HAVE 2005 - IEEE International Workshop on
Haptic Audio Visual Environments and their Applications
Ottawa, Ontario, Canada, 1–2 October, 2005

Comparing ARTag and ARToolkit Plus Fiducial Marker Systems

Mark Fiala
National Research Council of Canada, NRC 1200 Montreal RD, Ottawa, Canada K1A-0R6
e-mail: mark.fiala@nrc-cnrc.gc.ca

Abstract – Fiducial marker systems are systems of unique patterns and computer vision algorithms that help solve the *correspondence problem*, automatically finding features in different camera images that belong to the same object point in the world. Fiducial marker systems consist of patterns that are mounted in the environment and automatically detected in digital images using an accompanying detection algorithm, useful for *Augmented Reality* (AR), robot navigation, 3D modeling, and other applications. This paper compares the two recently developed systems ARTag and ARToolkit Plus on their reliability, detection rates, and immunity to lighting and occlusion. Processing in fiducial systems are defined as two stages, *unique feature* detection and *verification/identification*. The systems are compared considering these stages, experimental results are shown.

Keywords – *fiducial marker self-identifying augmented reality ARToolkit ARToolkit-Plus ARTag.*

I. INTRODUCTION

Fiducial Marker Systems are computer vision systems that are designed to locate special markers added to the environment. These systems are useful for a whole host of robotics, augmented reality, photo-modeling and other applications where the capability to find correspondences in an image with points in 3D space, or correspondences between two images, is needed.

Correspondences can be found from "natural" features in images, these are unique features of pixel neighborhoods that are found without adding markers to the environment. Recent advances in feature detectors such as Lowe's SIFT operator [15] have raised expectations for the use of natural features. However, there will likely be a place for marker based systems for some time. Some scenes are not textured or unique enough and in many applications it is not overly inconvenient to mount passive marker patterns, such as instrumented rooms or hand-held 3D input devices. The use of markers can greatly increase the confidence in image-world correspondences over natural feature based methods with only a small number of pixels used, allowing highly reliable computer vision systems necessary for applications such as spacecraft docking [11]. It is likely that fiducial marker systems will find permanent uses in future systems, as well as providing temporary support for other vision systems that will migrate to marker-less systems as they mature.

Fiducial marker systems will be a building block of many systems using computer vision, several example applications are shown in Fig. 1. They are based on the marker system reporting markers detected in each image, a simple example of detection and overlay is shown in Fig. 2.

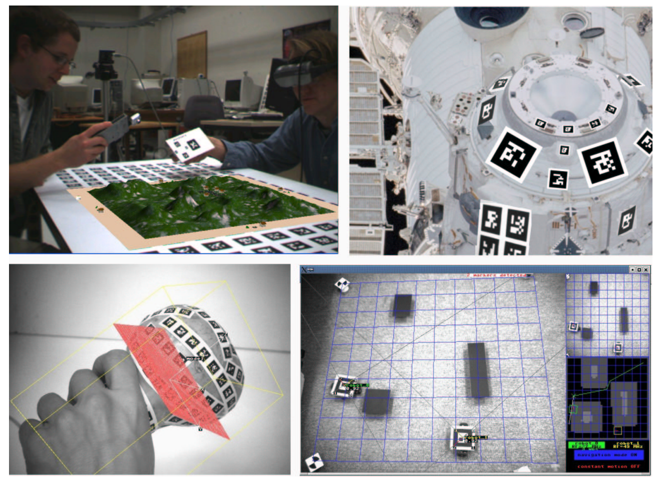

Fig. 1. Fiducial marker system *applications. (Upper left) Augmented reality application, terrain appears to sit on the table covered with a marker array. (Upper Right) Proposed fiducial use for spacecraft docking, relative pose can be reliably determined. (Lower left) HCI example of a 3D input device, pose of arbitrary shaped object determined by markers on its surface. (Lower right) Robot control, a stationary camera looks down on two robots with markers on their tops. First three examples use ARTag, last uses ARToolkit.*

Fiducial marker systems are only recently maturing, and provide vision systems that address the need for reliable computer vision systems able to work in varied applications, with varying cameras and lighting conditions. This paper seeks to compare two recent systems, ARTag and ARToolkit Plus, that are possibly the best suited for use in robust vision systems, as well as providing a brief summary of other fiducial markers systems from the literature and industry.

Fig. 2. *Fiducial marker systems automatically locate markers in images. Image shows ARTag marker system.*

II. ARTOOLKIT, ARTAG, AND ARTOOLKIT PLUS

In chronological order, ARToolkit was developed first followed by ARTag and then ARToolkit Plus. Each was inspired by the previous, and all three are available for download and use by researchers and system builders. Each comes with a set of markers, or tools to create them, and the computer vision algorithms to find and recognize the markers in digital images.

All three systems use square marker patterns mounted on a planar surface, they all have a square border with identifying information inside. The detection process for all three can be broken into two stages; *unique feature* detection and *verification/identification*. They first search the image for the unique feature, quadrilateral shapes that might be a perspective view of a marker's border and then check the interior (verification/identification) to determine if the feature is likely to be a marker or just some object in the environment. If it is deemed to be a marker its ID is determined. The output of the fiducial marker detection algorithms is a list of marker ID's and their corner locations in sub-pixel image coordinates. All three use the four corners of the quadrilateral to define a homography to allow perspective distortion to be addressed when examining the interior.

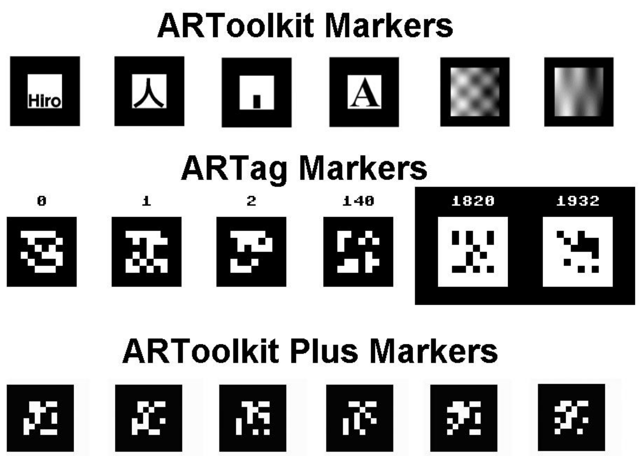

Fig. 3. *ARToolkit, ARTag, and ARToolkit Plus markers. ARToolkit (top row) was developed first and uses correlation for verification/identification. ARToolkit markers are user defined patterns, whereas ARTag and ARToolkit Plus markers are pre-designed.*

A. ARToolkit

Although *ARToolkit* is not compared in this paper, it is an important predecessor to the two systems compared; *ARTag* and *ARToolkit Plus*. ARToolkit [13] is a graphics toolkit that contains a fiducial marker system, it is widely used due to its availability (it is freely downloadable and works on several platforms). ARToolkit has enabled many vision systems, both those in the intended area of application of *Augmented Reality* [1], and as general purpose fiducial for use in areas such as human computer interaction (HCI) [4] and landmarks for robot navigation [9]. The markers have a black border on a white background and contain a greyscale image interior that is compared to a set of stored prototypes by correlation, several examples of ARToolkit markers are shown in Fig. 3.

With ARToolkit, the *unique feature* stage is performed by morphology operations on a binary image obtained through binarization. ARToolkit first finds black quadrilateral borders by finding connected groups of pixels below a set level, the contours of these groups are found and those contours with four straight sides are identified as potential markers.

This reliance on the image thresholding step is one weakness in ARToolkit, requiring controlled lighting so that a single greyscale level is sufficient to separate the black and white parts of the patterns. The most common modification of ARToolkit is a pre-processing stage of automatic thresholding, which improves the results but increases processing time and is susceptible to the *aperture problem* where the size of the marker and automatic thresholding window creates artifacts.

The *verification/identification* stage relies on correlation, the other drawback of ARToolkit. The rightmost two markers in Fig. 3 are DCT based markers from Owen *et al.*[5] designed to reduce inter-marker confusion with ARToolkit. The reference template images should be taken with the same camera under the same lighting as in the final application, in part to compensate for the non-linearity of the pattern printing process and camera irradiance digitization. The result is that ARToolkit suffers a high rate of *false positive* detections and frequently confuses one marker for another. However, in controlled circumstances it can be made to work well, and the interior patterns can be arbitrary shapes more pleasing and meaningful for human viewers than the next systems.

B. ARTag

ARTag was inspired by ARToolkit, but replaced both processing stages. The *unique feature* of quadrilateral outlines is found with an edge-based approach instead of morphology on a binary thresholded image, and the *verification/identification* stage uses a robust digital encoding method instead of correlation. ARTag was also inspired by the industrial 2D-barcode *Datamatrix* (Fig. 4) for the use of digital encoding. ARTag enjoys reduced sensitivity to lighting variation and immunity to partial occlusion as a benefit of its edge based method. an improvement over ARToolkit. The possibly more important improvement is the adoption of digital techniques for verification/identification that lead to vanishingly low false positive and inter-marker confusion rates. The use of error correction theory allows the markers to still be detected despite some errors in reading the digital pattern, providing a lower *false negative* rate than both ARToolkit and ARToolkit Plus as shown in Section C. ARTag and ARToolkit are directly compared in [12] where several metrics such as the *false positive*, *false negative* and *inter-marker confusion* rates are compared.

ARTag is described in [8], [10]. Like other fiducial marker systems, ARTag consists of a library (middle row of Fig. 3) of 2D patterns which are detected in digitized camera images. With ARTag 2002 markers can be used, they are specially designed by optimizing parameters in the digital algorithms to provide the most uniqueness between markers. The use of *Hamming distances* was used, and is used later herein when comparing to ARToolkit Plus.

C. ARToolkit Plus

The robustness of ARTag was recognized, and its reliability spurred the creation of at least one competing system; ARToolkit Plus. *ARToolkit Plus* [1] is a new system, while inspired by ARTag, does not contain all ARTag's features such as edge-based detection and robust digital processing such as error correction. ARToolkit Plus uses the same unique feature detection step, and only differs from ARToolkit in the verification/identification step.

ARToolkit Plus markers look very similar to ARTag markers, some are shown in the bottom row of Fig. 3, ARTag's digital 6x6 array structure and reliance on only bi-tonal levels (instead of varying grey levels of the interior pattern as in the original ARToolkit) is repeated in ARToolkit Plus. However, ARToolkit Plus markers encode the interior differently.

This paper focuses on comparing ARToolkit Plus and ARTag and seeks to show that ARTag is a more robust fiducial marker system.

III. OTHER FIDUCIAL MARKER SYSTEMS

Fiducial marker systems consist of the printed patterns and the computer vision algorithms to identify them in an image. There are several fiducial marker systems available, several are shown below in Fig. 4, of which ARTag is the system used in this paper. Most likely use the two step process of first locating a *unique feature*, followed by *verification/identification*.

Many use a quadrilateral boundary as a unique feature, where the four corners can be used to create a *homography* to sample the interior in a perspectively correct manner. A symbolic (digital) encoding method is used with the interior divided up into several *cells* which are coloured white or black. The rotation of the quadrilateral markers must be found in order to correctly interpret the interior, Canon's marker [7] and Binary Square Marker [3] use the four corner cells in their 3x3 or 4x4 cells respectively leaving 5 or 12 digital bits left to encode identity. *Cybercode* [17] uses a black bar to locate one side, and can be extended to an arbitrary size but does not provide information that allows it to be detected under perspective distortion. The ARVIKA markers (SCR, IGD, HOM) [2] use a 4x4 array, information is not available for these proprietary markers but it is clear that they can store either 16 or 20 bits.

"Blob" analysis can be used instead of quadrilateral outlines, connectivity is performed on a binarized image to identify adjacent and enclosed regions. Several commercially available circular fiducial marker systems exist using a single broken annular ring around a circular dot such as Photomodeler's "Coded Marker Module" [3]. Naimark and Foxlin [16] describe a system extending the number of rings carrying 15 bits of digital information. Bencina's "reacTIVision" marker [2] creates a tree of topology of what blobs are contained inside other blobs, trees matching a set library are used to find markers.

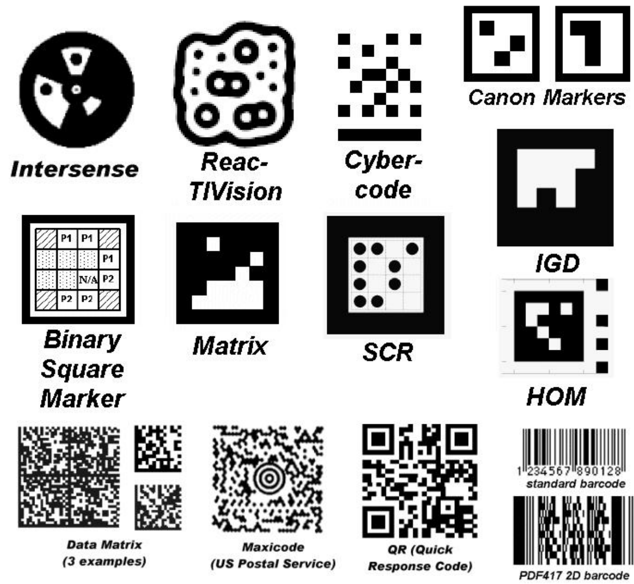

Fig. 4. *Other fiducial marker systems and barcode/2D-barcode systems.*

Most of these marker systems are not available for evaluation, due to non-disclosure or not being a fully developed system. There are several comparisons for the available and developed systems. Zhang [18] and Claus [6] perform a survey of several fiducial marker systems (ARVIKA markers and ARToolkit) with respect to processing time, identification, image position accuracy with respect to viewing angle and distance.

Also shown at the bottom of Fig. 4 are some other visual markers that at first appear similar to fiducial marker systems, but are intended for other uses and are not suitable for fiducial markers. The purpose is to carry information, not to identify in a camera under perspective distortion as is needed for fiducial marker systems, they typically require occupying a large region of the camera field of view and are not useful to be located at a distance.

IV. COMPARING *ARTAG* AND *ARTOOLKIT PLUS*

A. Step 1: Unique Feature Detection - Comparing Immunity to Occlusion and Lighting Variation

Both ARTag and ARToolkit Plus first search the input image with the quadrilateral boundaries as the unique feature. This affects the *false negative rate*, the rate of not finding a marker when it is indeed present. Only detected quadrilaterals that are found in this stage are passed to stage 2 to examine their interiors.

ARToolkit Plus uses the same thresholding scheme as the original ARToolkit and so suffers from the same sensitivity to lighting occlusion and variation as described in [12]. They

[1] http://studierstube.org/handheld_ar/artoolkitplus.php
[2] http://www.arvika.de/www/index.htm
[3] http://www.photomodeler.com

both require a complete unbroken quadrilateral border, unlike ARTag's ability to still hypothesize about borders broken by objects or the edge of the screen. Fig. 5 demonstrates this.

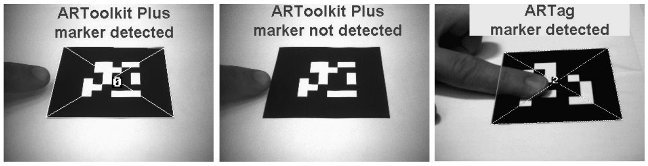

Fig. 5. *Marker detection with occlusion. ARToolkit Plus performance (Left) and (Middle) shows how detection is lost with even small occlusion of the boundary. (Right) shows how ARTag markers can be recognized even despite large occlusions.*

Likewise the quadrilateral detection fails in ARToolkit and ARToolkit Plus when there is variable lighting in an image. Similar sized arrays of ARToolkit Plus and ARTag markers were printed out and mounted on a surface in the same position with a fixed camera and light sources, and an image was captured for each of the two marker systems. Fig. 6 shows how only a subset of markers are found in the same image when different thresholds are applied. Fig. 7 shows the result with ARTag.

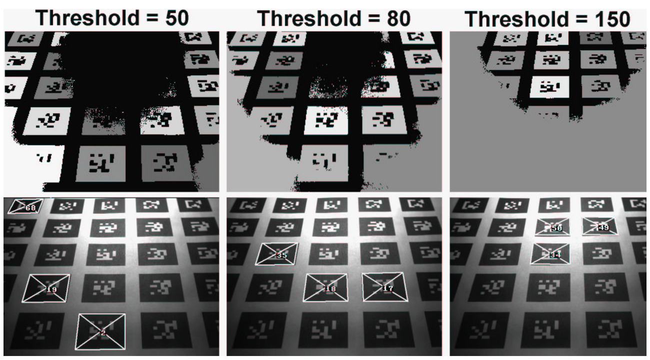

Fig. 6. *ARToolkit Plus marker detection with challenging lighting. ARToolkit Plus relies on a binarization stage using a single intensity threshold for the entire image. The top row shows an image thresholded at 3 different grey levels, the bottom row shows the markers detected at those thresholds. Only three markers are detected at each threshold.*

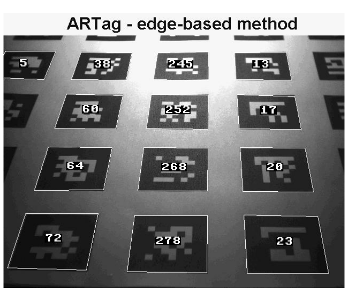

Fig. 7. *ARTag marker detection under identical lighting conditions.*

It is possible to simply call the ARToolkit Plus detection function several times for each image and amalgamate the results, but this would not solve the case where irradiance variations across a single marker could not be classified by one threshold. For example consider if lighting across a large marker caused the black in a bright end to be brighter than the white in darker end.

It should be noted that ARToolkit Plus does offer a type of automatic thresholding, one that is performed temporally rather than within a single image. An automatic threshold selection function is provided where the single global (image-wide) threshold is adapted according to the pixel distributions of successfully detected markers in a previous frame, and a randomly moving threshold when no markers are detected. This frees the user from manually setting a threshold but still wouldn't get better results than shown in Fig. 6. A local neighborhood thresholding algorithm, such as employed in [4] would yield better results, however this would encounter the aperture problem.

B. Step 2: Verification/Identification - Inter-marker Confusion Rates

As well as a low false positive rate, a good marker system should have a low rate of confusion between markers, the *inter-marker confusion rate*. The user should have a high confidence that one marker won't be mistaken for another. With proper marker system design the rate that this occurs can be minimized, ideally to very low levels that don't appear in practice. Owens *et al.* [5] explores this issue for the original ARToolkit.

The inter-marker confusion rate can be analyzed by considering the probability of mistaking one code of digital symbols for another. A measure of how easily two binary codes can be confused with each other is to calculate the *Hamming distance*[14], which is simply the sum of the differences between two digital sequences. The probability of an inter-marker confusion event can be calculated using knowledge of the Hamming distances within a marker set as described in [10]. This Hamming distance histogram is a metric to optimize in system design, *i.e.* to reduce inter-marker confusion. We seek to reduce the frequency of those of low values of Hamming distance, as that they represent cases when only a small number of bits need to be flipped to confuse one marker for another. The more that the non-zero part of the histogram can be pushed out to the right (if plotted as in this paper), the more immune to inter-marker confusion a marker set will be.

Hamming histograms are shown in Fig. 8, for a partial or complete usage of the library. ARTag has better histograms for both, especially so if only 50 markers from each system are used (in most applications the entire library will probably not be used). This translates into a much lower inter-marker confusion rate for ARTag. If only the first 50 are used we can see that there are cases where there is only a 4 bit difference between markers for ARToolkit Plus, whereas with ARTag the closest markers are at a 12-bit difference. With every single bit increase in the minimum Hamming distance is a geometrically decreasing confusion probability.

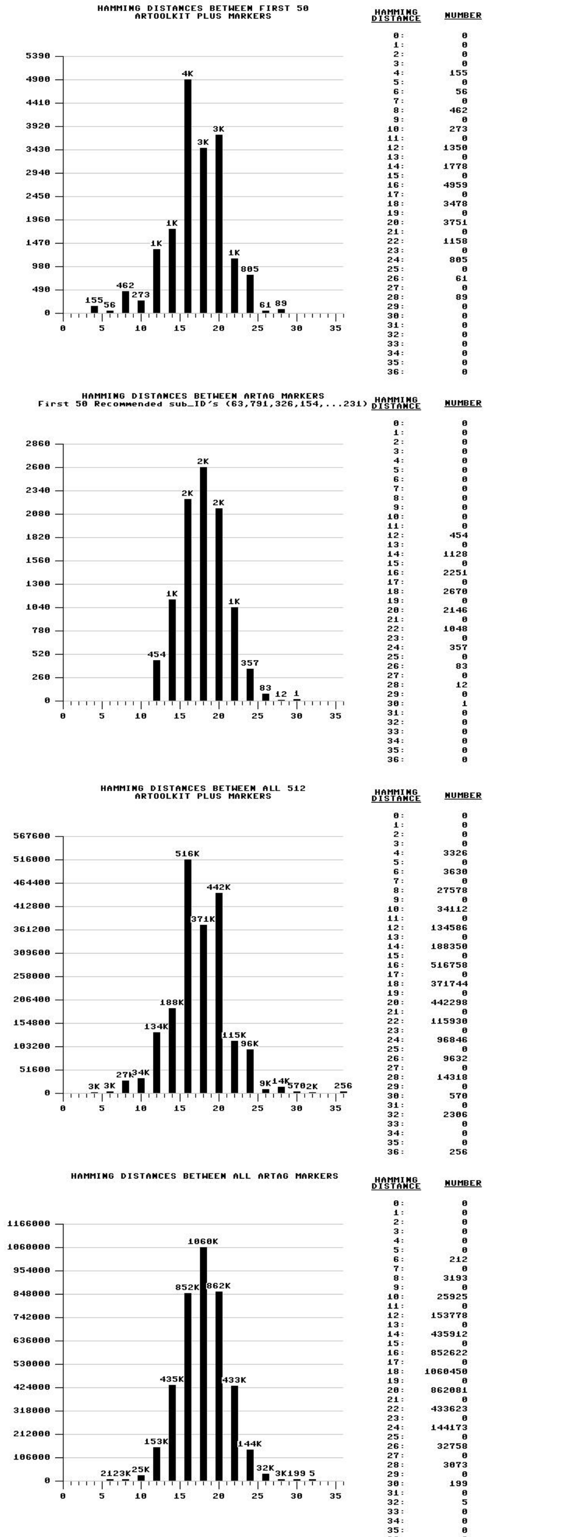

Fig. 8. *Hamming distance histograms, the less entries there are in the lower distance bins, the lower the* inter-marker confusion rate *will be. ARToolkit Plus is shown in the top 2 plots, ARTag on the bottom. The left plots show the histograms if only the first 50 markers in each system are used, the right plots are of all markers in the library. ARToolkit Plus has 512 possible markers, ARTag has 1001 (when only black borders on a white background are used).*

The relatively inferior Hamming histograms of ARToolkit Plus, and hence higher inter-marker confusion rate, is a direct result of how the two systems encode their digital content. Both start with a marker ID number, 10 bits for ARTag, 9 for ARToolkit Plus. These are then turned into a 36 bit code which is laid out in 6 rows to form the marker interior. How the shorter ID number is turned into the longer 36 bit code directly affects the inter-marker confusion rate. ARToolkit Plus merely repeats the ID 4 times to create 36 bits, and then XOR's the result with a fixed mask (011011011100001001101001110000100111), whereas ARTag uses digital convolution of the ID with checksum and error correction codes. ARTag's checksum and error correction codes were specifically chosen to create the best Hamming distance histogram (see [10]).

ARToolkit Plus uses an XOR mask to reduce Hamming distances between markers when rotated, however this does not help the un-rotated Hamming distance between markers since an XOR operation will not change the Hamming distance. Since each marker ID is simply repeated 4 times, half of all sequential ID pairs will have a Hamming distance of only 4 bits (incrementing a binary even to odd number involves changing only the lowest bit), creating the first non-zero histogram bin at HD=4. This first non-zero bin has the largest effect on the inter-marker confusion rate.

An experiment was performed verifying the difference in inter-marker confusion rates. An array of 24 markers was created for each system, and imaged under identical conditions where all markers were detected. One image was captured for ARToolkit Plus, and one for ARTag. Each image was then subjected to varying levels of additive gaussian noise. For each noise level (sigma=standard deviation) 10,000 iterations were performed and the inter-marker confusion rate plotted (Fig. 9).

C. *How often are markers detected: False Negative Rate*

How the robust unique feature detection and verification/identification stages are designed also has affect on another very important quantity: the *false negative* rate, which is the probability that a marker will be missed. *I.e.* what is the rate that a marker is present in the camera field of view, properly lit, but not detected? Ideally the false negative rate would be zero.

Image noise can affect both the detection of the quadrilateral boundary and the decoding of the interior pattern. The error correction coding scheme gives ARTag an advantage, it allows recovery from situations where noise has corrupted the digital code read from the marker interior. One would expect ARTag to have a lower false negative rate, and thus function better in non-ideal, noisy images.

An experiment was performed to analyze the effect of image noise on the false negative rate, similar to the experiment in Section B. The same pair of images were subjected to varying levels of additive gaussian noise and the false labelling of ID's for markers were recorded over 10,000 iterations for each noise

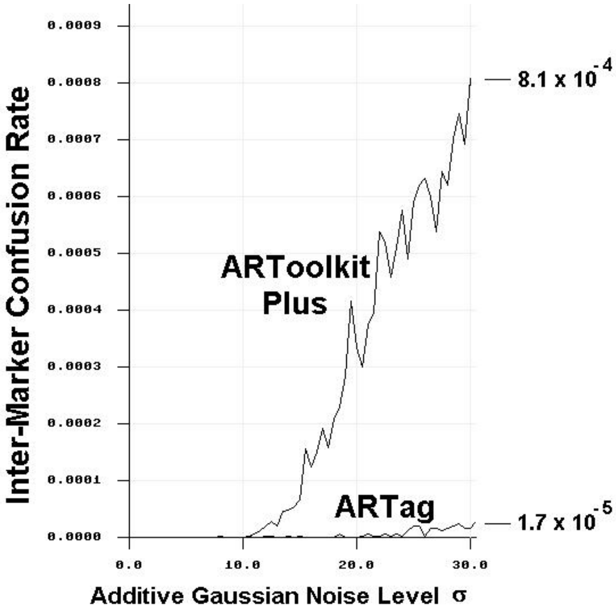

Fig. 9. *Effect of additive gaussian noise on inter-marker confusion. Plot of the inter-marker confusion rate as a function of gaussian std. dev. (in pixels).*

level. Fig. 10 verifies the hypothesis of ARTag having a lower false negative rate in the presence of noise.

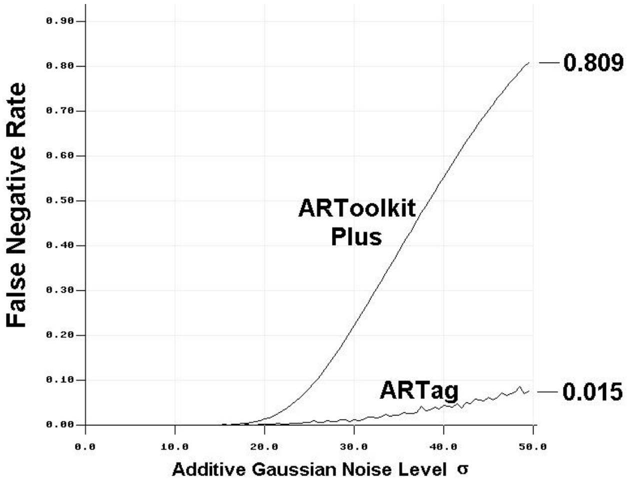

Fig. 10. *Effect of additive gaussian noise on marker detection. (Left 4 images) Input images used and example of detection with additive gaussian noise of std. dev. 40 pixels. Plot of the false negative rate as a function of gaussian std. dev. (in pixels).*

V. CONCLUSIONS

Fiducial marker systems are systems of unique patterns and computer vision algorithms that help solve the *correspondence problem*, and enable many computer vision applications. Two recently developed systems, ARTag and ARToolkit Plus, were compared for their performance by analyzing their design. Image processing in fiducial systems was defined in two stages, *unique feature* detection and *verification/identification*. It was shown how the edge-based approach of ARTag allows more unique features (quadrilateral contours) to be detected than ARToolkit Plus's binary morphology method based on greyscale thresholding. It was shown how ARTag is able to function despite lighting variations and partial occlusion. The digital processing that both systems use for the *verification/identification* stage was compared for the false positive and inter-marker confusion rates. Hamming distance analysis showed how ARTag achieves, by design, a lower possible inter-marker confusion rate.

[1] *2003 IEEE / ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003), 7-10 October 2003, Tokyo, Japan*. IEEE Computer Society, 2003.

[2] R. Bencina, M. Kaltenbrunner, and S. Jorda. Improved topological fiducial tracking in the reactivision system. In *Proceedings of the IEEE International Workshop on Projector-Camera Systems (Procams 2005)*, San Diego, USA, 2005.

[3] P. Boulanger. Application of augmented reality to industrial tele-training. In *Proc. of CRV'04 (Canadian Confernence on Computer and Robot Vision*, pages 320–328, May 2004.

[4] G. Roth C. McDonald and S. Marsh. Red-handed: Collaborative gesture interaction with a projection table. pages 773–778.

[5] F. Xiao C. Owen and P. Middlin. What is the best fiducial? In *First IEEE International Augmented Reality Toolkit Workshop (at ISMAR)*, Sep 2002.

[6] D. Claus and Fitzgibbon. Reliable fiducial detection in natural scenes. In *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, May 2004.

[7] H. Yamamoto D. Kotake, S. Uchiyama. A marker calibration method utilizing a priori knowledge on marker arrangement, Nov. 2004.

[8] M. Fiala. Artag, a fiducial marker system using digital techniques. In *CVPR'05*, pages 590 – 596.

[9] M. Fiala. Artoolkit applied to panoramic vision for robot navigation. In *Proc. of Vision Interface*, pages 119–127, June 2003.

[10] M. Fiala. Artag revision 1, a fiducial marker system using digital techniques. In *National Research Council Publication 47419/ERB-1117*, November 2004.

[11] M. Fiala. Artag fiducial marker system applied to vision based spacecraft docking. In *IROS 2005 IEEE/RSJ Workshop on Robot Vision for Space Applications (held with IROS'2005)*, pages 35–40, Aug 2005.

[12] M. Fiala. Fiducial marker systems for augmented reality: Comparison between artag and artoolkit. In *MIRAGE 2005 :Computer Vision/Computer Graphics Collaboration for Model-based Imaging, Rendering, image Analysis and Graphical special Effects*, INRIA Rocquencourt, France, Mar 2005.

[13] I. Poupyrev H. Kato, M. Billinghurst. *ARToolkit User Manual, Version 2.33*. Human Interface Technology Lab, University of Washington, 2000.

[14] R. Hamming. Error detecting and error correcting codes. In *Bell Systems Technology Journal*, volume 29, pages 147–160, Apr 1950.

[15] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60:91–110, 2004.

[16] L. Naimark and E. Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *ISMAR 2002: IEEE / ACM International Symposium on Mixed and Augmented Reality, Darmstadt,Germany*, Sept. 2002.

[17] J. Rekimoto and Y. Ayatsuka. Cybercode: Designing augmented reality environments with visual tags. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 1–10, Elsinore, Denmark, Oct 2000.

[18] X. Zhang, S. Fronz, and N. Navab. Visual marker detection and decoding in ar sytems: A comparative study. In *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 97–106, Sep 2002.