Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Performance analysis of visual markers for indoor navigation systems[#]

Gaetano C. LA DELFA[†1], Salvatore MONTELEONE[1], Vincenzo CATANIA[1],
Juan F. DE PAZ[2], Javier BAJO[3]

([1]*Department of Electrical, Electronics and Computer Engineering (DIEEI), University of Catania, Catania 95125, Italy*)
([2]*BISITE Research Group, Faculty of Science, University of Salamanca, Salamanca 37007, Spain*)
([3]*Department of Artificial Intelligence, Polytechnic University of Madrid, Madrid 28660, Spain*)
[†]E-mail: gaetano.ladelfa@dieei.unict.it

Received Oct. 8, 2015; Revision accepted Feb. 17, 2016; Crosschecked July 14, 2016

Abstract: The massive diffusion of smartphones, the growing interest in wearable devices and the Internet of Things, and the exponential rise of location based services (LBSs) have made the problem of localization and navigation inside buildings one of the most important technological challenges of recent years. Indoor positioning systems have a huge market in the retail sector and contextual advertising; in addition, they can be fundamental to increasing the quality of life for citizens if deployed inside public buildings such as hospitals, airports, and museums. Sometimes, in emergency situations, they can make the difference between life and death. Various approaches have been proposed in the literature. Recently, thanks to the high performance of smartphones' cameras, marker-less and marker-based computer vision approaches have been investigated. In a previous paper, we proposed a technique for indoor localization and navigation using both Bluetooth low energy (BLE) and a 2D visual marker system deployed into the floor. In this paper, we presented a qualitative performance evaluation of three 2D visual markers, Vuforia, ArUco marker, and AprilTag, which are suitable for real-time applications. Our analysis focused on specific case study of visual markers placed onto the tiles, to improve the efficiency of our indoor localization and navigation approach by choosing the best visual marker system.

Key words: Indoor localization, Visual markers, Computer vision
http://dx.doi.org/10.1631/FITEE.1500324 **CLC number:** TP391; TN2

1 Introduction

The massive, worldwide diffusion of the smartphone and its high hardware performance have contributed in recent years to the creation of the conditions for significant technological progress in the mobile consumer sector. Manufacturers continuously add new sensors to their latest devices, giving developers and startups the perfect instrument to create innovative applications and services, which have radically changed the citizen's way of life. Most of these applications and services are strictly related to the user position and context information: they are defined as location based services (LBSs), provide to the users a lot of functionalities based on their proximity to a specific point of interest, and are becoming very popular. In outdoor environments, the Global Positioning System (GPS) is almost a 'de facto' standard for positioning and navigation, but in indoor environments there does not actually exist a unique technology to solve the problem. For this reason LBSs inside buildings are not very common today. Various approaches and solutions have been proposed to address the challenge in a simple and

[#] A preliminary version was presented at the 6th International Symposium on Ambient Intelligence in University of Salamanca, Spain, June 3–5, 2015

ORCID: Gaetano C. LA DELFA, http://orcid.org/0000-0002-1842-5467

scalable way, and also a lot of commercial solutions are appearing on the market (Mautz, 2012). Among these, the most successful are those that take advantage of the hardware/sensors of the smartphone to extract contextual information and use them to localize the user. In Fig. 1, we give a brief and non-exhaustive summary of the main techniques used for indoor localization.

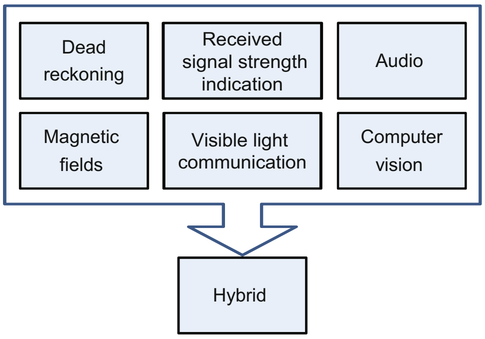

Fig. 1 Overview of the main indoor localization and navigation techniques

Dead reckoning systems use accelerometers, magnetometers, and gyroscope sensors embedded in the smartphones to provide a fast estimate of the user position. Because of high drift errors introduced by the sensors, usually a step counter is used to calculate the covered distance and a periodical recalibration is performed to reset from the error (Beauregard and Haas, 2006; Li *et al.*, 2012; Buchman and Lung, 2013; Liu *et al.*, 2013; Bajo *et al.*, 2015). Received signal strength indication (RSSI) systems exploit the RSSI of the radio signals present in the environment, typically, wireless-fidelity (Wi-Fi) signals, available for free in public buildings, or, recently, Bluetooth low energy (BLE) signals. They use triangulation and trilateration or, more frequently, information from a previously generated RSSI fingerprint database of the environment to estimate the position of the user (Fuchs *et al.*, 2011; Liu *et al.*, 2012; Buchman and Lung, 2013; Han *et al.*, 2014; Bajo *et al.*, 2015). Audio systems exploit controlled (usually malls, consumer stores, and museums are equipped with loudspeakers) or uncontrolled ambient sounds (for example, acoustic background fingerprint) to allow a simple smartphone to cheaply determine its location (Mandal *et al.*, 2005; Tarzia *et al.*, 2011). Magnetic field systems use the indoor ambient magnetic fields (caused, for example, by elevators, escalators, doors, or pillars) to build a magnetic map of the environment. This map is used by the smartphone to solve the indoor localization problem (Haverinen and Kemppainen, 2009; Subbu and Sasidhar, 2011). Visible light communication (VLC) systems exploit the susceptibility of light emitting diodes (LEDs) to amplitude modulation at high frequencies to transmit information into the environment. If the frequency is greater than a flicker fusion threshold, the lighting functionality is preserved because the modulation is not perceivable by the human eyes, and it is possible to perform accurate indoor positioning (Danakis *et al.*, 2012; Jovicic *et al.*, 2013). Recently, thanks to high performance cameras and high computational capabilities of new generation smartphones, researchers are focusing on computer vision systems which rely on complex, CPU-intensive (1) marker-less or (2) marker-based computer vision algorithms to determine the position of the user in the environment (Saito *et al.*, 2007; Arias and April, 2011; Chandgadkar and Knottenbelt, 2013). Usually, hybrid techniques and technologies are used to improve the accuracy, reduce costs, and enhance the performance of the whole indoor positioning system (Wang *et al.*, 2012; Zachariah and Jansson, 2012).

This work is an extension of our previous one (La Delfa *et al.*, 2015). In particular, in the present paper we extended the related work section, by going deeper in the presentation of some papers, and by introducing some other interesting indoor localization solutions, such as IndoorAtlas, which is based on the magnetic fields. Moreover, Section 4 further focuses on the presentation of the most important visual markers in the literature, shows the experimental results we obtained by analyzing three kinds of markers, Vuforia, ArUco marker, and AprilTag, and introduces a benchmarking tool useful for comparing the performance of such markers under different conditions.

2 Related work

Indoor navigation is now a very hot topic with a lot of research over the last decade, and there have even been some commercial solutions actually used in places such as museums or big shopping centers. Various solutions have been proposed in the literature. They can be classified, from the point of view of approach, into two main categories: the first exploits data coming from sensors to track the user or detect his/her position, and the second uses the information extracted from cameras and computer vision techniques to reach the same goal. In the following, we present an overview of proposals belonging to both categories.

2.1 Sensor-based approaches

Researchers from Duke University proposed UnLoc (Constandache *et al.*, 2010; Wang *et al.*, 2012) to face the problem by both resetting the drift error generated by the smartphone inertial sensors and avoiding any previous knowledge of the building. Their approach merges environmental sensing and dead reckoning to realize an indoor navigation system, based on the hypothesis that certain locations in indoor environments present—in the sensor domains—identifiable signatures (landmarks) generated by elevators, escalators, Wi-Fi, etc. They use dead reckoning to track the user, and periodically reset the error when the user encounters a landmark. The estimation of the landmark positions (initially unknown) and the identification of new ones are performed by elaborating the data coming from all the users: every new user improves the previous measurements. Jovicic *et al.* (2013) suggested the use of LEDs and VLC to localize the user inside an environment in an accurate way. On the transmitter side, the modification to the LED lighting infrastructure is cheap and simple—power-efficient switch-mode amplifiers are already present on the LED lamps, so the only cost comes from the programmable logic devices which drive the amplifier; on the receiver side, Harald Haas (one of the pioneers in this field) showed that it is possible to exploit the rolling shutter effect of complementary metal oxide semiconductor (CMOS) based camera sensors (Meingast *et al.*, 2005) to let a mobile phone decode the information transmitted by the LEDs' infrastructure (Danakis *et al.*, 2012). Several startups such as ByteLight proposed indoor navigation commercial solutions based on VLC. Other researchers and startups tried to use magnetic fields for positioning. One of these startups, IndoorAtlas (Haverinen and Kemppainen, 2009), inspired by the capability of some animals that use the Earth's magnetic fields for orientation detection and navigation, proposed a commercial indoor positioning solution based on a similar principle. They supposed that modern buildings often contain steel structures, which create a sort of magnetic fingerprint of the environment and then they exploited these anomalies to localize the user. Apple and Google included application programming interface (API) for indoor positioning in their software development kits (SDKs) for iOS and Android. Apple, in particular, provided an enhancement on its core location framework to let developers make an easy transition between outdoor and indoor navigation. Moreover, they provided for free a new portal to add or edit local business listings (with some features such as public access in the building, one million or more of visitors per year, Wi-Fi enabled, etc.): Apple Maps Connect. The listings (or corrected listings) appear on Apple Maps on the personal computer (PC) and the mobile, so the user can track himself/herself inside them. Behind the scenes, Apple (and Google, which offers a similar feature) uses mixed technologies such as Wi-Fi fingerprint, BLE iBeacons (Martin *et al.*, 2014; Villarrubia *et al.*, 2014), and dead reckoning to perform the indoor localization task. The list of approaches, techniques, and technologies to solve such a complex problem is actually very long.

2.2 Computer-vision-based approaches

Among all of these, as aforementioned, researchers are focusing on computer vision algorithms which use (1) marker-less approaches or (2) marker-based approaches for indoor localization. Marker-less approaches are used when visual markers are undesirable because of aesthetic reasons; they rely on what the camera sees to deduce the position of the user and usually require pre-knowledge of the environment: images are captured at predefined locations and processed to extract unique features. A database of these features is created, along with the associated camera position and orientation, and it is used as a visual fingerprint of the building (Aider *et al.*, 2005; Torres-Solis *et al.*, 2010; Bitsch Link *et al.*, 2012). Unfortunately, marker-less systems are really CPU-intensive and often require a considerable workload before they can start to work and need frequent recalibrations. Moreover, if the environment changes for some reason, depending on how relevant the changes are, the visual fingerprint needs to be recreated. Marker-based approaches rely on 2D visual markers, which can be easily decoded even by a low-cost smartphone, to let the user track his/her position inside buildings. Ecklbauer (2014) proposed the recognition of multiple custom ArtoolKit visual markers in a camera image to deduce the position of an Android smartphone, with no additional data sources, except the knowledge of the markers' positions. Moreover, in the first part of his thesis, he compared the detection/decoding performance of QR-code against that of ArtoolKit markers, showing some interesting measurements about speed detection and detection rate under different light conditions and for different sizes of markers. Mulloni *et al.* (2009) proposed a similar technique used for continuous navigation through a smartphone inside a building. Chandgadkar and Knottenbelt (2013) proposed a simple color-based 2D visual marker to obtain the user position and orientation and a step detection algorithm to track the user between two markers. The system relies on the robust OpenCV library for marker detection and for avoiding obstacles along the path. Despite the simplicity and scalability of these techniques, there are some drawbacks such as (1) the need for a line of sight, (2) the sensitivity to light changes, (3) the size of the marker, which must be as small as possible in order to be minimally invasive, (4) the fact that the APP does not work in real time, and (5) the cognitive workload for the user who has to look for the marker in order to auto-locate himself/herself (these procedures, if annoying for fully-sighted people, can become very difficult for people with visual deficits). To overcome these drawbacks, we proposed previously a hybrid approach which uses BLE to locate the user when there is no line of sight (and the localization APP is in background mode) or a 2D visual marker system deployed on the floor (in order to let the user auto-locate himself/herself without any cognitive workload: in fact when he/she launches the APP in navigation mode, the camera is in the palm of his/her hand and will be directed towards some part of the floor) to estimate the position with a good level of accuracy (depending on the density of tags) (La Delfa and Catania, 2014). To guarantee efficiency, accuracy, minimal invasivity, and real-time performance for our system, it is important to choose the right marker according to the particular situation of deployment. In the previous paper we proposed the use of an ArUco marker. We evaluated in a qualitative and empirical way the performance of three 2D visual markers: a Vuforia frame marker by Qualcomm (2014), an ArUco marker by the A.V.A. group from University of Cordoba (Garrido-Jurado *et al.*, 2014), and an AprilTag from the University of Michigan (Olson, 2011; Richardson *et al.*, 2013). In particular, we focused on the following features: (1) needed marker size for correct detection (the smaller the marker is, the less invasive the system will be); (2) detection and decoding speed (it is important to guarantee an almost real-time performance); (3) sensitivity to changes in light conditions, blurring (caused by fast movements of the user), and partial occlusion. We chose these markers because they are well-documented, opensource (AprilTag and ArUco), or with a free-to-use SDK (Vuforia) available, and have good global performance.

3 Visual markers deployed on the floor: requirements

To realize an efficient indoor navigation system using 2D visual markers deployed on the floor, a critical point is the choice of the visual marker which best fits the particular place of deployment. We focus on some of the intrinsic features of the system, and on how we can exploit them to improve the performance:

1. Almost uniform, prior-known background pattern of the floor as shown in Fig. 2a. It is possible to use this feature to improve the speed of the decoding algorithm and to reduce the physical size of the marker.

2. Almost fixed, prior-known size of the marker inside the frame as shown in Fig. 2b. It depends on the distance between the camera (which is on the palm of the hand) and the floor, and makes it easier and faster to find the marker in the frame, which brings an improvement to the detection speed.

3. Major probability for the marker to be in the upper part of the frame as shown in Fig. 2c, due to the fact that when the user launches the APP to navigate inside a building, he/she moves forward. It is possible to analyze a sub-portion of the frame and further improve the detection speed or use the saved time to apply some filters to the sub-portion in order to enhance the quality of the image.

4. Prior-known marker positions. So, it is possible to reduce the errors by considering that each decoded marker must be one in the boundary of the previously decoded marker.

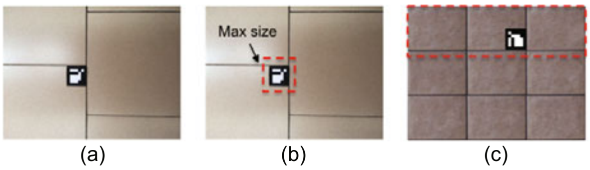

Fig. 2 Uniform background pattern of the floor (a), the maximum size of the marker inside the frame (b), and the major probability for the marker to be in the upper part of the frame (c)

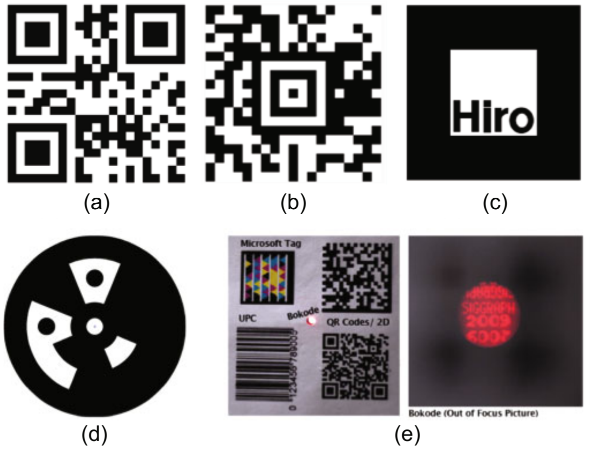

Fig. 3 Visual markers examples: (a) QR-code; (b) Aztec code; (c) ArtoolKit; (d) intersense; (e) Bokode compared to normal visual markers

The characteristics that the chosen marker must have, considering that the user is moving when using the system (usually with low speed), are:

1. Small size: This feature is required to reduce the invasivity of the system. We have to find the best compromise among size, speed of detection/decoding, and robustness of the algorithm.

2. Real-time detection: To make the auto-localization process through visual markers transparent for the final user, the detection must be as fast as possible.

3. Robustness to changes in light conditions: This feature is required because typically the marker will be deployed in highly dynamic environments characterized by the presence of other people, on/off switching of lights, shadows, etc.

4. Robustness in detecting blurred or out-of-focus markers, caused by movement which is too fast.

In the next section, we give a brief overview of the most famous markers and discuss specifically three 2D visual markers which best fit our requirements: Vuforia, ArUco marker, and AprilTag.

4 Real-time visual markers: Vuforia, ArUco, and AprilTag

A visual marker system is composed of a set of 2D visual markers and a computer vision algorithm to detect and decode each marker using a smartphone camera or other computer vision technologies. Today, thanks to their low cost, flexibility, and simplicity (a simple smartphone is able to generate and read most of them), there are several visual markers in the market. Fig. 3 shows some of them. The most well-known and used visual marker is probably the QR-code (Fig. 3a) (Denso, 2010): it can store up to 4296 alphanumeric characters, and contains a Reed-Solomon error correction algorithm (Wicker and Bhargava, 1994), which allows for decoding of even partially occluded or degraded QR-codes. Moreover, it is opensource, well-documented and there are hundreds of free libraries to generate and decode it. Unfortunately, it does not have real-time performance, which makes it unsuitable for our purposes. Aztec code (Fig. 3b) (Longacre and Hussey, 1997) is similar to the QR-code (large amount of stored data, Reed-Solomon error correction algorithm), but it does not need a white border to be correctly decoded. To guarantee real-time performance, usually a visual marker which stores just a simple binary code is used. An example (Fig. 3c) is the ArtoolKit marker (Kato and Billinghurst, 1999). Originally developed in 1999 by Hirokazu Kato, the ArtoolKit library relies on a template-matching algorithm to detect the marker. Thanks to that, the shape of an ArtoolKit marker can theoretically be any image, surrounded by a black square. Other than the classical square markers we also have circular markers, which are more robust against perspective distortion and more precise; intersense (Naimark and Foxlin, 2002), shown in Fig. 3d, is a commercial, patented, circular marker with high performance. Fig. 3e shows Bokode, an innovative marker invented by the MIT Media Lab (Mohan *et al.*, 2009), which is circular and has a diameter of just 3 mm. It can store a large amount of data, is readable from 4 m with a normal camera, and works by exploiting the Bokeh effect, which occurs when the camera is out of focus. The one in Fig. 3e is an active Bokode which uses a red LED as the light source, but the MIT Media Lab has also created a passive Bokode which can work without an LED or any power source (a simple reflector is used). The analysis of the visual markers' state of the art leads us to restrict the choice of the best one that fits our requirements (first of which is the detection speed) to three possible candidates, which are shown in Fig. 4.

We have chosen these markers also because they can be freely used through opensource and well-documented libraries (AprilTag and ArUco) or free SDKs (Vuforia) and they are portable to all the major platforms. In the following, we give an overview of their features, strengths, and weaknesses. We tested the markers in light (Fig. 5a), medium (Fig. 5b), dark floor patterns (Fig. 5c), and in various light conditions. To facilitate the detection we added a little white border around the markers. The tests were performed with an iPhone 5S. Moreover, to face the problem of the absence during the test phase of any tool to objectively evaluate/ compare the performance of each marker in a simple and rapid way, we started the design of a prototype to simplify the benchmarks. We give an overview of this in Section 4.4.

(a) (b) (c)

Fig. 4 Vuforia frame marker (a), ArUco marker (b), and AprilTag (c)

(a) (b) (c)

Fig. 5 Light (a), medium (b), and dark (c) floor patterns

4.1 Vuforia marker

Vuforia is an augmented reality multiplatform SDK developed and maintained by Qualcomm. It is very powerful and offers to the developers a lot of functionalities such as object recognition, image recognition, and shape and text recognition. Moreover, the Vuforia SDK can detect and estimate the pose (with respect to the camera) of a special visual marker called the frame marker (Fig. 4a), which we can use for our indoor localization purpose. There are 512 frame markers, which are not generated by the application but are distributed as an archive. Each one encodes an ID (an integer between 0 and 511) on the binary pattern along the border, needs an area around it (at least twice thicker than the frame marker border), free of graphical elements, and with a good contrast with respect to the black frame. It needs to be entirely visible on the camera image to be detectable, so it has no tolerance to partial occlusions. The internal part of the marker is not used by the algorithm, so it is possible to put inside an image or a logo, which makes the marker more esthetically good-looking than others (but it is important that the internal design should use a contrasting, bright image or pattern in order not to deteriorate the performance of the detection phase). Since we cannot have access to the source code, it is impossible to go deeper on the algorithms used by the SDK. However, by analyzing the API, we can deduce that (1) it is possible to set the size of the marker in the scene, and (2) there are three settings related to the performance of marker detection/decoding:

1. Mode-Optimize-Speed: This option provides a lower resolution (often 640×480 pixels, depending on the device) to achieve a higher frame rate and faster detection.

2. Mode-Optimize-Quality: This option provides a significantly higher resolution but a lower frame rate and a slower detection.

3. Mode-Default: This option is typically equivalent to Mode-Optimize-Speed.

We performed some detection/decoding tests for markers of different sizes (6.5 cm×6.5 cm, 5.0 cm×5.0 cm, 3.2 cm×3.2 cm) and different distances between the marker and the camera (80, 100, and 120 cm), in movement and with the smartphone in the palm of the hand, by setting the Mode-Optimize-Quality option. We repeated the tests in several lighting conditions. Our tests show that a marker size of 6.5 cm×6.5 cm or 5.0 cm×5.0 cm gives good real-time performance for light (Fig. 5a), medium (Fig. 5b), and dark (Fig. 5c) floor patterns in good and average light conditions. The performance gradually gets worse for poor light conditions, and if the size of the marker is reduced to 3.2 cm×3.2 cm (while the distance between the marker and the camera is increased), as shown in Table 1. In summary, despite the quite good overall performance and the fact that the SDK is well maintained by a big company such as Qualcomm, the system has some drawbacks: (1) the source code is not accessible, so it is impossible to modify the algorithms in order to exploit the features of the floor, (2) the number of markers is fixed, which leads to a low flexibility, and (3) it is not possible to much reduce the size of the marker.

Table 1 Qualitative evaluation of Vuforia marker performance

Marker size (cm×cm)	Marker–camera distance (cm)	Light condition*		
		Good	Average	Poor
6.5×6.5	80	+	+	−
	100	+	+	−
	120	+	+	−
5.0×5.0	80	+	+	−
	100	+	+	×
	120	+	+	×
3.2×3.2	80	+	−	×
	100	−	−	×
	120	−	×	×

* '+' indicates that the marker is always decoded, '−' indicates that sometimes the marker is not decoded, and '×' indicates that the marker is never decoded

4.2 ArUco marker

ArUco is a square visual marker realized by the AVA group from the University of Cordoba (Garrido-Jurado *et al.*, 2014). It can be decoded through the ArUco library, which is cross-platform (because it is openCV-based), and opensource (Berkeley software distribution (BSD) license). The library is written in C++, but it has a Java version and a Python version available. Moreover, it seems to be well maintained by the research group (last update was on July 29, 2016). Different from other similar systems, ArUco does not provide a predefined set of markers: it is possible to generate the desired number of markers, with the desired number of bits (n) encoded inside each of them. The library maximizes the inter-marker distance (to avoid the situation in which a few erroneous bits in the detection lead to a wrong, but valid, marker) and the number of bit transitions (so there is a smaller probability of confusing the marker with objects inside environments), and, based on the dictionary of generated markers, proposes an error correction algorithm that allowed by the corrections of a number of errors greater than the current state of the art. It is also possible to estimate the pose of the marker with respect to the camera. To be detectable, an ArUco marker must be entirely visible on the camera image, but it is possible to manage the occlusion by using the ArUco markers board. Since ArUco does not have a fixed number of bits, the performance of the detection/decoding algorithm varies depending on this parameter, which can be set according to the requirements of our use case: small areas can be covered with few markers, which means that the n can be reduced, which in turn leads to a faster detection/decoding phase. As before, for Vuforia, we performed some detection/decoding tests for different sizes of markers (but the same size of Vuforia markers) and different distances between the marker and the camera, under the same conditions. We chose to generate 512 AruCo markers, with $n = 4$. We also set the capture resolution to 640×480, and the focus mode to an optimal value. We repeated the tests in several lighting conditions, and in three types of floor: light (Fig. 5a), medium (Fig. 5b), and dark (Fig. 5c). Our tests show that ArUco works very well in any light condition for a marker size of 6.5 cm×6.5 cm or 5.0 cm×5.0 cm with a distance between the marker and the camera of 80, 100, and 120 cm with any type of floor pattern. The performance is a little worse (Table 2) (but better than that of Vuforia) if we reduce the size of the marker to 3.2 cm×3.2 cm or increase the distance between the marker and the camera to 120 cm, for average and poor light conditions. ArUco source code is accessible for the developer, so it is possible to modify the algorithms in order to adapt them to the scenario described in Section 3. Also, the fact that the number of markers and bits can be set as different values considerably increases the high flexibility of the system. In conclusion, ArUco is a good choice for an indoor localization system with visual markers deployed on the floor, when the requirements are high flexibility and good real-time performance.

4.3 AprilTag

AprilTag is a square visual marker developed for robotic applications by Olson (2011), in the APRIL Robotic Laboratory at the University of Michigan. The opensource library allows the detection of an AprilTag in an image, the

Table 2 Qualitative evaluation of ArUco marker performance

Marker size (cm×cm)	Marker–camera distance (cm)	Light condition*		
		Good	Average	Poor
6.5×6.5	80	+	+	+
	100	+	+	+
	120	+	+	+
5.0×5.0	80	+	+	+
	100	+	+	+
	120	+	+	+
3.2×3.2	80	+	+	−
	100	+	+	−
	120	+	−	×

*'+' indicates that the marker is always decoded, '−' indicates that sometimes the marker is not decoded, and '×' indicates that the marker is never decoded

decoding of the ID of the marker, and the estimation of its 3D pose and orientation with respect to the camera. The library is written in pure C with no external dependencies, and appears to be well-documented and well-maintained (last version update was on Mar. 18, 2015), robust to changes in the light condition and view angles, and with good real-time performance. We performed some detection/decoding tests by choosing the recommended pre-generated markers family 36h11 (36 bit markers with the minimum hamming distance between codes of 11), which consists of 518 markers, and using the same marker size, marker–camera distances, and conditions as in the previous Vuforia and ArUco cases. For the tests, we used the AprilTag iOS application, developed by Olson (2011) and available on the US Apple Store for free. The application allows the user to set some parameters:

1. Decimation (1–4): It allows the reduction of the resolution of the analyzed image.

2. Refined Tag Positions (On/Off): If it is set to On, the algorithm spends more time trying to precisely localize tags.

3. Refined Tag Decodes (On/Off): If it is set to On, the algorithm spends more time trying to decode tags.

4. Camera Focus (from 0 to 1): It allows one to arbitrarily set the focus to a given value.

Since the most important requirement for our scenario is the detection/decoding speed, we set both Refined Tag Positions and Refined Tag Decodes options to Off, the Camera Focus to the optimal value for our scenario, and Decimation to the maximum value. The results show that AprilTag works very well in all tested light conditions, for almost all tested sizes and marker–camera distances, and on any type of floor, as shown in Table 3. We also analyzed the library performance with different orientations of the marker with respect to the smartphone's camera. Experimental results from AprilTag's authors show that, if ϕ is the angle between the marker's normal vector and the vector to the camera ($\phi = 0°$ means that the marker is parallel to the camera, and $\phi = 90°$ means that the marker is out of view of the camera), the library is able to detect markers for a large range of ϕ, approximately from 0° to 80°, as verified by Olson (2011) who used a ray tracer to generate images with a known ground truth. These results were confirmed by our qualitative tests performed with markers deployed on the floor in different light conditions and for different marker sizes and distances between the marker and the camera, by varying ϕ from 0° to 90° with a step of 15°. As shown in Table 4, if the library is able to detect the marker for $\phi = 0°$, then it is almost always able to detect such a marker for a ϕ which ranges from 0° to 75°. Such tests were not performed for a marker size of 3.2 cm×3.2 cm because in this case, sometimes (i.e., in poor light conditions) the marker is not decoded.

Table 3 Qualitative evaluation of AprilTag performance

Marker size (cm×cm)	Marker–camera distance (cm)	Light condition*		
		Good	Average	Poor
6.5×6.5	80	+	+	+
	100	+	+	+
	120	+	+	+
5.0×5.0	80	+	+	+
	100	+	+	+
	120	+	+	+
3.2×3.2	80	+	+	+
	100	+	+	+
	120	+	+	−

*'+' indicates that the marker is always decoded, and '−' indicates that sometimes the marker is not decoded

Finally, we faced the problem of false positives (marker detected but decoded wrongly). According to our tests, the false positive rate for our situation of deployment is really low, which makes the whole system very robust to such a kind of error. Moreover, in a hybrid system which combines different localization technologies, these errors can be avoided, simply by considering that if the system knows the area where the user is located, the decoded marker may belong to a small subset of markers (i.e., all the markers deployed on such an area): the probability that the marker is wrongly decoded and concurrently belongs to this subset is close to zero.

Table 4 Qualitative evaluation of AprilTag performance for various ϕ's

ϕ (degree)	Light condition*		
	Good	Average	Poor
0	+	+	+
15	+	+	+
30	+	+	+
45	+	+	+
60	+	+	+
75	+	−	×
90	×	×	×

* Marker–camera distance: 80, 100, and 120 cm; marker size: 6.5 cm×6.5 cm and 5.0 cm×5.0 cm. '+' indicates that the marker is always decoded, '−' indicates that sometimes the marker is not decoded, and '×' indicates that the marker is never decoded

In conclusion, the availability of the source code (which allows the developer to modify the algorithms in order to adapt them to the floor features), the speed of the system, the robustness to errors, and the small marker size make AprilTag the best choice for an indoor, marker-based localization system when flexibility on the number of markers is not required.

4.4 Benchmarking tool concept

In our analysis we need to measure how Vuforia, ArUco, and AprilTag libraries respond to the variation of parameters such as the lighting condition, floor pattern, blurring levels, and marker–camera distances. Moreover, it is necessary to perform all these measurements in movement, with the device on the palm of the hand. To speed up our future work on visual markers, and make it more accurate, objective, and repeatable, we began the design phase of a simulation software tool that, through a visual interface, and with the device (i.e., its camera) in a fixed position, will make it possible to:

1. simulate the variation of the environmental testing parameters (luminosity and blurring level),
2. simulate the variation of the parameters related to the marker (size of the marker, marker–camera distance, and background—in terms of color and pattern—where the marker is placed), and
3. simulate marker detection/decoding with the device in movement.

The development of this tool (the concept is shown in Fig. 6) will lead to a significant reduction of the benchmarking time, and we think it will help many researchers and developers choose the marker that best fits their requirements.

Fig. 6 Benchmarking tool rendering

5 Conclusions and future work

In this paper, we addressed the problem of choosing the best marker for an indoor navigation system with visual markers deployed on the floor. While focusing to some extent on marker-based computer vision approaches, we analyzed the particular use case of markers deployed on the floor. The analysis led us to choose three visual markers which have features and performance that match our scenario: Vuforia marker, ArUco marker, and AprilTag. Among them, AprilTag and ArUco have very good overall real-time performance in any tested light condition and floor pattern, for all tested marker sizes. They are also opensource and cross-platform. While AprilTag seems to be a little quicker than ArUco and allows more reduction in the marker size than ArUco (while preserving overall performance), ArUco gives better flexibility because it allows the generation of the exact number of markers we require, with the exact desired number of bits. We are planning to realize a proof of the concept of our indoor localization system using both ArUco and AprilTag, to test the approach better in real situations with both markers, and to exploit the features of the floor. The goal is to reduce the size of the marker (so the system can be less invasive) and enhance the speed. Moreover, we are investigating the possibility of mixing this technique with dead reckoning, to track the user between markers, and with BLE, to (1) perform a raw background localization and (2) segment large areas into smaller areas in order to reuse the same set of markers. Also, we are working on the realization of a set of tools for rapid and accurate benchmarking.

References

Aider, O.A., Hoppenot, P., Colle, E., 2005. A model-based method for indoor mobile robot localization using monocular vision and straight-line correspondences. *Robot. Auton. Syst.*, **52**(2):229-246.

Arias, S., April, S., 2011. Visual Tag Recognition for Indoor Positioning. MS Thesis, Universitat Politècnica de Catalunya, Catalonia, Spain.

Bajo, J., de Paz, J.F., Villarrubia, G., *et al.*, 2015. Self-organizing architecture for information fusion in distributed sensor networks. *Int. J. Distrib. Sens. Netw.*, **11**(3):1-13. http://dx.doi.org/10.1155/2015/231073

Beauregard, S., Haas, H., 2006. Pedestrian dead reckoning: a basis for personal positioning. Proc. 3rd Workshop on Positioning, Navigation and Communication, p.27-35. http://dx.doi.org/10.1186/1687-6180-2014-65

Bitsch Link, J.A., Gerdsmeier, F., Smith, P., *et al.*, 2012. Indoor navigation on wheels (and on foot) using smartphones. Proc. Int. Conf. on Indoor Positioning and Indoor Navigation, p.1-10. http://dx.doi.org/10.1109/IPIN.2012.6418931

Buchman, A., Lung, C., 2013. Received signal strength based room level accuracy indoor localisation method. IEEE Int. Conf. on Cognitive Infocommunications, p.103-108. http://dx.doi.org/10.1109/CogInfoCom.2013.6719222

Chandgadkar, A., Knottenbelt, W., 2013 An Indoor Navigation System for Smartphones. MS Thesis, Imperial College London, London, UK.

Constandache, I., Choudhury, R.R., Rhee, I., 2010. Towards mobile phone localization without war-driving. Proc. IEEE Int. Conf. on Computer Communications, p.1-9. http://dx.doi.org/10.1109/INFCOM.2010.5462058

Danakis, C., Afgani, M., Povey, G., *et al.*, 2012. Using a CMOS camera sensor for visible light communication. IEEE Global Communications Conf., p.1244-1248. http://dx.doi.org/10.1109/GLOCOMW.2012.6477759

Denso, W., 2010. QR-Code Standard. Available from http://www.denso-wave.com/qrcode/qrstandard-e.html

Ecklbauer, B.L., 2014. A Mobile Positioning System for Android Based on Visual Markers. PhD Thesis, University of North Texas, Hagenberg, Austria.

Fuchs, C., Aschenbruck, N., Martini, P., *et al.*, 2011. Indoor tracking for mission critical scenarios: a survey. *Pervas. Mob. Comput.*, **7**(1):1-15. http://dx.doi.org/10.1016/j.pmcj.2010.07.001

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., *et al.*, 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Patt. Recogn.*, **47**(6):2280-2292. http://dx.doi.org/10.1016/j.patcog.2014.01.005

Han, D., Jung, S., Lee, M., *et al.*, 2014. Building a practical Wi-Fi-based indoor navigation system. *IEEE Pervas. Comput.*, **13**(2):72-79. http://dx.doi.org/10.1109/MPRV.2014.24

Haverinen, J., Kemppainen, A., 2009. A global self-localization technique utilizing local anomalies of the ambient magnetic field. Int. Conf. on Robotics and Automation, p.3142-3147. http://dx.doi.org/10.1109/ROBOT.2009.5152885

Jovicic, A., Li, J., Richardson, T., 2013. Visible light communication: opportunities, challenges and the path to market. *IEEE Commun. Mag.*, **51**(12):26-32. http://dx.doi.org/10.1109/MCOM.2013.6685754

Kato, H., Billinghurst, M., 1999. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. Proc. 2nd IEEE ACM Int. Workshop on Augmented Reality, p.85-94. http://dx.doi.org/10.1109/IWAR.1999.803809

La Delfa, G.C., Catania, V., 2014. Accurate indoor navigation using smartphone, bluetooth low energy and visual tags. Proc. 2nd Conf. on Mobile and Information Technologies in Medicine, p.1-4.

La Delfa, G.C., Catania, V., Monteleone, S., *et al.*, 2015. Computer vision based indoor navigation: a visual markers evaluation. 6th Int. Symp. on Ambient Intelligence-Software and Applications, p.165-173. http://dx.doi.org/10.1007/978-3-319-19695-4_17

Li, F., Zhao, C., Ding, G., *et al.*, 2012. A reliable and accurate indoor localization method using phone inertial sensors. Proc. ACM Conf. on Ubiquitous Computing, p.421-430. http://dx.doi.org/10.1145/2370216.2370280

Liu, Y., Wang, Q., Liu, J., *et al.*, 2012. MCMC-based indoor localization with a smart phone and sparse WiFi access points. IEEE Int. Conf. on Pervasive Computing and Communications Workshops, p.247-252. http://dx.doi.org/10.1109/PerComW.2012.6197488

Liu, Y., Dashti, M., Zhang, J., 2013. Indoor localization on mobile phone platforms using embedded inertial sensors. 10th Workshop on Positioning Navigation and Communication, p.1-5. http://dx.doi.org/10.1109/WPNC.2013.6533266

Longacre, A., Hussey, R., 1997. Two Dimensional Data Encoding Structure and Symbology for Use with Optical Readers. US Patent 5 591 956.

Mandal, A., Lopes, C.V., Givargis, T., *et al.*, 2005. Beep: 3D indoor positioning using audible sound. IEEE 2nd Consumer Communications and Networking Conf., p.348-353. http://dx.doi.org/10.1109/CCNC.2005.1405195

Martin, P., Ho, B.J., Grupen, N., *et al.*, 2014. An iBeacon primer for indoor localization: demo abstract. Proc. 1st ACM Conf. on Embedded Systems for Energy-Efficient Buildings, p.190-191. http://dx.doi.org/10.1145/2674061.2675028

Mautz, R., 2012. Indoor Positioning Technologies. Südwestdeutscher Verlag für Hochschulschriften.

Meingast, M., Geyer, C., Sastry, S., 2005. Geometric models of rolling-shutter cameras. Computer Vision and Pattern Recognition, ePrint Archive, arXiv:cs/0503076. Available from http://arxiv.org/abs/cs/0503076

Mohan, A., Woo, G., Hiura, S., *et al.*, 2009. Bokode: imperceptible visual tags for camera based interaction from a distance. *ACM Trans. Graph.*, **28**(3):98.1-98.8. http://dx.doi.org/10.1145/1576246.1531404

Mulloni, A., Wagner, D., Barakonyi, I., *et al.*, 2009. Indoor positioning and navigation with camera phones. *IEEE Pervas. Comput.*, **8**(2):22-31. http://dx.doi.org/10.1109/MPRV.2009.30

Naimark, L., Foxlin, E., 2002. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. Proc. 1st Int. Symp. on Mixed and Augmented Reality, p.27-36. http://dx.doi.org/10.1109/ISMAR.2002.1115065

Olson, E., 2011. AprilTag: a robust and flexible visual fiducial system. Proc. IEEE Int. Conf. on Robotics and Automation, p.3400-3407. http://dx.doi.org/10.1109/ICRA.2011.5979561

Qualcomm, 2014. Qualcomm Vuforia. Available from https://developer.vuforia.com/

Richardson, A., Strom, J., Olson, E., 2013. AprilCal: assisted and repeatable camera calibration. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, p.1814-1821. http://dx.doi.org/10.1109/IROS.2013.6696595

Saito, S., Hiyama, A., Tanikawa, T., *et al.*, 2007. Indoor marker-based localization using coded seamless pattern for interior decoration. IEEE Virtual Reality Conf., p.67-74. http://dx.doi.org/10.1109/VR.2007.352465

Subbu, P., Sasidhar, K., 2011. Indoor Localization Using Magnetic Fields. PhD Thesis, University of North Texas, Texas, USA.

Tarzia, S.P., Dinda, P.A., Dick, R.P., *et al.*, 2011. Indoor localization without infrastructure using the acoustic background spectrum. Proc. 9th Int. Conf. on Mobile Systems, Applications, and Services, p.155-168. http://dx.doi.org/10.1145/1999995.2000011

Torres-Solis, J., Falk, T.H., Chau, T., 2010. A Review of Indoor Localization Technologies: Towards Navigational Assistance for Topographical Disorientation. *In*: Molina, F.J.V. (Ed.), Ambient Intelligence. InTech Open Access Publisher, Rijeka, Croatia, p.51-83. http://dx.doi.org/doi:10.5772/8678

Villarrubia, G., Bajo, J., de Paz, J.F., *et al.*, 2014. Monitoring and detection platform to prevent anomalous situations in home care. *Sensor*, **14**(6):9900-9921. http://dx.doi.org/10.1145/1999995.2000011

Wang, H., Sen, S., Elgohary, A., *et al.*, 2012. No need to war-drive: unsupervised indoor localization. Proc. 10th Int. Conf. on Mobile Systems, Applications, and Services, p.197-210. http://dx.doi.org/10.1145/2307636.2307655

Wicker, S.B., Bhargava, V.K., 1994. Reed-Solomon Codes and Their Applications. IEEE Press, Piscataway, NJ, USA.

Zachariah, D., Jansson, M., 2012. Fusing visual tags and inertial information for indoor navigation. IEEE/ION Position Location and Navigation Symp., p.535-540. http://dx.doi.org/10.1109/PLANS.2012.6236924