

# An Approach to Marker Detection in IR- and RGB-images for an Augmented Reality Marker

Aaronkumar Ehambram, Patrick Hemme and Bernardo Wagner

<sup>1</sup>*Institute of Systems Engineering - Real Time Systems Group, Leibniz Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany*

**Keywords:** Computer Vision, Augmented Reality, Kinect, Intel RealSense, Pose Estimation, Retroreflective Markers, Sensor Fusion.

**Abstract:** We introduce an augmented reality marker based on ArUco markers (Garrido-Jurado et al., 2014) that can be detected in RGB- and IR-images by using retroreflective material. Due to active perception by IR-capable camera systems the negative impact of external disturbances like change of light conditions on the marker detection is minimized. By the parallel processing architecture of RGB- and IR-images redundancy stabilizes the detection. As different retroreflective materials influence the image quality depending on the camera system, we also examined different retroreflective materials and compared the performance of the Kinect V2 and the Intel RealSense D435 regarding the detection probability depending on the geometrical distance of the augmented reality marker to the camera.

## 1 INTRODUCTION

Localization, pose estimation and mapping are some of the most important tasks for autonomous systems. Applications like autonomous robots (Sim and Little, 2006) and unmanned vehicles (Broggi and Dickmanns, 2000) are highly dependent on the reliability of pose estimation tasks that can be solved effectively by using cameras. Pose estimation algorithms that use natural features such as keypoints and texture often need significant computational requirements and are subject to filter stability issues. Using artificial landmarks reduces to a certain degree the need of computational requirements due to specialized processing steps that lead to an optimized extraction of the points of interest. Additionally, the usage of artificial landmarks enables us as users also to adapt the markers to our needs. An improvement in pose estimation based on artificial landmarks requires a very stable and precise recognition of the markers. One way of gaining stability and precision is redundancy. Many robotic systems as described in (Fankhauser et al., 2015) are equipped with cameras that provide in addition to monocular and stereo vision depth-images and IR-images. Widely spread camera systems like the Kinect (Zhang, 2012) and the Intel RealSense (Draeos et al., 2015) use such technologies making a huge bundle of sensor data available. The usage

of more than one sensor as a source of information for recognition will lead to a more stable detection. In this paper we suggest a new physical topology of augmented reality markers based on ArUco markers (Garrido-Jurado et al., 2014) that enable a detection in the RGB- and IR-image by using retroreflective materials. Therefore, we examined two different retroreflective materials and evaluated their impact on the image data provided by two different cameras - the Kinect V2 and the Intel RealSense D435. In our evaluations we concentrate on the detection quality depending on the distance of the markers to the camera. The main contribution of this paper is the proposal of an optical fiducial augmented reality marker detection based on RGB- and IR-image data. In the following we first discuss related works in Section 2, then we present our suggested physical marker topology based on the ArUco markers and shortly picture a parallel architecture enabling the detection in RGB- and IR-image in Section 3. In Section 4 we describe our evaluation system and discuss the experimental results and end with conclusions in Section 5.

## 2 RELATED WORK

In the literature several fiducial marker systems have been proposed. Some of these are discussed in the

following in order to determine the goals of this paper. Some approaches use planar circular markers that encode the identification in circular sectors or in concentric rings (Naimark and Foxlin, 2002). Due to their primitive construction such markers usually provide only one correspondence point located at the center. That is why more than one marker is needed for pose estimation. Motion Capture is a marker system using retroreflective materials. The markers are often spheres (Rossner, 2014) with a retroreflective surface that can easily be used for segmentation in the IR-image as described in (Dorfmueller and Wirth, 1998). The usage of such markers is very common in Motion Capture because of the stability of the detection. As special IR-cameras are used, the marker system is optimally improved for the detection algorithms. However the identification of the markers involves a complex detection process obtained from the relative marker positions. A very robust alternative to the previously discussed approaches is a square-based planar marker system. One of the most popular square-based fiducial marker systems is ARToolKit (Kato and Billingham, 1999). ARToolKit markers are similarly composed to ArUco markers by a wide black border with an inner image which is stored in a database of valid patterns. Although ARToolKit was extensively used the template matching approach for marker identification leads to high false positive and inter-marker confusion rates. Also the high sensitivity to changing light conditions due to fixed global threshold to detect squares negatively influences the detection algorithm. The ARTag system (Fiala, 2010) is based on the same principal of square-based fiducial systems using binary codes for identification. The improvement of the ARTag system is established by an edge-based square detection method that is more robust to lighting and partial occlusion and by a binary coding scheme that includes checksum bits for error detection and correction. Unfortunately the proposed marker dictionary is fixed to 36 bits and the maximum number of erroneous bits that can be corrected is two. The ArUco library (Garrido-Jurado et al., 2014) provides a stable marker detection based on the ARToolKit and an algorithm for generating configurable marker dictionaries following a criterion to maximize the inter-marker distance and the number of bit transitions. The detection is based on contour extraction and polygone detection. The ArUco detection algorithm is used in our implementation due to its accurate performance.

None of the marker systems that use planar markers make use of special material for detection improvement. As mentioned, on the one hand Motion Capture systems show a very reliable detection scheme of

the markers due to the use of retroreflective material and IR-cameras. But on the other hand square-based fiducial planar markers enable very robust and compact artificial landmark detection methods in RGB-images. That is why we fuse both marker topologies to one novel marker topology based on ArUco markers that are able to be detected in RGB- and IR-image data.

### 3 RGB- AND IR-IMAGE BASED MARKER DETECTION

The marker detection proposed in this paper is based on the ArUco marker detection presented in (Munoz-Salinas, 2012) and (Garrido-Jurado et al., 2014). Our proposed marker topology and the parallel processing architecture are described in the following.

#### 3.1 Enabling IR-image Marker Detection

In order to use the ArUco detection algorithm with IR-image data, it is necessary to change the physical topology of the marker. While in the RGB-image the colours and contrast of the used material and the light conditions have the main impact on the image quality, IR-images captured by active IR-cameras are highly influenced by the reflection characteristics of the material. The higher the reflection rate, the brighter the corresponding spots appear in the image due to high pixel values. Consequently, the usage of retroreflective material for the white bits on the marker leads to the desired characteristics. As the black bits of the markers are equipped with diffuse reflecting and the white bits are coated with retroreflective material, the necessary contrast between the white and black areas in the IR-image is maintained leading to high gradients along the contours that have to be extracted. It stabilizes the detection, since the whole detection algorithm is based on the optimized gradient image.

#### 3.2 Parallel Processing of RGB- and IR-image

As the marker detection with retroreflective material works with RGB- and IR-image data, it is expedient to use both data in order to stabilize the detection through redundancy. In Figure 1 the ROS (Quigley et al., 2009) based architecture of our fused detection is depicted which we describe in the following in detail. In Figure 1 the involved nodes that represent the processes that run simultaneously are marked with

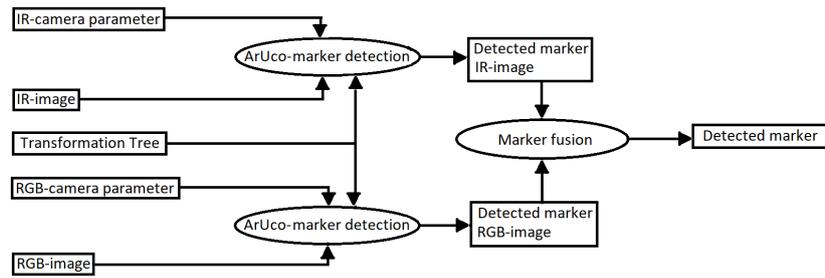


Figure 1: Fused marker detection from the RGB- and IR-image.

circular boundaries. The corresponding topics are marked in rectangular boundaries. In total three nodes are necessary for parallel processing of the RGB- and IR-image: Two nodes process the ArUco-marker detection. The one applies the detection algorithm on the RGB-image and the other on the IR-image. For the detection algorithm camera parameters, transformations and the image data have to be provided. The output of the detection nodes are the detected marker positions and their identification numbers. At this point four different constellations are possible:

1. In both images no markers were detected.
2. In both images markers were detected.
3. Only in the RGB-image markers were detected.
4. Only in the IR-image markers were detected.

Regarding the first case, no further processing steps are necessary. In the second case two marker positions for one corresponding marker in the scene are published. Ideally, both published poses are equal. But due to sensor inaccuracy and noise both poses are not always equal. In order to calculate the correct pose of the marker a further node is used, that fuses the different marker poses. As RGB- and IR-cameras have a similar detection accuracy when retroreflective markers are used (see Section 4), neither the RGB- nor the IR-camera based detection should be favored. Consequently, the correct pose of the marker is calculated by averaging the published pose data of the corresponding markers. Accordingly if the RGB-image based pose is described by

$$\mathbf{P}_{\text{RGB}} = (x_{\text{RGB}} \ y_{\text{RGB}} \ z_{\text{RGB}} \ \alpha_{\text{RGB}} \ \beta_{\text{RGB}} \ \gamma_{\text{RGB}})^T \quad (1)$$

and the IR-image based pose is described by

$$\mathbf{P}_{\text{IR}} = (x_{\text{IR}} \ y_{\text{IR}} \ z_{\text{IR}} \ \alpha_{\text{IR}} \ \beta_{\text{IR}} \ \gamma_{\text{IR}})^T \quad (2)$$

the published pose data is

$$\bar{\mathbf{p}} = \frac{\mathbf{P}_{\text{RGB}} + \mathbf{P}_{\text{IR}}}{2}. \quad (3)$$

The calculated pose is then published to the corresponding topic. The third and the fourth case describe the situations in which the marker fusion node

subscribes to the detected marker topics but no correspondences between both data can be determined. Because of the high true-positive accuracy of the ArUco marker detection it is expedient to publish the detected pose of the corresponding marker to the detected marker topic.

## 4 EVALUATION

For evaluating the performance of the detection of the marker presented in Section 3 we measured the detection probability depending on the distance of the marker. For sake of comparability and evaluation of the degree of detection performance improvement we also printed the ArUco marker on common white printing paper. All markers have a border size of 15 cm. In order to perform a reproducible measurement of the detection probability we built a measuring system which is pictured in Figure 2. The marker is fixed on a frame that is mounted on a linear rail system. The angle of perspective in respect to the camera system is variable. The camera angle in respect to the normal vector of the floor was measured to 30°. In order to maintain comparability of the different cameras, identical environmental conditions are necessary. Therefore, both cameras are mounted on the measuring system and the camera data is recorded in parallel. In order to record the distance of the marker the frame is equipped with two poles, that can be detected by the laser scanner mounted on the measuring

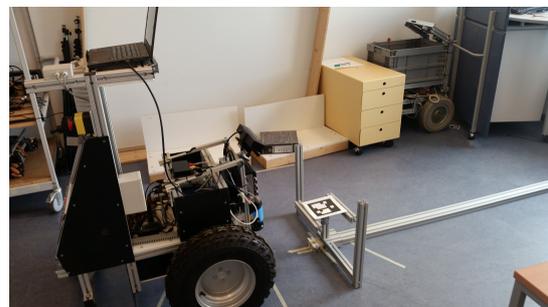


Figure 2: Setup for measuring the detection probability.

system with the cameras. The detection probability is approximated by a sampling method. Therefore, the length of 2.5 m of the linear rail system is quantized into 10 cm steps. Each step represents a bin. For generating the samples both cameras are run with 30 fps while the frame of the rail system, on which the markers are mounted, is slowly moved back and forth varying the distance between marker and cameras. As the distance is tracked by the laser scan data during the whole experiment the amount of captured frames  $F_{captured,i}$ , the amount of frames in which the markers are detected  $F_{detected,i}$  and the corresponding distance  $d_i$  can be assigned to each bin  $B_i$ . Consequently,  $B_i$  can be described by

$$B_i = \{F_{captured,i}, F_{detected,i}, d_i\}. \quad (4)$$

The detection probability  $p_i(B_i)$  for each bin  $B_i$  can be approximated by dividing the amount of the frames in which the markers are detected by the number of captured frames according to equation 5

$$p_i(B_i) = \frac{F_{detected,i}}{F_{captured,i}}. \quad (5)$$

Three different parameters have strong influence on the detection probability measurement depending on the distance:

1. The markers tilt angle changes the reflection behaviour of the materials.
2. The different retroreflective materials lead to different detection performances.
3. The cameras show different behaviours and provide very different images.

For investigating the influence of the tilt angle three different angles were chosen for the experiments that are explained in Subsection 4.1 in more detail. We used two different retroreflective materials for the experiments that have different reflection characteristics. For better comparability paper printed ArUco-markers are tested additionally. For reader convenience we want to introduce the terms *Retro\_1* and *Retro\_2* for the different retroreflective materials. As such retroreflective materials are used in road construction the reflectance characteristics are determined by technical regulations. The following mentioned technical reflectance data are based on the german regulations "Richtlinien fuer die Sicherung von Arbeitsstellen an Strassen" - in short RSA (Schoenborn and Schulte, 1995). *Retro\_1* is a microprisms based material that leads to higher reflectance values than foils with glass bead construction, which results in better viewing of microprismatic signs in the dark. The *Retro\_1* material is a foil that is constructed with a 3M pattern and C construction type. The reflectance

behaviour can be assigned to the reflectance class RA1 according to (Schoenborn and Schulte, 1995). *Retro\_2* is a material with encapsulated catadioptric acting glass beads, which are partially embedded in a plastic layer. The *Retro\_2* material is a foil that is constructed with a 3M pattern and B construction type. The reflectance behaviour can be assigned to the reflectance class RA2 according to (Schoenborn and Schulte, 1995). In our experimental investigations we used the Kinect V2 camera and the Intel RealSense D435. While the Intel RealSense D435 shows economical benefits the used infrared projector uses structured-light that negatively influences the marker detection in the IR-image. Due to the time-of-flight technology of the Kinect and the homogenous infrared projection the IR-image data provided by the Kinect leads to an improved detection behaviour. But the time-of-flight technology enables only the perception of infrared light that has the particular frequency modulation that matches the emitter frequency. Consequently, the Kinect does only percept the infrared light that it projects with its emitter into the scene. In contrast the Intel RealSense D435 also perceives the surrounding infrared light enabling the usage of external light projectors. Although the Intel RealSense D435 is more flexible external disturbances by IR-emitters may lead to unwanted noise. In the following the experimental results will be presented in detail.

#### 4.1 Markers Tilt Angle

In order to measure the influence of the tilt angle the different markers are mounted on the frame that is shown in Figure 2. For each marker both cameras are used for detecting the markers. Since the angle of perspective of the cameras in respect to the normal vector of the floor was measured to 30° the markers tilt angle is set to 0°, 15° and 30° in respect to the tangent vector of the floor level. Due to very similar results of the 18 experiments (3 markers times 2 cameras times 3 tilt angles) only one parameter constellation will be discussed in the following. The diagram portrayed in Figure 3 shows the results of the tilt angle experiments for the Kinect and the *Retro\_1* marker. The abscissa measures the horizontal distance of the marker to the Kinect camera and the ordinate shows the detection probability determined according to the method described in Section 4. As the IR- and RGB-images are captured synchronously the frame counter is increased if both data are received. A frame is marked as a detected frame if either in the IR- or in the RGB-image markers are found as described in Section 3.1. The results of three different experiments are integrated into the plot 3. The red graph corresponds

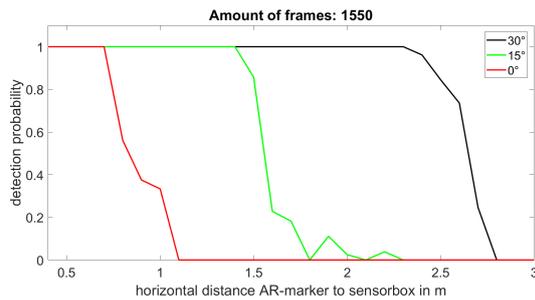


Figure 3: Detection probability depending on the tilt angle of marker *Retro\_1* for the Kinect.

to the detection probability for a tilt angle of  $0^\circ$ , the green graph to  $15^\circ$  and the black graph to  $30^\circ$ . According to the plot 3 the marker detection shows the best performance with a tilt angle of  $30^\circ$ . While the detection with a tilt angle of  $30^\circ$  has 100 % accuracy till a distance of  $2.4\text{ m}$  the 100 % detection probability for  $15^\circ$  is beneath  $1.5\text{ m}$  and for  $0^\circ$  beneath  $0.75\text{ m}$ . As for all experiments the tilt angle of  $30^\circ$  shows the best performance we fix the parameter to  $30^\circ$  for further discussion.

## 4.2 Retroreflective Materials and Cameras

For determining the best camera-material-constallation providing the best performance in respect to the marker detection we want to present the experimental results to the investigations to the different marker materials and cameras. Therefore, two different environmental conditions are considered in detail. While diffused daylight 4.2.1 disturbs the IR-camera of the Intel RealSense D435 as described in Section 4 and reflection negatively influences the RGB-image data the experiments in an underexposed room in Section 4.2.2 stress the benefits of using IR-camera data and show the behavior of the different retroreflective materials in respect to the cameras.

### 4.2.1 Diffused Daylight

The experimental results shown in Table 1 and 2 were recorded during diffused daylight conditions. While the Table 1 shows the detection probability of the different camera-material-constellations, Table 2 shows snapshots of the markers captured by the respective cameras. Each cell of the chart 1 is made of similar diagrams: The abscissa measures the horizontal distance of the marker to the respective camera and the ordinate shows the detection probability determined according to the method described in Section 4. The green graph describes the detection probability of the markers based on the RGB-image, the black graph

refers to the IR-image and the red graph shows the detection probability when both frames are parallel considered as mentioned in Section 3.1. The RGB-image based marker detection with the Kinect is according to Table 1 independent of the material. This fact is also depicted in Table 2. As all RGB-images have a very similar appearance the edge based detection algorithm performs very similar. In contrast, the IR-image based marker detection performance strongly depends on the marker material: While the ArUco-marker printed on paper is detected very accurately till a distance of  $1.5\text{ m}$ , the *Retro\_1* marker is detected till a distance of nearly  $2.5\text{ m}$  though the detection is not as accurate as with paper due to the lower detection probability. The *Retro\_2* marker has the worst performance regarding IR-image based detection. The very inaccurate detection is caused by the reflection characteristics of the encapsulated catadioptric acting glass beads. As pictured in Table 2 black spots appear in the regions where the retroreflective material is located and due to the inhomogeneous bright regions the detection algorithm fails. Concerning the results of the Kinect camera in respect to the merged marker detection it is obvious that using the *Retro\_1* marker shows the most benefits due to the very accurate RGB-image detection and the capability to detect the marker in far distances based on the IR-image.

The Intel RealSense has similar detection results regarding the RGB-images. But in Table 2 we can see that the detection probability based on RGB-images is clipped to 0 % below a distance of  $0.5\text{ m}$ . This behaviour is caused by the experimental setup due to the differing position of the Intel RealSense in respect to the Kinect. Besides the neglectable impact caused by the setup the RGB-image based detection is similarly accurate as with the Kinect. But the IR-image based detection shows its best performance with paper. Both retroreflective materials lead to poor results with maximum distances of  $1.2\text{ m}$ . The reason for the described detection behaviour is shown in Table 2: Due to the structured-light technology of the Intel RealSense surrounding infrared light is also perceived by the camera. Consequently, the diffuse reflecting paper appears in the IR-image accurately, leading to good detection. But both retroreflective materials are over illuminated, so that the inner structure can not be extracted and the detection fails. Conclusively, the Intel RealSense in combination with paper printed ArUco markers has the best performance under the diffused daylight conditions.

### 4.2.2 Underexposed Room

The experimental results of the detection probability in a low illuminated room are described in the Tables

Table 1: Detection probability measurement results with diffused daylight.

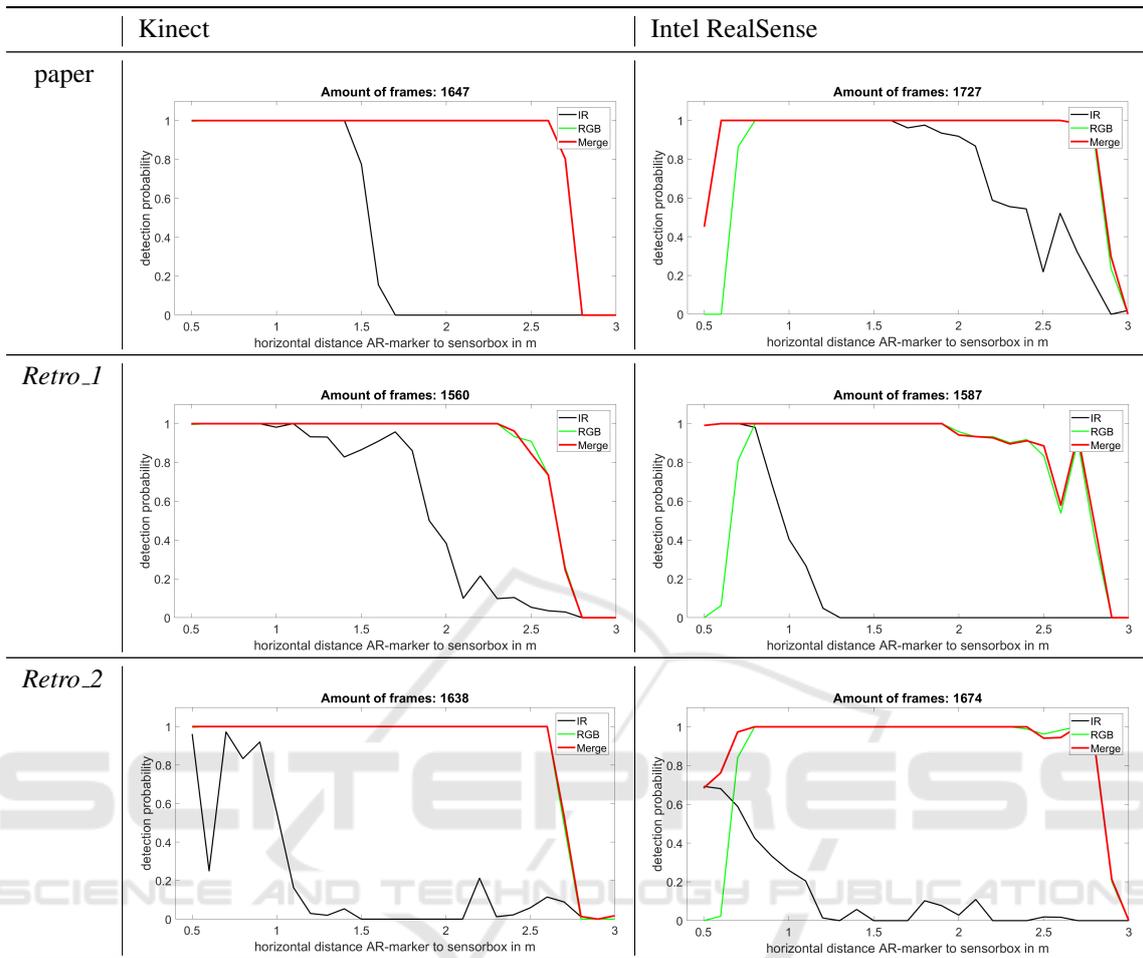


Table 2: Detection probability measurement pictures with diffused daylight.

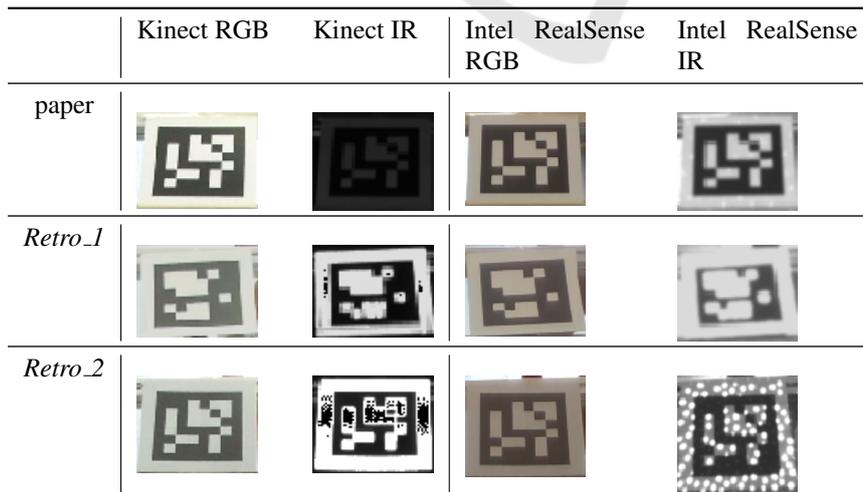


Table 3: Detection probability measurement results in an underexposed room.

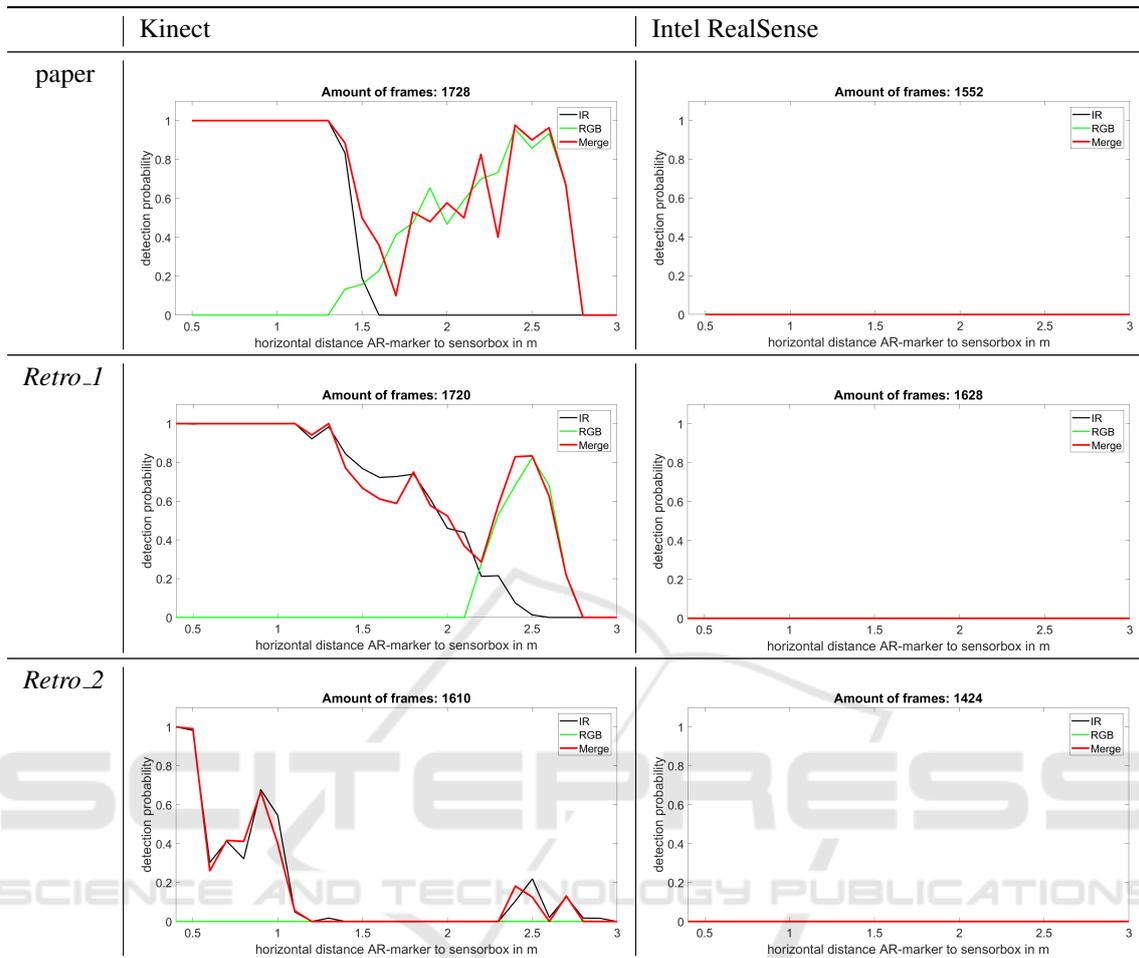
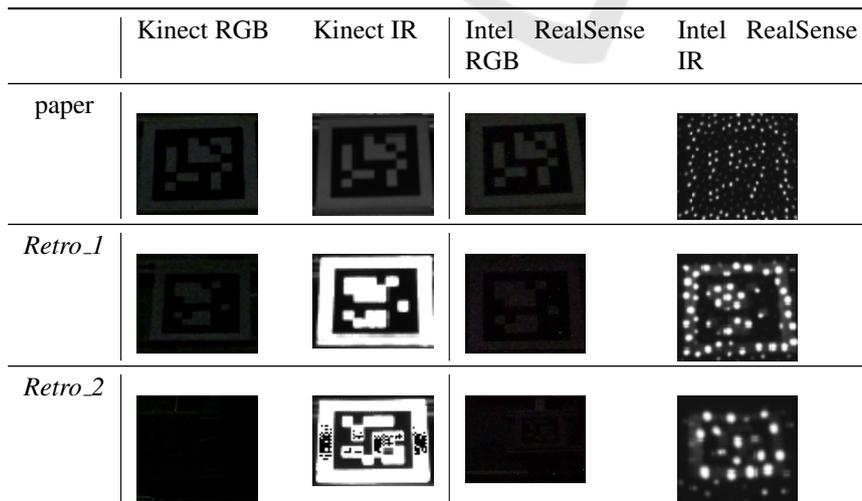


Table 4: Detection probability measurement pictures in an underexposed room.



3 and 4. The results of the Intel RealSense reveal that in no frames markers were detected. The reason for the very poor detection performance based on the data provided by the Intel RealSense can be seen in Table 4: Due to the reduced illumination in the room also the infrared light by external emitters is reduced, so that the structured-light dominates in the IR-images. That is the reason why the markers are pictured with bright dots. As the detection algorithm is based on the gradient image, the necessary edges for extracting the marker are not calculated appropriately. According to the expectations the RGB-image based detection does not work fine with low illumination. But interestingly the Kinect still detects the markers since a small light emitter was placed in the room. Consequently, the exposure adjustment of the Kinect supports the marker detection. The Kinect shows a reasonable performance with the *Retro\_1* marker regarding the IR-image. The detection behaviour with IR-images of the paper printed marker in the underexposed room is similar to the results described in Table 1. The detection of the *Retro\_2* marker is due to the already explained effects of the reflectance not accurate.

Conclusively the best marker detection performance can be achieved with the Kinect camera and the *Retro\_1* marker. Due to the very accurate detection with daylight and the reasonable behaviour on the IR-image data a parallel detection in both images and a following merging step stabilizes the detection due to redundancy.

## 5 CONCLUSION

We presented an augmented reality marker based on ArUco markers that can be detected in RGB- and IR-images. The usage of retroreflective material improves the marker detection in IR-images remarkably. Due to active perception by IR-capable camera systems the negative impact of external disturbances like change of light conditions on the marker detection is minimized. As different retroreflective materials influence the image quality depending on the camera system, we also examined two different retroreflective materials and compared the performance of the Kinect V2 and the Intel RealSense D435 regarding the detection probability depending on the distance of the augmented reality marker to the camera. According to our results the Kinect together with the microprisms based material for the markers shows the best detection performance. Since the marker detection is parallel processed on the IR- and RGB-image, it is stabilized due to redundant sensor fusion.

## REFERENCES

- Broggi, A. and Dickmanns, E. D. (2000). Applications of computer vision to intelligent vehicles. *Image and Vision Computing*, 5(18):365–366.
- Dorfmueller, K. and Wirth, H. (1998). Real-time hand and head tracking for virtual environments using infrared beacons. In *Modelling and Motion Capture Techniques for Virtual Environments*, pages 113–127. Springer.
- Draeos, M., Qiu, Q., Bronstein, A. M., and Sapiro, G. (2015). Intel realsense= real low cost gaze. In *ICIP*, pages 2520–2524.
- Fankhauser, P., Bloesch, M., Rodriguez, D., Kaestner, R., Hutter, M., and Siegwart, R. Y. (2015). Kinect v2 for mobile robot navigation: Evaluation and modeling. In *2015 International Conference on Advanced Robotics (ICAR)*, pages 388–394. IEEE.
- Fiala, M. (2010). Designing highly reliable fiducial markers. *IEEE Transactions on Pattern analysis and machine intelligence*, 32(7):1317–1324.
- Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., and Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292.
- Kato, H. and Billinghurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94. IEEE.
- Munoz-Salinas, R. (2012). Aruco: a minimal library for augmented reality applications based on opencv. *Universidad de Córdoba*.
- Naimark, L. and Foxlin, E. (2002). Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 27. IEEE Computer Society.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan.
- Rossner, H.-C. (2014). Retro-reflective marker. US Patent 8,915,599.
- Schoenborn, D. and Schulte, W. (1995). RSA Handbuch-sicherung von Arbeitsstellen an Strassen, Band 1: RSA mit Kommentar.
- Sim, R. and Little, J. J. (2006). Autonomous vision-based exploration and mapping using hybrid maps and rao-blackwellised particle filters. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 2082–2089. IEEE.
- Zhang, Z. (2012). Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10.