

## POPULAR STRATEGIES AND METHODS FOR USING AUGMENTED REALITY

Dawid PIECHACZEK<sup>1\*</sup>, Ireneusz JÓZWIAK<sup>2</sup>

<sup>1</sup> University of Science and Technology, Wrocław; dawid.piechaczek@pwr.edu.pl,  
ORCID: 0000-0001-6670-7568

<sup>2</sup> University of Science and Technology, Wrocław; ireneusz.jozwiak@pwr.edu.pl,  
ORCID: 0000-0002-2160-7077

\* Correspondence author

**Abstract:** Augmented reality (AR) is a modern technology which integrates 3D virtual objects into the real environment in real time. It can be used for many purposes, which should improve different processes in daily life. The paper will analyze the areas in which this technology is currently used. First, the history of the development of augmented reality will be recalled. Then, this technology will be compared to virtual reality because these terms are often incorrectly used interchangeably. This paper describes the tools and popular platform solutions related to augmented reality. The most common problems related to the use of this technology will be discussed, including popular approaches concerning optical and video combining methods. The existing applications and their potential in solving everyday problems will be analyzed. Finally, the perspectives for the development of augmented reality and its possibilities in the future will be discussed. This paper provides a starting point for using and learning about augmented reality for everyone.

**Keywords:** augmented reality, graphical elements, image processing, image recognition.

### 1. History and development of augmented reality

Augmented reality is defined as a system which combines the real world with the virtual one. All processes should work in real time and enable every motion in 3-dimensional space. In brief, augmented reality gives the opportunity to set virtual models in real conditions. For years, the AR existed only in the minds of scientists.

In 1901, L. Frank Baum (Augment, 05.05.2018) claimed that it is possible to display electronic view, which overlays data into the real life. It took over 60 years to put this idea into practice. In 1968, Ivan Sutherland invented the first display system based on a head-mounted device (Augment, 05.05.2018). This stunning system generated graphics and showed simple

drawings for users. Unfortunately, the complexity of the images was limited by computing capabilities of computers.

Next, Myron Krueger (Augment, 05.05.2018) built a laboratory called Videoplace in 1974. In a special room, people were surrounded by an artificial reality and the system responded to their every action and movement. A full mobile device for augmented reality was first built by Steve Mann (Augment, 05.05.2018), who packed an Apple II computer with a small viewfinder into his rucksack. The view from a camera was presented in real time with special contextual information (Mann et al., 2005).

In the early 90's, the term "augmented reality" was mentioned by Tom Caudell (Caudell and Mizell, 1992). From that moment, the AR technology development gradually began to take pace. The phenomenon of miniaturization caused the spread of technology to small, portable devices. In the meantime, new solutions for phones such as a camera and GPS have started to gain popularity. Since 2010, a tendency to create augmented reality apps primarily for smartphones was observed.

Nowadays, the augmented reality is applicable to many devices and the estimated value of the AR business is 1.1 billion dollars (Augment, 05.05.2018). It should be noted that the augmented reality is not a virtual reality. Unfortunately, it happens in literature and the Internet that these concepts are confused with each other. We differentiate the following realities:

- virtual reality – fully replaces the real world. All elements in front of the user's eyes are virtual, most often they are 3D elements created in a graphic program,
- augmented reality – a type of system which applies virtual elements to the real world of the user. The user still has the chance to observe their real surroundings,
- mixed reality combines a virtual and real world to generate new environments where their objects can interact with each other. The term hybrid reality can also be found in literature.

Another known concept is the concept of augmented virtuality, where elements from the real world, in real time, may appear in the virtual world, e.g. user's hands.

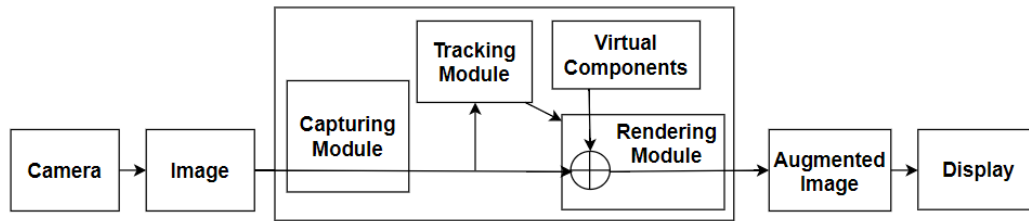
## **2. Hardware, software and algorithms in augmented reality**

The spread of smartphones with high resolution screens had a big impact on the development of augmented reality. These devices and portable laptops allowed for the free use of technology. It is enough to go back to earlier years, where the prototypes which used the AR had to be hung on the ceiling or occupied a large part of the room. Every device supporting the operations of augmented reality must have the necessary hardware parts, such as: a processor, display, sensors and input devices (Mann et al., 2005). The processor is an important part of every device. It ensures fast operation and high efficiency. A good quality

processor prevents the application from lagging and affects the response of the application in real time. The clock speed and the number of cores are important parameters when choosing a processor. The first enables the carrying out of a number of operations per second, depending on the frequency. The second parameter allows one to increase the number of threads on which these operations are performed. In modern solutions, it is no longer troublesome to look for a compromise between the timing and the number of cores. Due to the different ways of rendering views there are many types of displays present on the market.

Nowadays, monitors, optical projectors, handheld devices are used. Moreover, there is a known case of display systems worn on the human body and this is not science fiction. Head-mounted display (HMD) is a really popular type of augmented reality display. It was even used by Ivan Sutherland 40 years ago. This solution is based on a device which is worn on the forehead like a helmet. The device renders images of the real and virtual world and combines them together over the user's field of view. HMD adapts to the movements of the human head. Such devices can use various commercial units for displaying the image: from well-known, commercial solutions to a mobile phone, where the image is supported by the appropriate selection of lenses. Displays for augmented reality may also look like glasses. In some versions, the glasses are equipped with a webcam that records the image and then redisplay it through the eyepieces. Other glasses project the image reflected off the surfaces or through the eyewear's lenspieces. Head-up display (HUD) is a transparent surface with displayed virtual elements (Caudell and Mizell, 1992). This technology is primarily associated with a fighter pilot cabin, where very important information about the position and opponents were often displayed. It was first used in the 1950s and it is considered a precursor of augmented reality.

Nowadays, it looks like a headset with a special screen instead of micro. Sensors are an indispensable element that ensures interaction between a person and the system. The most popular is the use of an accelerometer, which examines the position of the equipment in relation to the body. This sensor is present in many smartphones that are currently available on the market. Other devices that are applied to the head are also equipped with sensors which check the position of the head. Another sensor responsible for determining the position of the device is GPS. It is used when trying to embed augmented reality on a larger surface than a single room. The basic device used in augmented reality technology is the camera. It is most often used to connect the virtual and real world. The image from the camera is usually displayed on the screen and additionally enriched by graphic elements. Moreover, input devices are used to provide more data that flows into the device. Such devices are: a pointer, stylus, wand, glove or other body wear (Siltanen, 2012). These products help to clarify, among others, the position of a specific part of the body or the exact movement. Particularly, they help to recognize gestures made by people. Moreover, devices supporting speech processing are used for the remote usage of devices.



**Figure 1.** Flowchart for simple AR system (Siltanen, 2012).

Augmented reality software must effectively combine all received information and process it. Figure 1 shows a flowchart for a model AR system. In the capturing module, the image from the camera is captured. Instead of the camera, any other input device which records movement or sound can be used. It is important to evaluate the coordinates well, regardless of the image received from the camera. This process is called image registration and it uses many methods related to image tracking techniques. This process occurs most often in the tracking module. Tracking techniques are divided into two parts. The first one is searching for points that are important for image processing and error detection. Then, the flow of images from cameras is examined. At this point, you can use many algorithms that recognize images, including the detection of edges or corners and thresholding. In the second stage, the system of coordinates obtained from the image is used. In some of the images they are already well known, but for some, they should be calculated additionally. For noise that prevents the coordinates from being read, location and mapping (e.g. SLAM) are used. The rendering module has the possibility of combining virtual graphics with real image and displaying it on the screen.

In order to facilitate the programming of applications using augmented reality, SDK packages such as Vuforia and ARToolKit are available (Peddie, 2017). Cloud solutions are also available to speed up the processing. They are often available as ready-made plugins for programming environments which just need to be plugged in and can be used right away. They usually include the entire configuration of the target device on which, for example, goggles or a telephone will be displayed, which significantly simplifies the implementation. The program is most often developed in the Unity software, which is the most popular solution for creating two- and three-dimensional computer games. In addition, you can use them on many platforms, such as Android and iOS. A single screen in the program is called a scene. The application may have many scenes, but each of them will require a separate render. Each element on the stage is called a game object and has a component containing information about the position, details and scripts related to the object. Tags are used to easily control the object. Object groups can be exported into pre-made packages, which are called prefab. They are ready for late use and most manufacturers provide such ready packages for programmers. An important element of the software is the camera, which at the moment shows the selected element or section of the scene. The camera position can be varied, depending on the movement, position, etc. The dynamics and behavior of objects are conditioned by scripts. In the case of Unity, they are written in C# language and ensure the universality of the platform

and compatibility with the majority of devices available on the market. In the augmented reality, it is very important to determine what the user is looking at and where they are.

In practice, the device often has to adjust to the environment and calibrate the image. Tracking enables constant calculation of the relative position and allows you to process data in real time. This is one of the most required behaviors in the world of virtual reality. Researchers are constantly competing in finding better and better methods for image processing and evaluation. That is why vision tracking systems are based on tracking the environment and the behaviors occurring in them. This results in a lot of data to be processed. An interesting algorithm used in the detection of hand and head movements is the tracking algorithm. It is based on recognizing motion or detecting skin color in an image. CAMSHIFT is now a popular choice when it comes to this type of algorithm (Bradski, 1998). It supports the tracking of movements in real time and is based on color probability distributions. The algorithm is relatively simple to use, the computational complexity is efficient and it can recognize images on complex scenes.

However, the algorithms used to process gestures and movements are not always necessary. Sometimes a device may have problems with recognition in an unknown environment and this is challenging. Therefore, there are a lot of techniques which are often used to limit the increase in environmental data and help to easily estimate the coordinates. One of the solutions that allow to overcome the described difficulties is to use special, predefined symbols that will be easily recognized by the system (Siltanen, 2012). The symbol usually has special characters or signs that are easily captured using pattern recognition or image processing methods. Once detected, it allows to adjust the position and image configuration in the camera accordingly. This method of image detection in the environment is called marker-based tracking and is commonly used with augmented reality systems. In addition, feature-based and model-based solutions are popular, too. In the first approach, the system recognizes the optical features of the image and learns by observing the movements between the frames. In the second approach, the system has a previously created model of the scene or its parts and compares it with the current image, looking for the best matches. However, this article will focus on the marker-based method. One of the most important features of a good marker is the ability to easily detect. Therefore, it should be detectable in different circumstances. Differences in contrast are much easier to detect than differences in color using vision techniques (Hartley and Zisserman, 2000). Therefore, the use of black and white markers is optimal and fully justified. The marker-based system should easily evaluate the coordinates. Any four points should be sufficient to assess the position of the camera. Ideally, they should form a simple figure, for example a square. Additionally, the figures located in the corners should be relatively strong because on their basis it is possible to calculate the intersections.

In the detection process, the intensity of the image is assessed at the outset. If the image format is different, then conversion occurs. What is important, the marker detection should also take place on an image in shades of gray. Next, the marker borders are detected and an

evaluation of how the view will be taken is carried out. For this, different, very well-known techniques are used, such as thresholding or line fitting. Edge detection is a laborious process, which is why most often only sampling is performed during preprocessing and the intersection of edges at the corners of the square is detected (Hirzer, 2008). Small mistakes when detecting a marker position in the 2D position can significantly affect the calculation of coordinates, which is why this stage is very important. There are many reasons for such errors; among them are noise, bad threshold or pixel quantization error. The problem may also be the detected large number of squares in the image, so a lot of memory is needed to process the image, but this has also already been solved.

Methods of accepting or rejecting a given case were invented. They are designed to cover the number of squares found in the field that are not markers. The calculation of the number of pixels in the figure circuit is a fast and efficient process that can be carried out during labeling. On the other hand, calculating the size and distance of an element is longer and more labor-intensive. Therefore, it is worth using the method of the first pixel located on the edges to quickly calculate the area and diagonals of the figures. For example, in the ARToolKit SDK, it is assumed that markers will be detected from a suitable distance (Peddie, 2017). Therefore, all possible markers further away from the calculated one will not be considered. This simplifies the calculation and narrows the number of squares found in the image. In addition, it distinguishes the size of markers that are within the allowed limits of the standard. Then, for further analysis, it allows only those that have met all the criteria related to sizes. Depending on the type of marker, the program can recognize the whole content immediately or analyze it in depth. The simplest markers have a black image in the middle of a white background, which is surrounded by a black field. The system identifies the marker by comparing the divided image segments with the marker templates (Siltanen, 2012). When comparing, the application checks the marker with each template. When the best fit is assessed, the identity of the marker is determined. An example of markers that use templates is the picture on the left side of Figure 2.



**Figure 2.** Examples of markers used for image detection. One white hole is visible on the left marker. On the right there is a binary maker, there are many white holes and strong edges that are used to quickly calculate the acceptance criteria.

Another example of the markers used are called 2D barcodes (Figure 2). They are a grid of black and white cells containing data. During detection of the image, the values of the means of these cells are calculated (Siltanen and Hyväkkä, 2006). There are two types of barcodes: pictures containing only the id, so called ID markers, and data markers that have more data. There are many standards of 2D barcodes. Currently, the most popular are QR Code, Data

Matrix and PDF417. Their initial and main purpose is to use for tagging and logistics, but they are increasingly used in the field of augmented reality. There are also markers in the shape of wheels. They are called circular markers and are shown in a perspective view. The position of the center of the ellipse is easier to capture from the position of the center of the square. There is a heuristic of unambiguous rejection of non-markers that checks the number of intensity changes in two directions that are perpendicular (Siltanen and Hyv akk a, 2006). Another important step in using markers is to assess their pose. Three orientations ( $x, y, z$ ) and three angles ( $\alpha, \beta, \gamma$ ) are used to describe the position of the marker. Good quality equipment is useful when reading the item's position. What matters is the use of devices with image orientation parameters, a good focal length and a size sensor. Coordinates can detect common points that will be responsible for the marker's corners. The transformations that occur between the lens and the marker can be described by the formula (1):

$$X_C = TX_i \quad (1)$$

where:

$X_C$  is responsible for the points of coordinates,

T is the transformation matrix,

$X_i$  shows the coordinates of the image that is captured.

The T-matrix is usually presented as a 3x3 rotation matrix, and additionally, it has the coordinates of the transformation –  $t$ :

$$X_C = [R|t]X_i \quad (2)$$

then we get a homogeneous coordinate (3):

$$\begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \quad (3)$$

Most often, it can be assumed that cameras are currently devoid of problems with distortion of images, so the transformation matrix can be multiplied by the calibration matrix K. Simple mathematic transformations result in a matrix that presents parameters that represent the detected points. The perspective projection matrix, which should be multiplied by the parameters achieved, is helpful in their assessment. This method is called an iterative algorithm and is used to accurately assess the position of markers in an image. It is possible to configure many markers if you want to achieve a larger number of additional elements on the screen. As always, there are problems in the processing of markers' coordinates. The translation of coordinates  $x$  and  $y$  is much easier than translation  $z$ . The source of this problem is the camera system, which during changes in the plane with a much smaller movement than when doing it in the remaining planes. Based on the methods described here which use markers to obtain a viral image, it can actually be understood as the basics of augmented reality that are currently available to everyone. More complex issues are combined with tracking without prior knowledge. The SLAM approach is then used (Strasdat et al., 2012).

The algorithm is based on Bayes' filters, which are designed to help in drawing inferences about the current state and the future state according to the past. State estimation is always fraught with uncertainty. In addition, it is assumed that camera movement is non-linear in the filter-based method. Therefore, appropriate time intervals are used which are inaccurate when used. Another approach is to use key-frames that support the building of a 3D model. In the initial view of the scene, which is treated as a map, subsequent frames are merged into the corresponding model. In order to reduce the imperfection of the method, methods for monitoring density, e.g. DTAM or LSD-SLAM, are introduced. The first fix supports the processing of a large number of pixels in real time, but this is very computationally demanding.

In turn, LSD-SLAM supports the creation of density maps. Facilitated transformation (positional transformation and drift scale) of key frames is a saving against the usual rigid estimation. LSD-SLAM is the basis for the theory of the use of AR on mobile phones. In summary, it seems that the SLAM method is slightly less efficient and popular than the methods associated with key frames (Strasdat et al., 2012). Its possibilities related to creating models and tracking speed seem to be more limited. The approaches to creating virtual models in the real world presented in this article can be divided into two phases: recognition using markers and in the context of previous knowledge. In addition, the modeling phase, which is only partially described, is very often used. From the phases selected, it seems that marker-based identification is not computationally demanding and it is effective enough in most cases. The model phase uses the knowledge of the previously known model and tries to match it to the actual image. In turn, the possibility of eliminating image matching to the model without previous knowledge seems to be very promising but at the same time very difficult to calculate for ordinary devices available to customers.

### **3. Areas of life using augmented reality**

Augmented reality is still treated on the market as a technological novelty. However, more and more often it affects many areas of life. Frequently, augmented reality is used in the context of geodetic applications (Lee, 2012). Then it can serve as an innovative tool for quick orientation in the field. One may not be aware but augmented reality is often used in various types of navigation. It often uses an image of the road, sometimes a real image, on which it places lines that help reach a specific destination. Augmented reality is also used as a rich source of information about a given place that is visited. The big advantage of this solution is the mobility of the device, which can provide us with information. The issues described are also associated with tourism to a similar extent, where you can often find ready-made objects on the map that are possible to visit, even from the inside. The global success of augmented reality in the field of architecture and design is also well-known. Users of such applications can now



match furniture to their interiors and arrange the interior properly before purchase. It also helps in creating the concept of buildings and providing their visualization to important customers. There were also applications that were intended to show destroyed buildings after an earthquake. A great market for applications using augmented reality is commerce. There are a lot of applications being created, like the one for cosmetic companies to create make-up and for artists who want to sell their works. Users could then use the app to check how the particular work of art would look on the wall in their home. Augmented reality also proves to be helpful when users want to see the contents of the parcel they want to purchase and breaking the seal could result in the loss of the possibility of complaint. Then, using the appropriate application, users can compare the product with their requirements. There is also a case where the user did not have to see the product, but could taste it or smell it. Then, before buying a food product or perfume, the right substances gave the user the sensations they could have after receiving the product. One of the most important reasons for using augmented reality is the opportunity to learn through it.

A user who can touch anything without sacrificing additional funds to buy it can learn and gain a lot. Therefore, applications for learning astronomy, chemistry and medicine are very much appreciated by students (Lee, 2012). A very curious example is an application that contains information on all organs and human bones. Thanks to it, a medical student can get to know all important names and functions that the specific organ has in the body. More and more applications are created that affect different senses, e.g. hearing. It is common for music to be mixed or played with hardware that uses augmented reality. In addition, you can learn to play instruments or create music yourself on such devices that generate the same sounds as the devices in the real world. The last area of life that will be discussed in the paper and in which the augmented reality was originally tested and used is entertainment. Over the years, people have been spending more and more money on having fun. That is why a lot of games have recently been created that aim to bring the feelings associated with playing to a higher level. A particularly spectacular success was achieved by the game about creatures that inhabited the entire planet and the collector user had to find them and then catch them. The game has gained millions of fans and is a very good advertisement for the growing popularity of augmented reality.

#### **4. The future of augmented reality and conclusions**

The article presents the theory, examples of use and benefits of using augmented reality. It is certainly a very good tool for presenting data to the user. The perspectives for the development of augmented reality and its possibilities in the future were discussed. The pace of development and the money that is invested in this business make it an increasingly desirable

product on the market. It seems that nowadays the quality of sensors used in augmented reality can be improved, especially in terms of precision and price. In addition, it seems that, over time, solutions which use this technology will become more and more common and cheaper. Development in the field of programming may result in an even greater impact of the extended reality on everyday life than today. Certainly, the business associated with entertainment is still waiting for the development of this sector, which increasingly uses solutions from the borderline of augmented reality and virtual reality. It seems that if the devices' batteries will be able to last long enough and the devices are gradually integrated with solutions that enable quick calculations, which are based on solutions in cloud computing, mobile devices using augmented reality have a great chance for dynamic development. There are still many ideas that can benefit from augmented reality and help people in their daily lives. They are waiting for implementation and it is only a matter of time when it will happen.

## References

1. Bradski, G.R. (1998). Computer Vision Face Tracking for Use a Perceptual User Interface. *Intel Technology Journal*.
2. Caudell, T.P., Mizell, D.W. (1992). *Augmented Reality: an Application of Heads-up Display Technology to Manual Manufacturing Processes*. Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, vol. 2, p. 659.
3. Hartley, R.I., Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
4. Hirzer, M. (2008). *Marker detection for augmented reality applications*. Graz University of Technology.
5. Lee, K. (2012). Augmented Reality in Education and Training. *TechTrends: Linking Research and Practice to Improve Learning*, 56, 2, 13.
6. Mann, S., Fung, J., Aimone, Ch. (2005). *Designing EyeTap Digital Eyeglasses for Continuous Lifelong Capture and Sharing of Personal Experiences*. Alt.Chi, Proc. CHI.
7. Peddie, J. (2017). *Software Tools and Technologies*. Augmented Reality. Heidelberg: Springer.
8. Siltanen, S. (2012). *Theory and applications of marker-based augmented reality*. Kuopio: VTT.
9. Siltanen, S., Hyvääkä, J. (2006). Implementing a natural user interface for camera phones using visual tags. *7th Australasian User Interface Conference*, 50, 113. ACS.
10. Strasdat, H., Montiel, J.M., Davison, A.J. (2012). Visual SLAM: Why Filter? *Image And Vision Computing*, 30, 65-77.
11. *Augmented Reality Blog Homepage*, <http://www.augment.com/blog>, 05.05.2018.