

# Computer Vision Based Indoor Navigation: A Visual Markers Evaluation

Gaetano C. La Delfa, V. Catania, S. Monteleone,  
Juan F. De Paz and J. Bajo

**Abstract** The massive diffusion of smartphones and the exponential rise of location based services (LBS) have made the problem of localization and navigation inside buildings one of the most important technological challenges of the last years. Indoor positioning systems have a huge market in the retail sector and contextual advertising; moreover, they can be fundamental to increase the quality of life for the citizens. Various approaches have been proposed in scientific literature. Recently, thanks to the high performances of the smartphones' cameras, marker-less and marked-based computer vision approaches have been investigated. In a previous paper, we proposed a technique for indoor navigation using both Bluetooth Low Energy (BLE) and a 2D visual markers system deployed into the floor. In this paper, we present a qualitative performance evaluation of three 2D visual markers suitable for real-time applications.

**Keywords** Indoor location systems • Computer vision • Fiducial markers

---

G.C. La Delfa (✉) · V. Catania · S. Monteleone  
Department of Electrical, Electronics and Computer Engineering (DIEEI),  
University of Catania, Catania, Italy  
e-mail: gaetano.ladelfa@dieei.unict.it

J.F. De Paz  
BISITE Research Group, Faculty of Science, University of Salamanca,  
Plaza de la Merced S/n, Salamanca, Spain  
e-mail: fcofds@usal.es

J. Bajo  
Department of Artificial Intelligence, Polytechnic University of Madrid, Madrid, Spain  
e-mail: jbajo@fi.upm.es

V. Catania  
e-mail: vincenzo.catania@dieei.unict.it

S. Monteleone  
e-mail: salvatore.monteleone@dieei.unict.it

## 1 Introduction

The massive diffusion of the smartphone has contributed in the last years to create the conditions for a significant technological progress in the mobile consumer sector giving to developers and startups the perfect instrument to create innovative applications and services. Most of these applications and services are strictly related to the user position and context information: they are defined as Location Based Services (LBS) and are becoming very popular in the last years. Outdoor the GPS is almost a standard de facto for positioning and navigation, but in indoor environments does not actually exist a unique technology to solve the problem. Various approaches have been proposed in literature to address the challenge in a simple and scalable way, and also a lot of commercial solutions are appearing into the market. Among these, the most successful ones are those which take advantage of the hardware/sensors of the smartphone to extract contextual information and use them to localize the user. *Dead Reckoning systems* use accelerometer, magnetometer and gyroscope sensors embed into the smartphones to provide fast estimation of the user position [10]. *Received Signal Strength Indication systems* exploit the RSSI of the radio signals present in the environment (typically, WiFi signals, available for free in public buildings, or, recently, BLE signals) [6]. *Visible Light Communication systems* exploit the susceptibility of LEDs to the amplitude modulation at high frequencies to transmit information into the environment and perform accurate indoor positioning (without deteriorating the lighting functionality) [2, 8]. Recently, thanks to the high performances cameras and high computational capabilities of last generation smartphones, researchers are focusing on *Computer Vision systems* which rely on complex, (1) marker-less or (2) marker-based computer vision algorithms to determine the position of the user in the environment [1]. Usually, *Hybrid* techniques and technologies are used to improve the accuracy, reduce costs and enhance the performances of the whole indoor positioning system [14]. The remainder of this paper is organized as follows: in Sect. 2, we give an overview about the state of the art of the indoor localization approaches. In Sect. 3, we discuss about the features needed to build an efficient indoor localization and navigation system with 2D visual markers deployed onto the floor. Section 4 focus on analyzing three markers with respect to some parameters which are of interest for our use case. Finally we conclude with some considerations and ideas for future works.

## 2 Related Works

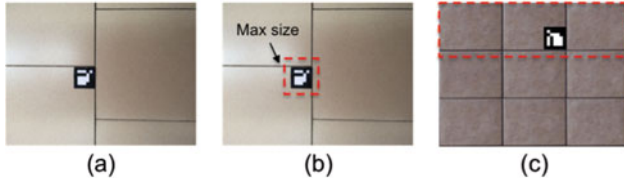
Nowadays, indoor navigation is a very hot topic and a lot of research has been made during the last decade. Researchers from Duke University proposed UNLOC [14] in which they merge environment sensing and dead reckoning (D.R.) to realize an indoor navigation system, based on the hypothesis that certain locations in indoor

environments presents - in the sensors domains - identifiable signatures (Landmarks) generated by elevators, escalators, WiFi, etc. They use D.R. to track the user, and periodically reset the error when the user encounters a landmark. In [8], the authors suggest the use of LEDs and Visible Light Communications (VLC) to localize the user inside an environment in an accurate way. On the Tx side, the modification to the LED lighting infrastructure is cheap and simple, on the Rx side, Harald Haas [2] et al. show that it is possible to exploit the rolling shutter effect of CMOS-based cameras to let a mobile phone to decode the information transmitted by the LEDs infrastructure. Apple and Google included API for indoor positioning in their SDKs for iOS and Android. They use mixed technologies such as WiFi fingerprints, BLE, iBeacons and D.R. to perform the indoor localization task. The list of approaches, techniques and technologies is actually very long. 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 purposes. Marker-less approaches are used when visual markers are undesirable due to aesthetic reasons; they rely on what the camera sees to deduce the position of the user and usually require a pre-knowledge of the environment. They are CPU-Intensive, and also require a considerable workload before they can start to work and frequent recalibrations. Marker-based approaches rely on 2D visual markers, which can be easily decoded even by a low-cost smartphone. B.L. Ecklbauer, in his thesis [3] uses 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. In the report "An Indoor Navigation System For Smartphones" [1] the authors propose 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. Despite of the simplicity and scalability of all these techniques, there are some drawbacks such as (a) the need of a line of sight, (b) the sensitivity to light changes, (c) the size of the marker which must be as smaller as possible in order to be minimally invasive, (d) the fact that the app does not work in real-time, and (e) the cognitive workload for the user who has to look for the marker in order to auto-locate himself. In order to face all these drawbacks, in our previous paper [4] we proposed a hybrid approach which uses BLE for locating the user when there is not a line of sight and a 2D visual markers system deployed onto the floor (in order to let the user to auto-locate himself without any cognitive workload: in fact when he launches the app in navigation mode, the camera is in the palm of his hand and will be directed towards some part of the floor) to estimate the position with a good level of accuracy. To guarantee accuracy, minimal invasivity and real-time performances, it is important to choose the right marker according to the particular situation of deploy. In the previous paper we proposed the use of Aruco marker. In the present paper, we evaluate in a qualitative and empirical way the performances of three 2D visual markers: (a) Vuforia frame marker [7], ArUco marker [5], AprilTag [13].

### 3 Visual Markers Deployed onto the Floor: Requirements

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

- Almost uniform, prior-known background pattern of the floor (Fig. 1a). It is possible to use this feature to improve the speed of the decoding algorithm and to reduce the marker size.
- Almost fixed, prior-known size of the marker inside the frame (Fig. 1b). It depends on the distance between the camera - which is on the palm of the hand - and the floor, and makes easier and faster to find the marker in the frame, which brings to a detection speed improvement.
- Major probability for the marker to be in the upper part of the frame (Fig. 1c), due to the fact that when the user launches the app to navigate, he 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 in order to enhance the quality of the image.
- Prior-known markers positions, so each decoded marker must be one in the boundary of the previously decoded marker.



**Fig. 1** (a) Uniform background pattern of the floor. (b) Max size of the marker inside the frame. (c) Major probability for the marker to be in the upper part of the frame

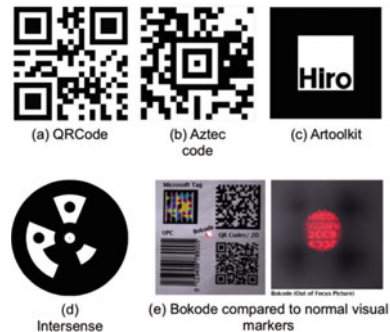
The characteristics that the chosen marker must have, considering that when the user use the system he is moving (usually with low speed), are:

1. *Small size*: it is required to reduce the invasivity of the system. We have to find the best compromise between 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 speed must be as fast as possible.
3. *Robustness to changes in light conditions*: it is required because typically the marker will be deployed in high dynamic environments characterized by the presence of other people, on/off switching of lights, shadows etc.
4. *Robustness in detecting blurred markers*: caused by too fast movement.

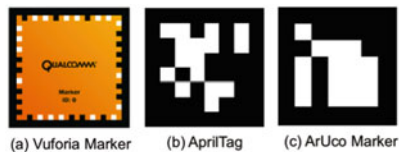
### 4 Real Time Visual Markers: Vuforia, ArUco, AprilTag

A visual marker system is composed by 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 there are several visual markers in the market. Figure 2a shows the QR-Code: it can store up to 4296 characters, and contains a Reed-Solomon error correction algorithm which let to decode even partially occluded QR-Codes. Moreover, it is opensource and very well-documented but it has not real-time performances which make it not good for our purposes. Aztec code (Fig. 2b) 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 performances, usually a visual marker which stores just a simple binary code is used. An example (Fig. 2c) is ArtoolKit marker [9]. Originally developed 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 have also circular markers, stronger to perspective distortion and more precise, such as Intersense [12] (Fig. 2d). Figure 2e shows a marker invented by the MIT Media Lab [11], circular and with a diameter of just 3 mm. It can store a large amount of data, it is readable from 4 meters with a normal camera and it works by exploiting the Bokeh effect which occurs when the camera is out-of-focus. The analysis of the state of the art bring us to restrict the choice of the best marker that fits our requirements to three possible candidates (Fig. 3). We have chosen these markers also because they can be freely used through opensource, well-documented libraries or free SDKs and they are portable in all the major platforms. In the following, we give an overview of their features, strengths and weaknesses. We test the markers (with an iPhone 5S) in light, medium and dark floor pattern, in various light conditions.

**Fig. 2** Visual markers examples



**Fig. 3** Vuforia frame marker, ArUco marker, AprilTag



## 4.1 Vuforia Marker

Vuforia is an augmented reality multi-platform SDK developed by Qualcomm. It is very powerful and offers to the developers a lot of functionalities (objects, images, shapes and text recognition). Moreover, it can detect (and estimate the pose) of a special marker called Frame Marker (Fig. 3a), which we can use for our indoor localization purposes. There are 512 markers, distributed as an archive. Each one encodes an ID on the binary pattern along the border, and needs an area around it, free of graphical elements, with a good contrast respect to the black frame. The internal part of the marker is not used by the algorithm so it is possible to put inside an image, which makes it more good looking than other ones. We performed some detection/decoding tests for different sizes of the markers:  $(6.5 \times 6.5)$ ,  $(5.0 \times 5.0)$ ,  $(3.2 \times 3.2)$  cm<sup>2</sup> and different distances marker-camera (80, 100, 120 cm), in movement and with the smartphone in the palm of the hand. We repeated the tests in several lighting conditions. Our tests show that a marker size of  $(6.5 \times 6.5)$  and  $(5.0 \times 5.0)$  cm<sup>2</sup>, give good real-time performances for light, medium and dark floor pattern in good and average light conditions. The performances get worse for poor light conditions and if we reduce the size of the tag to  $(3.2 \times 3.2)$  cm<sup>2</sup>. Despite of the good performances 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 brings to a low flexibility, and (3) it is not possible to reduce a lot the size of the marker.

## 4.2 ArUco Marker

ArUco is a square visual marker realized by the AVA group from the University of Cordoba [5]. It can be decoded through the C++ ArUco library, which is cross platform (because openCV based), and opensource. It seems to be well-maintained by the research group. Differently 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 and the number of bit transitions and proposes an error correction algorithm which lets to correct a number of errors greater than the state of the art. It is also possible to estimate the pose of the marker with respect to the camera. Since ArUco does not have a fixed number of bits, the performances of

the detection/decoding algorithm vary 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 bring to a faster detection/decoding phase. We generated 512 ArUco markers and performed tests for the same size as the Vuforia marker tests, at the same distance marker-camera, under the same conditions. Our tests show that ArUco works very well, in any light conditions for a marker size of  $(6.5 \times 6.5)$  and  $(5.0 \times 5.0)$  cm<sup>2</sup> with a distance marker-camera of 80cm, 100 and 120 cm in any type of floor pattern. The performances are a little bit worse if we reduce the size of the marker to  $(3.2 \times 3.2)$  cm<sup>2</sup>. ArUco source code is accessible for the developer: due to this, it is possible to modify the algorithms in order to adapt them to the scenario described in Sect. 3. Also the possibility to set the number of markers and bits increases a lot the flexibility of the system. In conclusion, ArUco is a good choice for an indoor localization system with visual markers deployed onto the floor, when the requirements are flexibility and real-time performances.

### 4.3 AprilTag

AprilTag is a square visual marker developed for robotic applications by Edwin Olson, at University of Michigan [13]. The opensource library lets to detect an AprilTag in an image, decode the ID of the marker, and estimate its 3D pose and orientation respect to the camera. The library is written in pure C with no external dependencies, and appears well-documented and well-maintained, robust to changes in light conditions and view angle, and with real-time performances. We performed some detection/decoding tests by choosing the recommended pre-generated markers family 36h11 (36 bit markers with minimum hamming distance between codes of 11), which consists of 518 different markers, and by using the same marker sizes, marker-camera distances and conditions of the previous Vuforia and ArUco cases. The results show that AprilTag works very well in all tested light conditions and for all tested sizes and marker-camera distances. The availability of the source code (which lets the developer to modify the algorithms to adapt them to the floor features), the speed of the system and the small marker size make AprilTag the best choice for an indoor, marker-based localization system when the flexibility about the number of markers is not required.

## 5 Conclusions and Future Works

In this paper, we have addressed the problem of choosing the best marker for an indoor navigation system with visual markers deployed onto the floor. We started with an overview on the state of the art, focusing on a marker-based computer vision approaches. Next we analyzed the particular use case of markers deployed onto the

floor, in order to find some features that can be exploited to improve the performances of the system. The analysis lead us to choose three visual markers which have features and performances that match with our scenario: Vuforia marker, ArUco marker and AprilTag. We tested them, in different light conditions and floor patterns, by considering three main parameters: their size, their distance from camera and the detecting/decoding speed. The results show that Vuforia has good performances if the marker size is equal or greater than  $(5.0 \times 5.0) \text{ cm}^2$ , but the drawback is that its SDK is not opensource. AprilTag and ArUco have very good overall real-time performances in any tested light conditions and floor patterns, for all tested marker sizes. They are also opensource and cross-platform. While AprilTag seems to be a little bit quicker than ArUco and lets to reduce, more than ArUco, the size of the marker (while preserving overall performances), ArUco gives more flexibility because lets to generate the exact number of marker we require. We are planning to realize a proof of concept of our indoor localization system using both ArUco and AprilTag, in order to test better the approach in a real situations with both the markers, and to exploit the features of the floor. The goal is to reduce the size of marker (so the system can be less invasive) and enhance the speed. Moreover, we are investigating the possibility to mix this technique with D.R., to track the user between markers, and BLE localization.

## References

1. A. Chandgadkar, W. Knottenbelt, *An Indoor Navigation System for Smartphones* (Imperial College, London, 2013)
2. C. Danakis, M.Z. Afgani, G. Povey, I. Underwood, H. Haas, Using a CMOS camera sensor for visible light communication, in *Workshops Proceedings of the Global Communications Conference, GLOBECOM 2012* (Anaheim, California, USA, 2012), pp. 1244–1248, <http://dx.doi.org/10.1109/GLOCOMW.2012.6477759>. Accessed 3–7 Dec 2012
3. B.L. Ecklbauer, A mobile positioning system for android based on visual markers. Ph.D. thesis, Hagenberg, Austria (2014)
4. G.C. La Delfa, V. Catania, Accurate indoor navigation using smartphone, bluetooth low energy and visual tags, in *Proceedings of the 2nd Conference on Mobile and Information Technologies in Medicine* (2014), [http://mobmed.org/download/proceedings2014/mobileMed2014\\_paper\\_6.pdf](http://mobmed.org/download/proceedings2014/mobileMed2014_paper_6.pdf)
5. S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* **47**(6), 2280–2292 (2014)
6. D. Han, S.H. Jung, M. Lee, G. Yoon, Building a practical wi-fi-based indoor navigation system. *IEEE Pervasive Comput.* **13**(2), 72–79 (2014). <http://dx.doi.org/10.1109/MPRV.2014.24>
7. Q. Incorporated, Qualcomm vuforia (2014), <https://developer.vuforia.com/>
8. A. Jovicic, J. Li, T. Richardson, Visible light communication: opportunities, challenges and the path to market. *IEEE Commun. Mag.* **51**(12), 26–32 (2013). <http://dblp.uni-trier.de/db/journals/cm/cm51.html#JovicicLR13>
9. H. Kato, M. Billinghurst, Marker tracking and hmd calibration for a video-based augmented reality conferencing system, in *IWAR '99: Proceedings of the 2Nd IEEE and ACM International Workshop on Augmented Reality* (IEEE Computer Society, Washington, DC, USA, 1999), pp. 85–94. <http://dl.acm.org/citation.cfm?id=857202.858134>



10. Y. Liu, M. Dashti, J. Zhang, Indoor localization on mobile phone platforms using embedded inertial sensors, in *WPNC* (IEEE, 2013), pp. 1–5. <http://dblp.uni-trier.de/db/conf/wpnc/wpnc2013.html#LiuDZ13>
11. A. Mohan, G. Woo, S. Hiura, Q. Smithwick, R. Raskar, Bokode: imperceptible visual tags for camera based interaction from a distance, in *SIGGRAPH '09: ACM SIGGRAPH 2009 Papers* (ACM, New York, NY, USA, 2009), pp. 98:1–98:8. <http://doi.acm.org/10.1145/1576246.1531404>
12. L. Naimark, E. Foxlin, Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker, in *ISMAR '02: Proceedings of the 1st International Symposium on Mixed and Augmented Reality* (IEEE Computer Society, Washington, DC, USA, 2002), pp. 27–36. <http://dl.acm.org/citation.cfm?id=850976.854961>
13. E. Olson, AprilTag: a robust and flexible visual fiducial system, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2011), pp. 3400–3407
14. H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, R.R. Choudhury, No need to war-drive: unsupervised indoor localization, in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys '12)*. (ACM, New York, NY, USA, 2012), pp. 197–210. <http://doi.acm.org/10.1145/2307636.2307655>