Lopez, 2014, p. 34–35.)

### 3 PRACTICAL IMPLEMENTATION

This section will provide information on the working solution of the above mentioned techniques, methods and algorithms. It although gives an understanding how the system is controlled. At the beginning, the technical requirements are considered. After that, the devices included in this work will be specified.

#### 3.1 Requirements

Technical requirements for every engineering system can be crucial. Engineer should clearly understand tasks for which he designs the system. It is an obligatory condition for a stable system, so it does not depend on size and quantity of elements.

The technical requirements for vision system of tele operated mobile assembly robot are:

- Video resolution – HD (at least 1280x720) at 30 FPS
- Latency – 1 ms maximum
- Distance range – 500 m (at least)
- View options: Image in image, quadratic view, Multiview, 3D vision, fast switching between any video source
- Type of transmission – radio-based technology

#### 3.2 Cameras

The quantity of cameras is four. This means that two will be installed on the top of the robot as eyes. Other two will be on the manipulators to watch at the process of doing the operation more precisely.

It should be mentioned, that there are plenty of cameras on the market. They can use different technologies of capturing reality (CMOS and CCD). The purposes of usage are also different: industrial cameras (with very high FPS), professional TV cameras for recordings (celebrations, talk shows, CCTV cameras with built-in encoding, high IP and sensors like movement, broadcast cameras for streaming video to the server or vision mixer remotely. The sizes are varied dramatically from 30 cm in length to 2 cm. Everything depends how one will use a camera. It can be attached on some PTZ (pan-tilt-zoom) platform, kept in hands or even fly with quadcopter.

Algorithm for choosing camera for an application:

- 1) Choose the type (amateur, professional, broadcast etc.);
- 2) Choose the output of the camera (SDI, HDMI, Ethernet);
- 3) Choose the quality type and megapixels (HD, SD, FPS);
- 4) Choose the dimensions of the camera and lenses;
- 5) Choose the brand or producer.

Our choice is broadcast cameras with small sizes. Their main advantage over others in our task that this type of cameras can continuously send data remotely by cable or wirelessly. It has special output that immediately uses the data from camera image converter. This type of cameras cannot record. They are created only for video streaming.

In Figure 3.1, the choice of main cameras is presented. The advantages of this camera are small size, specialization on broadcasting, very high resolution (up to 1920x1080) and frames per second (FPS), motion detection, reasonable price.



**Figure 3.1.** Main camera – Marshall CV500 MB (B&H Photo Video Digital Cameras, Photography, Camcorders, 2015a).

The most important technical specifications of the camera are presented in the Table 3.1. It has lens that can be replaced if more precision is needed in the image. Maximum frames per second equals to 59.94 that is much higher than, for instance, television standard PAL (FPS=30) or modern movie like 'The Hobbit' (FPS=48). (Fenlon, 2014.)

*Table 3.1. Specifications of Marshall CV-500 MB (B&H Photo Video Digital Cameras, Photography, Camcorders. 2015a).*

Sensor	1/3" 2.2MP CMOS
Lens	3.7mm M12 mount lens(interchangeable)
Frame Rate (FPS)	1080p60/50fps 1080i60/50 1080p30/25 720p60/50
Interface	1 x BNC (composite out) 1 x BNC (HD/3G-SDI) 1 x RS-485
Mount	1/4"-20
Power	Power Consumption: 150mA Voltage: 12VDC Connector: 2 x wire screw terminal
Dimensions (H x W x D)	

The next cameras will be Marshall CV200-MB presented in Figure 3.2. They will be attached to the manipulators of the robot. The operator needs these cameras for creating the full understanding of the environment and parts.



**Figure 3.2.** Additional camera – Marshall CV200 MB (B&H Photo Video Digital Cameras, Photography, Camcorders, 2015a).

In Table 3.2 the specifications are presented. This camera is a very reliable solution because of IP 67. (B&H Photo Video Digital Cameras, Photography, Camcorders, 2015b). The most remarkable specification is that IP67 (International Protection Marking on IEC



standard 60529). According to DSMT (2015) '6' means that "No ingress of dust; complete protection against contact (dust tight)". 'DSMT (2015) also mentioned that 7' means that "Ingress of water in harmful quantity shall not be possible when the enclosure is immersed in water under defined conditions of pressure and time (up to 1 m of submersion)" It also should be mentioned, that the range of operating temperatures allows working in harsh conditions. The power consumption is low and suitable for the manipulators attachment.

*Table 3.2. Specifications of Marshall CV200 MB (B&H Photo Video Digital Cameras, Photography, Camcorders, 2015b).*

Imaging Sensor	2.1 MP 1/3" Panasonic CMOS Sensor
Total Pixels	2010(H) x 1108(V) / 2,227,080 pixels
Active Pixels	1944(H) x 1092(V) / 2,122,848 pixels
Scanning System	Progressive Scan (Internal Sync System)
Video Output	HD-SDI, 3G-SDI, CVBS
Resolution	1920 x 1080i at 59.94 fps 1920 x 1080p at 59.94/29.97 fps 1280 x 720p at 59.94/29.97 fps
Horizontal Resolution	> 1000 TVL
Lens Mount	3.6 mm F2.0 M12 Mount - interchangeable, without cap
Electronic Shutter	Manual (1/30 or 1/25 to 1/30,000 sec)
AWB	AUTO / AUTO EXT/ PRESET / Manual (Red-Gain, Blue-Gain)
DNR	DNR (OFF/LOW/MIDDLE/HIGH)
Wide Dynamic Range	WDR (OFF/LOW/MIDDLE/HIGH)
Mirror Image	Mirror, Horizontal or Vertical Flip
Operating Temperature	-20 to 70°F (-10 to 50°C)
Operating Humidity	Under 90° Non-condensing
Power Supply	12 VDC, 1.0 A
Power Consumption	12 VDC 150 mA, 1.8 W
Dimensions (W x H x D)	0.7 x 0.7 x 2.4" (1.9 x 1.9 x 6.0 cm)
Weight	22 grams

To summarize, four cameras were chosen for attaching to the robot. They will take accumulators like an energy source for its operation. The quality of image for an operator gives the ability to work efficiently on the site.

### 3.3 Transmitter-Receiver Sets

As it is mentioned in section 1.4 it was decided to use wireless transmitter-receiver set with MIMO-OFDM technologies. Transmitter is situated on the robot and receiver is situated in the control room. The model name and type is Swit S-4914T/R SDI/HDMI 700m. This company was chosen among three other big companies like Teradek and Paralinx. These companies have the same types of devices but because of the big names in TV production industry they have overestimated prices. That is why SWIT brand is a good alternative with its very reasonable prices for devices.

This company has started from producing batteries for cameras. Now they have different products such as displays, multiviewers, video monitors and on-camera light. The company has participated in the NAB2015 show, which is the most important technical conference for broadcasters in the world. (SWIT Electronics, 2015a.)



**Figure 3.3.** Swit S-4914 transmitter-receiver set (SWIT Electronics. 2015b).

In the Figure 3.3 the devices are presented. These devices use MIMO and OFDM technologies which were discussed in previous section. This means that the latency of the video on operator's display will be less than 1 ms. Although the size of the transmitter is

created for on-camera use. This is a very important aspect, which gives the opportunity to attach 4 transmitters to the robot's body easily. Each set of transmitter-receiver supports only one HD-SDI input/output. (SWIT Electronics, 2015b.)

Technical features of Swit S-4914 T/R according to SWIT Electronics (2015b):

- “Professional WHDI (MIMO/OFDM) technology
- Uncompressed 3G/HD/SD-SDI and HDMI
- 2-ch SDI/HDMI embedded audio transmission
- 700 meters (LOS) transmission distance
- Less than 1ms latency
- 5.1-5.9GHz license-free operation frequency
- User configurable frequency bands
- Multicast – 1 transmitter to multi receivers
- Less than 63mW radio power
- AES 256bit encryption
- Mini transmitter designed for HDV camcorder
- Various DV plates for option
- All-metal housing, durable
- Pure hardware connection; Plug and Play”

The main technical features are presented in above. Moreover, the system supports and transmits the video stream up to 1920x1080p@60 FPS. This means that quality will be exclusively high. The effect of telepresence should be obtained. The power consumption is 6,5 W. 700 meters (LOS) means that this the range in the line of sight. If the obstacles indoors create problems, the range will be minimized to required 300 meters.

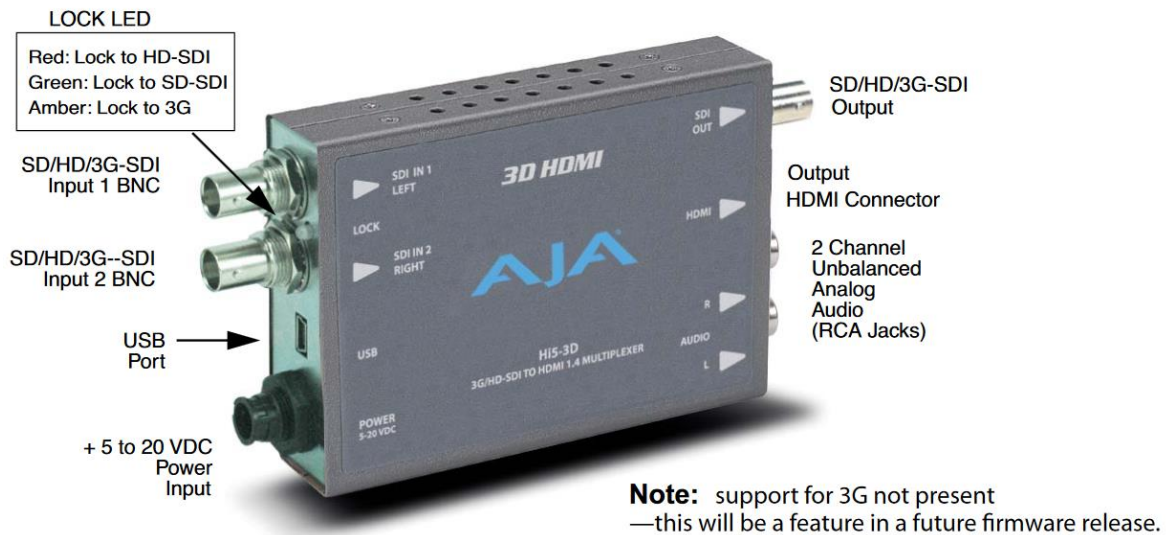
### 3.4 3D Generation and platform

It is decided that the 3D image will be generated by standalone device. This will speed up the operation of whole vision system. Although, it will save the computing time and increase operability.

The market of 3D generating devices is not yet very crowded by the products and solutions. Practically all of the 3D videos and cinemas were created in special software. For example, MAGIX Movie Edit Pro, which can handle with this task successfully. However, in this work

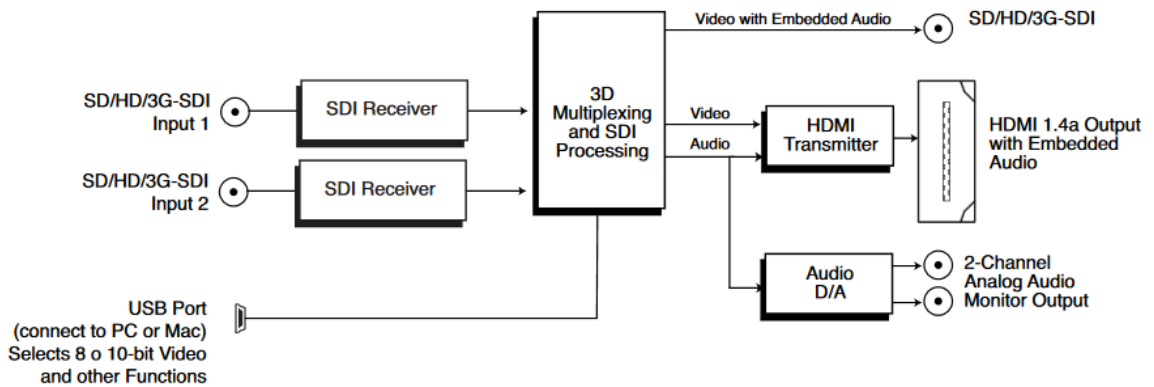
I am interested in a device situated in the control room. The main requirement is to generate smooth 3D image from the video stream in real-time.

There were two alternatives. One of them is Teranex 3D. It is a specialized device with various inputs on different purposes, which converts the input stream from 3D to 3D. Nevertheless, after checking the price it was decided to search more. Another option is AJA Hi5-3D. It is a mini-converter with only 2 SDI connections and 1 HDMI output. It is very ergonomically designed. The best thing is that the price is significantly lower than for Teranex 3D. The appearance of this device can be found in Figure 3.4.



**Figure 3.4.** Appearance of AJA Hi5-3D converter (AJA, 2011, p. 5).

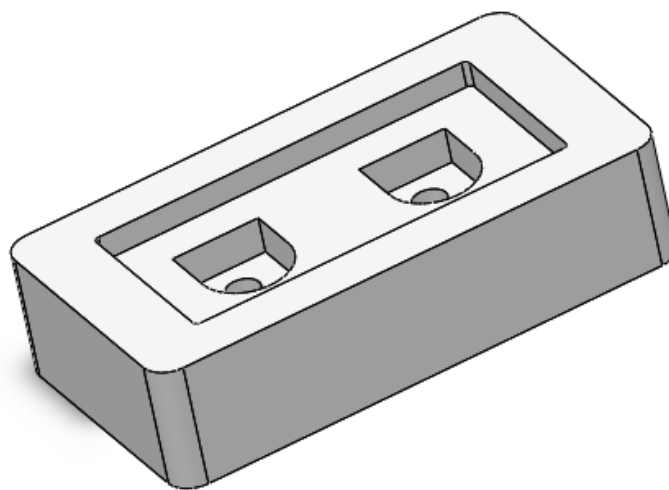
The block diagram depicted in Figure 3.5 shows the general and simplified overview of what is happening inside the box. The device catches the signal from two sources and after that the processing goes in the block related for producing the 3D image. In the end, an image converted to HDMI.



**Figure 3.5.** General block scheme of 3D converter (AJA, 2011, p. 3).

The device is using HDMI 1.4a which means support of 3D. After connecting to TV set with 3D capability, converter automatically finds the best way to produce stereoscopic video. In this study the converter was tested with Sony TV. It gives smooth and real-time video while a viewer wearing active 3D glasses.

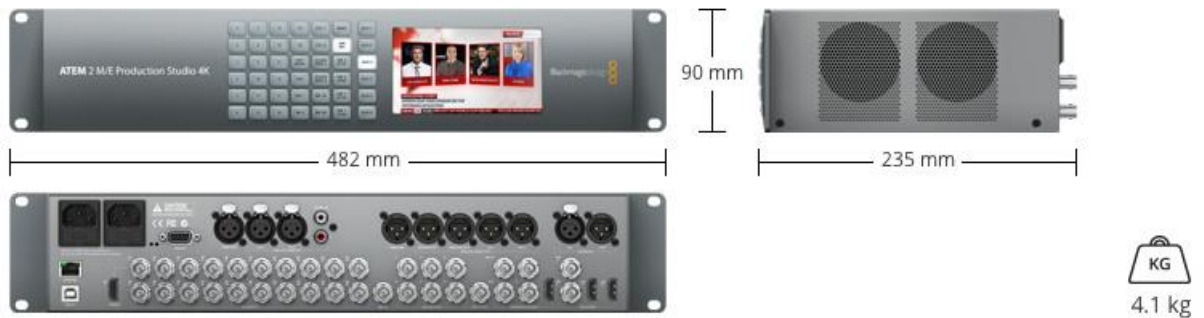
Another issue in creation of 3D image is the fixing of cameras. For synchronizing cameras and getting the right image with a stereoscopic effect one needs right displacement two cameras in the relation with each other. Experimentally, it was decided to connect cameras together by specially designed platform. Model of it, created in SolidWorks is shown in Figure 3.6. It has holes for screws to adjust each camera's fitting.



**Figure 3.6.** Design of platform for cameras.

### 3.5 Video Switcher

For the mobile robot, the choice is Blackmagic Design ATEM Production Studio 4K. It is a reliable vision mixer for live video switching on high speeds without loss of quality. It was produced for TV industry. The mixer is able to work with up to 20 inputs. Figure 36 shows the front and rear views of the device.



**Figure 3.7.** The front and rear views of the video switcher (Blackmagic, 2015).

This video switcher allows producing broadcast live multi camera production. It supports a wide variety of video sources. In the case of mobile robot teleoperation, the switcher will ensure smooth and fast video transition and several output display. In control room, the switcher will be controlled through Ethernet connection. There are 20 SDI and 1 HDMI inputs. The most important feature that this device has 4 SDI outputs for display because it is necessary to have 4 outputs. If it is needed, the device also has an opportunity to play audio stream from cameras. (Blackmagic, 2015.)

Although, with this video mixer the special software is included. Its name is ATEM Software Control. With this software, it becomes possible to switch between sources, to control cameras and even create special visual effects. All of these features can be done immediately. In Figure 3.8 the process of switching is shown. Everything is live and fast.



**Figure 3.8.** Software in operation (Blackmagic, 2015).

In this part, the video switcher was chosen for routing the video sources on to different screens with projectors. This system allows the user to see as detailed picture as possible for having the whole information about the situation in field but sitting remotely in control room without any danger.

### 3.6 Frame grabber

After vision mixer there is a need to feed the PC with video stream for further operations in software. To do this one, frame grabber will be used. It is an obligatory component of practically every machine or computer vision system. Main task of a frame grabber is to acquire images created by camera. Originally, frame grabbers were created to convert analog stream to digital, but these days they can help to convert between formats. Such devices can be connected internally or externally. (Sciammarella & Sciammarella, 2012, p. 246.)

In the case of mobile assembly robot it was decided to use external frame grabber with some widely used interface. Although, it was a need to have the easiest installation as possible on Windows OS. It was not very hard task because there is only one company at this market. The company specializes on various types of frame grabbers. In Figure 3.9 the operation principle is shown. The name of the device is AV.io SDI by Epiphan. It works with HD video up to 1080i. It has SDI input. It can be connected via USB 2.0 and USB 3.0. The speed will be high enough to be smooth and without delays. Moreover, the device is using driverless installation. It means that Windows recognizes it like another web camera. Frame

grabber although uses DirectShow. This will work great with LabVIEW and OpenCV libraries. (Epiphan Video, 2016.)



**Figure 3.9.** Scheme of frame grabber's operation (Epiphan Video, 2016).

### 3.7 Program with use of LabVIEW and OpenCV libraries

Before that section, the discussion was mainly about used hardware. However, for successful operation there is a need of software in which it is possible to create stable program for the pattern recognition and machine learning algorithms operation.

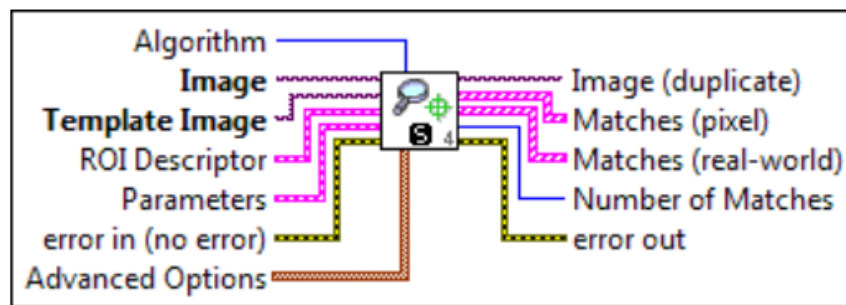
It was decided to use LabVIEW IDE with the OpenCV libraries. First of them has special add-on for doing operations over video. It is called Vision Development Module. It has all necessary codecs for grabbing video from the source. It uses IMAQ drivers for this. In Figure 3.10 there is an example of creating image with IMAQ initialization with image example. In this figure, one frame is shown from the file. By the function 'Grayscale (U8)' it converts an image to this format.



**Figure 3.10.** Creation of grayscale image in LabVIEW (Alor-Hernandez et al., 2011, p. 302).

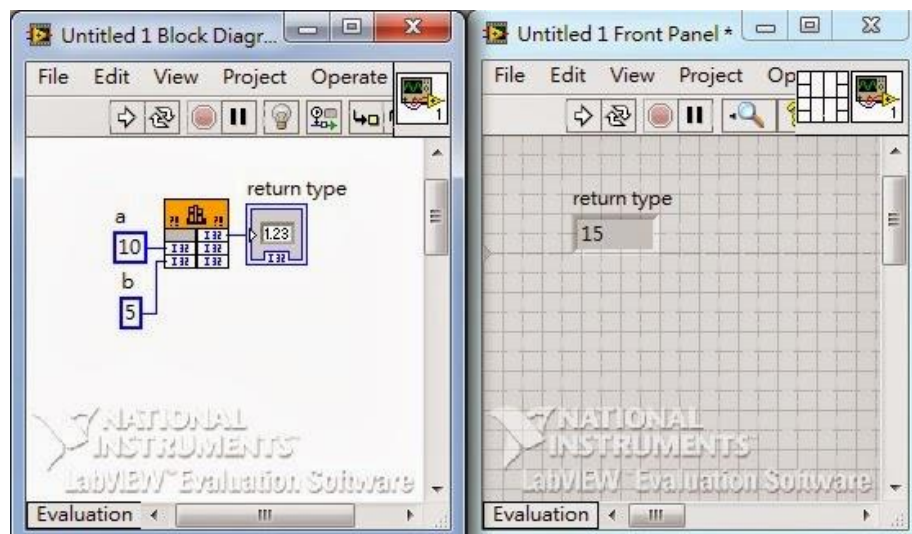


Although, LabVIEW has built-in functions to find templates, for instance. The principle was discussed in 1.7.1. IDE has blocks like Match Pattern 4, which is shown in Figure 39. Furthermore, used inputs (left, upper and bottom sides) and outputs (right side) will be discussed. In 'Algorithm' input can specify what technique to use for recognition (discussed in 1.7.1). 'Image' is a video stream fed to the block. 'Template Image' is a IMAQ created image (Figure 3.11) from the file of any image file format. 'Matches (real-world)' gives an array of match clusters, which can be afterwards used for detecting position of overlaid rectangle.



**Figure 3.11.** Match pattern block in LabVIEW (Zone NI, 2013b).

OpenCV functionality is connected to the LabVIEW using DLL (Dynamic Link Library) format. It is a library with code inside; in our case, C++ code with Open CV libraries connected. DLL files help to create separate modules (Microsoft, 2007). In this project, it is used to add some features to the program. For example, machine learning algorithm. The dll is a piece of code, which looks like normal block in LabVIEW. It is called by the function 'Call Library Function Node'. See Figure 3.12 for understanding the principle.

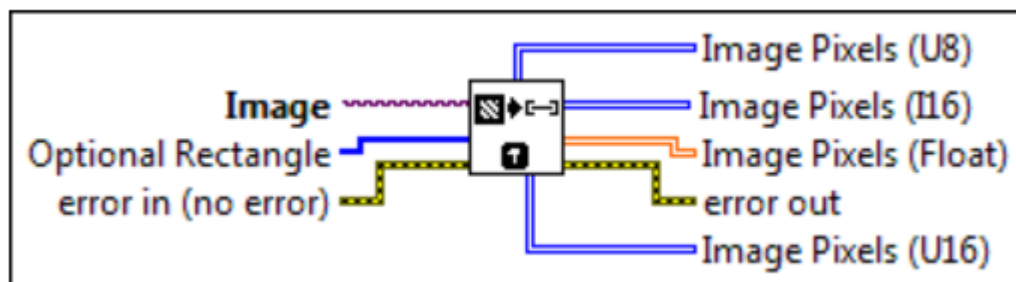


**Figure 3.12.** DLL example in LabVIEW.

The problem here is that LabVIEW and OpenCV are using different types of image formats. It means that somehow there is a need for conversion. In this case, the basic understanding of image can help. Image is a matrix of  $m \times n$  size. Each element is a point of the image (pixel). Each element can have value in dependency of image intensity. It is represented as

$$I = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad (8)$$

It is possible to use a LabVIEW function 'IMAQ ImageToArray VI', which gives as an output a 2D array with, values of every pixel. Consider Figure 3.13 for more information. After using this block, it is possible to use output of it as an input for openCV function where some learning algorithms will be done.

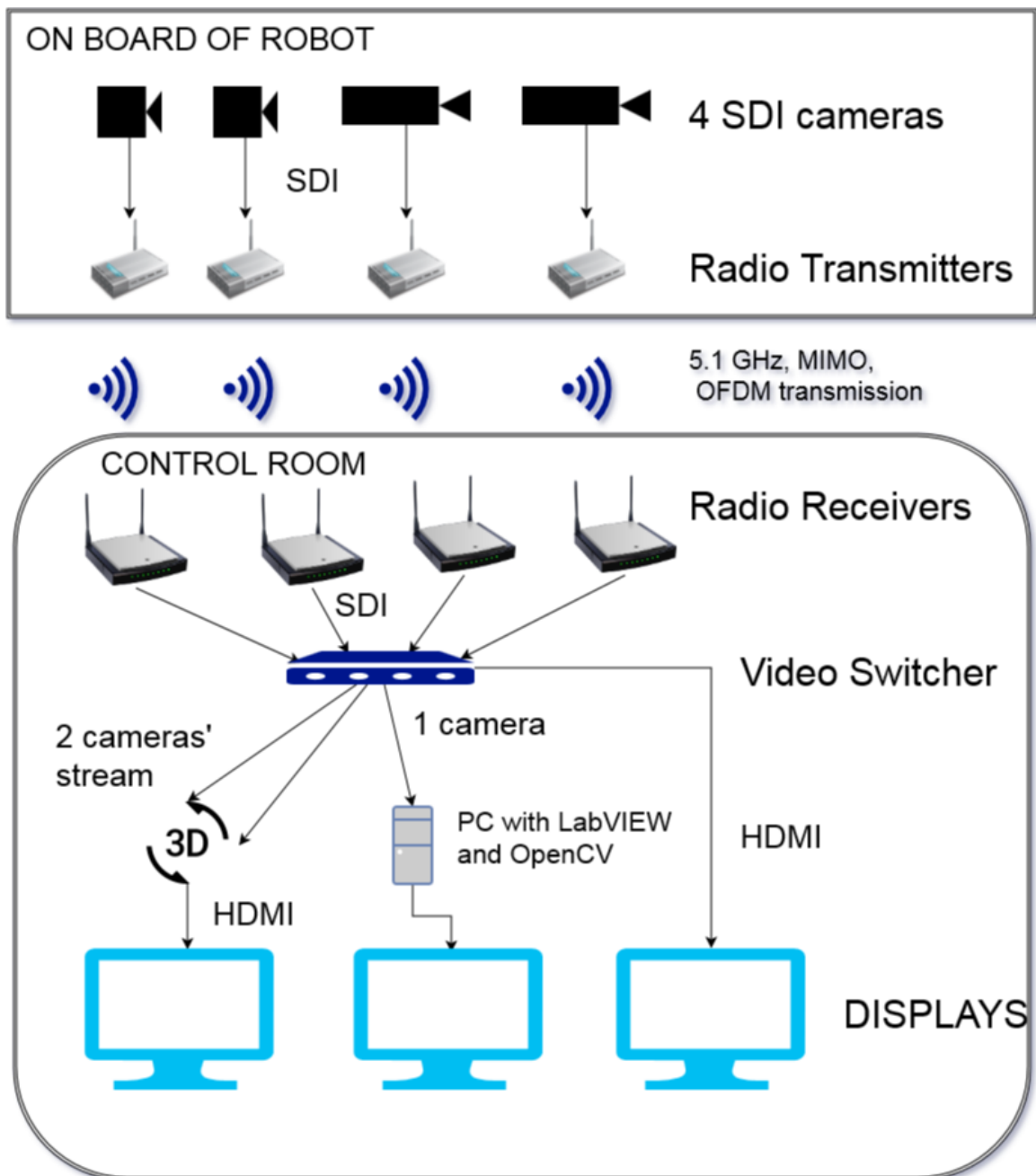


**Figure 3.13.** Image to Array function (Zone NI, 2013c).

In Appendix 1, the algorithm of work is enclosed. The concept is given in a block diagram view which gives clear understanding how elements of the advisory system are work. From the initialization to the video showing, every step is covered.

#### 4 RESULTS AND ANALYSIS

The main result of this thesis is a fully working vision advisory system which includes both hardware and software solutions. The scheme in Figure 4.1 shows the interaction between different parts of the system. In the Appendix 1 it is possible to find the algorithm for working software program of this vision system.



**Figure 4.1.** General scheme of the machine vision system.

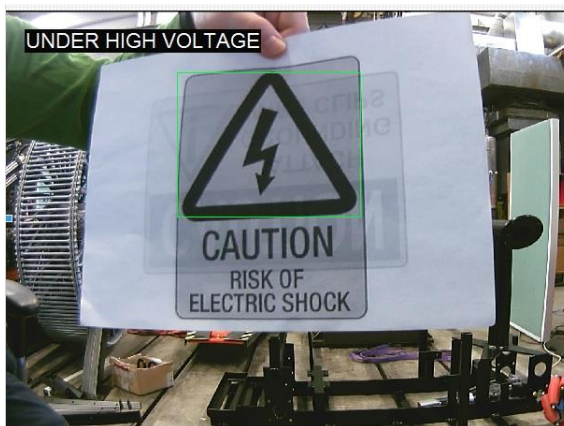
The images presented in Figure 4.2 (a, b, c, d) shows screenshots from display. Three of them is about template matching and another is about detection of people's faces. As one can notice, system gives a piece of advice. It is a string in the left corner of the image.



A



b



C



d

**Figure 4.2.** Results of the pattern recognition.

One more feature, which is implemented is fully working real-time generation of 3D vision with the special device. The platform was designed and manufactured by 3D printing for adjusting cameras for working together. The result of it can be used by operator to gain telepresence experience.

The system reliability is not 100% because the recognition in some situations can be hard. For instance, in laboratory experiments light conditions can be as perfect as possible. Moreover, outside humidity, extreme temperatures and other radio signals can cause difficulties.

## 5 CONCLUSION AND FUTURE WORK

In the modern industry, system created after this thesis can be fully used in automotive industry, hazardous areas, mines and other underground activities. It cannot be denied that there will be some difficulties in retuning system to particular application but there is a room for improvement.

The machine vision for mobile robot was provided as an advisory system for an operator. It is created as a full featured product, which can be used in different purposes. Functions of this system are:

- ✓ Ability to use multiple displays
- ✓ 3D vision capability
- ✓ Latency is less than 1 ms and range is 700 meters;
- ✓ Detection signs, colors, people in the almost real-time frames.

In future, more experiments can be done on site during actual work of the machine. This will be very helpful for optimizing algorithms and coefficients involved in recognition processes. Although, other functionality can be added because system is flexible enough.

## LIST OF REFERENCES

AJA. 2011. [Online Document] Hi5-3D HD-SDI Multiplexer to HDMI 1.4a and SDI Video and Audio Converter. [Referred 08.04.16]. Available: [https://www.aja.com/pdf/support/manuals\\_conv/AJA\\_manual\\_Hi5-3D.pdf](https://www.aja.com/pdf/support/manuals_conv/AJA_manual_Hi5-3D.pdf)

Alor-Hernandez G., Posada-Gomez R., Sandoval-Gonzalez O.O., Sibaja A.M. & Portillo-Rodriguez O. 2011. [Chapter 14:] Digital Image Processing Using LabView. In: Folea, S. (editor). Practical Applications and Solutions Using LabVIEW™ Software. Croatia: InTech. pp. 297–316.

Ayache, N. 1991 Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception. Cambridge: MIT press. 326 p.

B&H Photo Video Digital Cameras, Photography, Camcorders. 2015a. Marshall Electronics CV500-MB2 2MP Miniature Broadcast Camera with HD-SDI. [B&H Photo webpage]. [Referred 08.04.2016]. Available: [http://www.bhphotovideo.com/c/product/1058330-REG/marshall\\_electronics\\_cv500\\_mb2\\_2mp\\_miniature\\_hd\\_sdi.html](http://www.bhphotovideo.com/c/product/1058330-REG/marshall_electronics_cv500_mb2_2mp_miniature_hd_sdi.html)

B&H Photo Video Digital Cameras, Photography, Camcorders. 2015b. Marshall Electronics CV200-MB 2.1MP Full-HD Broadcast HD-SDI Lipstick POV Camera [B&H Photo webpage]. [Referred 08.04.2016]. Available: [http://www.bhphotovideo.com/c/product/1119913-REG/marshall\\_electronics\\_cv200\\_mb\\_2\\_1mp\\_full\\_hd\\_broadcast.html](http://www.bhphotovideo.com/c/product/1119913-REG/marshall_electronics_cv200_mb_2_1mp_full_hd_broadcast.html)

Beer, J.M., Fisk, A.D., Rogers, W.A. 2014. Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. *Journal of Human-Robot Interaction*, 3(2). pp. 74–99.

Biglieri, E. 2005. Coding for wireless channels. New York: Springer Science+Business Media, Inc. 427 p.

Bing, B. 2015. Next-Generation Video Coding and Streaming. John Wiley & Sons. 427 p.

Black Ice Software. 2016a. RGB Color Space - Color Space Conversion. [Black Ice Software webpage] [Referred 08.04.2016]. Available: <http://www.blackice.com/colorspaceRGB.htm>

Black Ice Software. 2016b. HSV Color Space - Color Space Conversion. [Black Ice Software webpage] [Referred 08.04.2016]. Available: <http://www.blackice.com/colorspaceHSV.htm>

Blackmagic. 2015. Software. [Blackmagic webpage]. [Referred 08.04.2016]. Available: <https://www.blackmagicdesign.com/uk/products/atem/software>

Bölcskei, H., 2004. Principles of MIMO-OFDM wireless systems. Chapter in CRC IEEE Wireless Communications, August 2006, pp. 1–33.

Briechle K., Hanebeck U.D. 2001. Template Matching using Fast Normalized Cross Correlation. Proceedings of SPIE – The International Society for Optical Engineering, vol. 4387, pp. 95-102 Available: [http://i81pc23.itec.uni-karlsruhe.de/Publikationen/SPIE01\\_BriechleHanebeck\\_CrossCorr.pdf](http://i81pc23.itec.uni-karlsruhe.de/Publikationen/SPIE01_BriechleHanebeck_CrossCorr.pdf)

Campbell, S. 2011. Active vs. passive 3D: What is better? [PC world webpage]. [Referred at 08.04.2016]. Available: [http://www.pcworld.idg.com.au/article/389806/active\\_vs\\_passive\\_3d\\_what\\_better/](http://www.pcworld.idg.com.au/article/389806/active_vs_passive_3d_what_better/)

Dalal, N., Triggs, B. 2005. Histograms of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA. [Referred at 08.04.2016].

Davies, E.R. 2012. Computer and Machine Vision: Theory, Algorithms, Practicalities. UK: Academic Press. 912 p.

Dhake A. M. 1995. Television and Video Engineering. New Delhi: McGraw-Hill. 190 p.

DSMT. 2015. Resources. IP rating chart. [DSMT webpage]. [Referred 08.04.2016]. Available: <http://www.dsmt.com/resources/ip-rating-chart>

Ebner, M. 2007. Color Constancy. New Jersey: Wiley & Sons. 397 p.

EBU – TECH 3299. 2010. [Online Document] High Definition (HD) Image Formats for Television Production Genève: International Organization for Standardization. p.10 Available in pdf-file: <https://tech.ebu.ch/docs/tech/tech3299.pdf>

Epiphan Video. 2016. Features [Epiphan Video webpage] [Referred 08.04.2016]. Available: <http://www.epiphan.com/products/avio-sdi/>

Fenlon W. 2014. 48 FPS and Beyond: How High Frame Rate Films Affect Perception. [online document]. [Referred 08.04.2016]. Available: <http://www.tested.com/art/movies/452387-48-fps-and-beyond-how-high-frame-rates-affect-perception/>

Fernando, A., Worrall, S. T., Ekmekciođlu, E. 2013. 3DTV: Processing and Transmission of 3D Video Signals. John Wiley & Sons. 224 p.

Fernando S. 2013. Color Detection & Object Tracking [web document]. [Referred 08.04.2016]. Available: <http://opencv-srf.blogspot.fi/2010/09/object-detection-using-color-seperation.html>

Gelsema E.S., Kanal L.N. 2001. Pattern Recognition in Practice II. Elsevier Science. 571 p.

Geronimo D., Lopez, A.M. 2014. Vision-based Pedestrian Protection Systems for Intelligent Vehicles. Springer Verlag. 114 p.

Gurevich G., Nice K., Wilson T.V. 2014. How Digital Cameras Work. [web document]. [Referred 08.04.2016]. Available: <http://electronics.howstuffworks.com/cameras-photography/digital/digital-camera.htm>

Holt, A. & Huang, C.Y., 2010. 802.11 Wireless Networks: Security and Analysis. New York: Springer. 212 p.

Hubble. 2016. A short introduction to astronomical image processing. [Hubble webpage]. [Referred 08.04.2016]. Available: [https://www.spacetelescope.org/projects/fits\\_liberator/improc/](https://www.spacetelescope.org/projects/fits_liberator/improc/)



Hui Liu and Guoqing Li, 2005. OFDM-Based Broadband Wireless Networks Design and Optimization, Hoboken, New Jersey: Wiley & Sons. 251 p.

IEEE 802.11. 2016. OFFICIAL IEEE 802.11 WORKING GROUP PROJECT TIMELINES - 2016-02-02. [IEEE webpage]. [Referred 08.04.2016]. Available: [http://grouper.ieee.org/groups/802/11/Reports/802.11\\_Timelines.htm](http://grouper.ieee.org/groups/802/11/Reports/802.11_Timelines.htm)

ITechnospot. 2015. So what's after 4G-LTE? [ITechnospot webpage]. [Referred 08.04.2016]. Available: <http://www.itechnospot.com/blog/techy-news/so-whats-after-4g-lte/>

Li, Z., Su, Y.X.C. 2015. Intelligent Networked Teleoperation Control, New York: Springer. 282 p.

Litwiller, D., 2001. CCD vs. CMOS. Photonics Spectra, 35(1), 1-4 pp.

Litwiller, D. 2005. CMOS vs. CCD: Maturing Technologies, Maturing Markets. Photonics Spectra, 39(8), 1-5 pp.

Martin, D. 2014. A Practical Guide to Machine Vision Lighting - Part I. 1-4 p. [Referred 08.04.2016]. Available in pdf-file at: <http://www.ni.com/white-paper/6901/en/pdf>

Martin, D. 2015a. A Practical Guide to Machine Vision Lighting - Part 2. 1-5 p. [Referred 08.04.2016] Available in pdf-file at: <http://www.ni.com/white-paper/6902/en/pdf>

Martin, D. 2015b. A Practical Guide to Machine Vision Lighting - Part 2. 5 p. [Referred 08.04.2016] Available in pdf-file at: <http://www.ni.com/white-paper/6903/en/pdf>

McEwen, A., Cassimally H. 2014. Designing the Internet of Things. UK: Wiley. 336 p.

Merriam-Webster. 2016. Pixel [Merriam-Webster webpage]. [Referred 08.04.2016] Available: <http://www.merriam-webster.com/dictionary/pixel>

Microsoft. 2007. What is a DLL? [Microsoft Support webpage]. [Referred 08.04.2016] Available: <https://support.microsoft.com/en-us/kb/815065>

National Instruments India. 2014. Webcast Wednesday # 33 | Image Processing using LabVIEW. Youtube video. [Referred 08.04.2016] Available: <https://www.youtube.com/watch?v=4wzwhAllx5o>

Nava Rodríguez, N.E. 2011. Advanced Mechanics in Robotics Systems, London: Springer. 105 p.

Pender, K. 2012. Digital Colour in Graphic Design. Taylor & Francis. 224 p.

Poynton, C. 1996. A Technical Introduction to Digital Video. John Wiley & Sons. 1-31 pp. [Referred 08.04.2016] Available in pdf-file: <http://www-inst.eecs.berkeley.edu/~cs150/Documents/BasicVideoPrinciples.pdf>

Rysavy, P. 2014. Challenges and considerations in defining spectrum efficiency. Proceedings of the IEEE, 102(3), pp.386–392. [Referred 08.04.2016]

Satinet. 2016. All about HDMI: cables, connectors and adapters. [Satinet webpage]. [Referred 08.04.2016]. Available: <http://satinet.info/vse-o-hdmi-kabeli-razemy-i-perekhodniki/>

Sciammarella, C.A., Sciammarella F. M. 2012. Experimental Mechanics of Solids. John Wiley & Sons. 743 p.

Scopus. 2016a. Analyze Search results. Year. [Scopus webpage]. [Referred 08.04.2016]. Available: <http://www-scopus-com.ezproxy.cc.lut.fi/term/analyzer.uri?sid=409233055B3EF29F617F69F8398535BA.aqHV0EoE4xllF3hgVWgA%3a10&origin=resultslist&src=s&s=TITLE-ABS-KEY%28pattern+recognition+%26+machine+vision%29&sort=plf-f&sdt=b&sot=b&sl=51&count=2401&analyzeResults=Analyze+results&txGid=0>

Scopus. 2016b. Analyze Search results. Subject area. [Scopus webpage]. [Referred 08.04.2016]. Available: <http://www-scopus-com.ezproxy.cc.lut.fi/term/analyzer.uri?sid=409233055B3EF29F617F69F8398535BA.aqHV0EoE4xllF3hgVWgA%3a10&origin=resultslist&src=s&s=TITLE-ABS-KEY%28pattern+recognition+%26+machine+vision%29&sort=plf-f&sdt=b&sot=b&sl=51&count=2401&analyzeResults=Analyze+results&txGid=0>

Scully, T. 2013. 7 Reasons Why HD-SDI is Better than HDMI. [ChurchMag webpage]. [Referred 08.04.2016]. Available: <http://churchm.ag/hdmi-vs-hd-sdi/>

Studytonight. 2016. Transmission Mediums in Computer Networks. [Studytonight webpage]. [Referred 08.04.2016]. Available: <http://www.studytonight.com/computer-networks/transmission-mediums>

SWIT Electronics. 2015a. Home. [online document]. [Referred 08.04.2016]. Available: <http://www.swit.cc/>

SWIT Electronics. 2015b. Wireless system. [online document]. [Referred 08.04.2016]. Available: <http://www.swit.cc/products/show.aspx?id=193>

Teledyne Dalsa. 2014. CCD vs. CMOS. [Teledyne Dalsa webpage]. [Referred 08.04.2016]. Available: <https://www.teledynedalsa.com/imaging/knowledge-center/appnotes/ccd-vs-cmos/>

van Ballegoie, E. 2012. LG 55LM760S TV review: higher mid-range TV. [UK Hardware webpage] [Referred 08.04.2016]. Available: <http://uk.hardware.info/reviews/3280/5/lg-55lm760s-tv-review-higher-mid-range-tv-passive-3d-easy-on-the-eyes>

Visualistan. 2015. How safe is that Wifi? #infographic. [Visualistan webpage]. [Referred 08.04.2016]. Available: <http://www.visualistan.com/2014/04/how-safe-is-that-wifi-infographic.html>

Wikipedia. 2013. Serial Digital Interface. [Wikipedia webpage]. Updated 17.03.2016. [Referred 08.04.2016]. Available: [https://en.wikipedia.org/wiki/Serial\\_digital\\_interface](https://en.wikipedia.org/wiki/Serial_digital_interface)

Wikipedia. 2015. List of video connectors. [Wikipedia webpage]. [Referred 08.04.2016]. Available: [https://en.wikipedia.org/wiki/List\\_of\\_video\\_connectors](https://en.wikipedia.org/wiki/List_of_video_connectors)

Wikipedia. 2016. Pyramid (image processing) [Wikipedia webpage]. [Referred 08.04.2016]. Available: [https://en.wikipedia.org/wiki/Pyramid\\_%28image\\_processing%29](https://en.wikipedia.org/wiki/Pyramid_%28image_processing%29)

Woodford, C. 2007. Webcams. [Explainthatstuff webpage]. Updated May 2014. [Referred 08.04.2016]. Available: <http://www.explainthatstuff.com/webcams.html>

Xiao, S.Q. & Zhou, M.T., 2008. Millimeter Wave Technology in Wireless PAN, LAN, and MAN, CRC Press. 448 p.

Zone NI. 2013a. Pattern Matching Techniques. [National Instruments webpage]. [Referred 08.04.2016]. Available: [http://zone.ni.com/reference/en-XX/help/372916P-01/nivisionconcepts/pattern\\_matching\\_techniques/](http://zone.ni.com/reference/en-XX/help/372916P-01/nivisionconcepts/pattern_matching_techniques/)

Zone NI. 2013b. Match Pattern 4 VI. [National Instruments webpage]. [Referred 08.04.2016]. Available: [http://zone.ni.com/reference/en-XX/help/370281U-01/imaqvision/imaq\\_match\\_pattern\\_4/](http://zone.ni.com/reference/en-XX/help/370281U-01/imaqvision/imaq_match_pattern_4/)

Zone NI. 2013c. IMAQ ImageToArray VI [National Instruments webpage]. [Referred 08.04.2016]. Available: [http://zone.ni.com/reference/en-XX/help/370281P-01/imaqvision/imaq\\_imagetoarray/](http://zone.ni.com/reference/en-XX/help/370281P-01/imaqvision/imaq_imagetoarray/)

Machine vision algorithm.

