



**HAL**  
open science

## Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions

Lilian Calvet, Pierre Gurdjos, Carsten Griwodz, Simone Gasparini

► **To cite this version:**

Lilian Calvet, Pierre Gurdjos, Carsten Griwodz, Simone Gasparini. Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Jun 2016, Las Vegas, United States. pp.562 - 570, 10.1109/CVPR.2016.67 . hal-01420665

**HAL Id: hal-01420665**

**<https://hal.science/hal-01420665>**

Submitted on 20 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

# Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions

Lilian Calvet<sup>1</sup>, Pierre Gurdjos<sup>2</sup>, Carsten Griwodz<sup>1</sup>, and Simone Gasparini<sup>2</sup>

<sup>1</sup>Simula Research Laboratory, Oslo, Norway ✉{lcalvet, griff}@simula.no

<sup>2</sup>University of Toulouse, France ✉{pgurdjos, simone.gasparini}@enseeiht.fr

## Abstract

Using fiducial markers ensures reliable detection and identification of planar features in images. Fiducials are used in a wide range of applications, especially when a reliable visual reference is needed, e.g., to track the camera in cluttered or textureless environments. A marker designed for such applications must be robust to partial occlusions, varying distances and angles of view, and fast camera motions. In this paper, we present a robust, highly accurate fiducial system, whose markers consist of concentric rings, along with its theoretical foundations. Relying on projective properties, it allows to robustly localize the imaged marker and to accurately detect the position of the image of the (common) circle center. We demonstrate that our system can detect and accurately localize these circular fiducials under very challenging conditions and the experimental results reveal that it outperforms other recent fiducial systems.

## 1. Introduction

The term *fiducial marker*, or simply *fiducial*, refers to a set of (coplanar) points encoded in a planar pattern allowing a reliable detection and identification across views. A *fiducial marker system* is a (set of) fiducial marker(s) coupled with dedicated computer vision algorithms solving the detection and identification problems. This is used in a variety of applications, both in computer vision and robotics, ranging from camera calibration to augmented reality or visual SLAM. The choice of fiducials is of crucial importance within this framework as markers must provide reliable visual references in the scene that can be used to estimate, e.g., the camera position or its motion. Such framework requires that the fiducial marker system be robustly and accurately detectable even under very challenging conditions, such as, e.g., when the markers are partially or largely occluded, or seen under highly skewed angles or from long

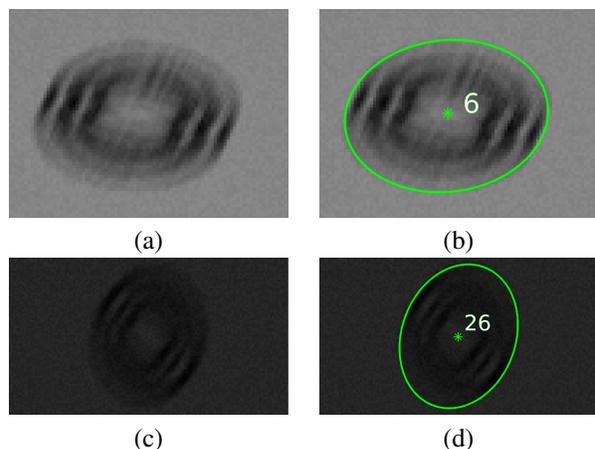


Figure 1. (a,c) Synthetic images of circular fiducials under very challenging shooting conditions *i.e.*, perturbed, in particular, by a (unidirectional) motion blur of magnitude 15px. (b,d) Using the proposed fiducial system, markers are correctly detected and identified with an accuracy of 0.54px and 0.36px resp. in (a) and (c) for the estimated imaged center of the outer ellipse whose semi-major axis (in green) is equal to 31.9px and 34.5px resp.

distances, when the illumination is very poor or irregular, or when the camera undergoes very fast motions generating blur.

In this paper, we present a robust, highly accurate and *theoretically-founded* fiducial system, which is highly tolerant to all of the mentioned challenges, as shown in Figure 1. Its markers are based on concentric black rings on a white background, extending the one-ring markers introduced by Gatrell *et al.* in [10]. The geometric properties of the concentric circles delivered by their edges are exploited to accurately detect the image of the circle common center, thus providing a highly reliable feature point that can be used for tracking and motion estimation. The thickness of the rings can be used to encode the information of the marker, typically a unique ID, thus providing a simple and reliable

method for recognizing the different markers placed in the scene.

The detection method proposed in this work relies on the flow conservation property: the ingoing amount of gradient magnitudes through an arc of the outermost ellipse must be equal to the outgoing amount of gradient magnitudes through an arc of the innermost ellipse. This property also holds in presence of motion blur since, as described in [16], the image of a circular fiducial is not affected by the blur along the line perpendicular to the direction of the motion blur. Consequently, the proposed circular fiducial can be detected and identified without explicitly un-blurring the image before the detection stage. Furthermore, since any concentric circle-pair encodes –through the circular points– the Euclidean structure of its supporting plane, the perspective distortion can be removed by estimating its rectifying homography. While reaching the performance of other recent fiducial systems under favorable conditions, we show that the proposed fiducial system clearly outperforms these systems under highly challenging (*i.e.*, more realistic) conditions.

The paper presents the system theoretical foundations and describes both the circular fiducial detection step and its subsequent validation, the latter simultaneously delivering the position of the imaged center along with the marker ID. A comparison of its performance with a variety of recent marker system under conditions of varying difficulties is finally provided.

## 2. Related work

The most widely used fiducial systems use bitonal patterns, usually made of two main components: low frequency elements for marker detection, *e.g.* a square black border [8, 12, 18], and high frequency elements for information encodage, *e.g.* an internal region filled with  $6 \times 6$  grid of black and white cells [8, 18].

Thanks to the relatively simple detection algorithms allowing fast detection at high frame rates even on mobile phones, these systems have gained popularity in many augmented reality applications, supported by many freely available libraries, such as ARToolkit [12] and ARToolkit-Plus [18].

However, this conventional approach suffers in presence of motion blur. Whereas low frequency components remain localizable to some extent, high frequency components are not preserved, which prevents the extraction of the marker identity. A trivial solution for this problem is to increase the size of the markers in the scene. However, for many applications, this can be invasive and therefore unacceptable. When processing video streams, temporal continuity can be exploited to improve the detection performance and robustness: the markers can be tracked across the frames, *e.g.* as implemented in ARToolkit [12]. However, such tracking al-

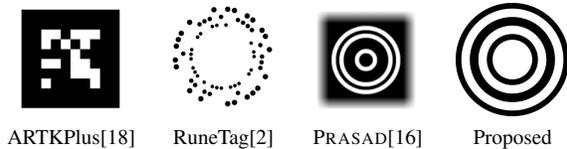


Figure 2. Prior fiducials and our proposal. From left to right, the last two are designed to be detected under motion blur conditions. Only the proposed detection algorithm is able to exactly match a feature point (its center) across a collection of images, even in presence of motion blur.

gorithms fail when the camera moves rapidly, and require re-initialization whenever the imaged marker is no longer detected.

Lately, fiducial systems that are robust to motion blur have been proposed. The *mono-spectrum marker* [17] consists of low frequency components in the form of coloured-“degraded” dots lying on a black square background. Regions of the markers are distinguished from other regions by means of a frequency spectrum specific to the marker. At the end of the localization process, the four marker corners are extracted as the centers of the dots located at each corner. A drawback of such an approach is that it is assumed that the image of a dot center corresponds to the center of the imaged dot. However, this underlying assumption is not valid under a projective camera model, thus preventing feature matching across a collection images. In practice, this can significantly affect the accuracy of the retrieved pose, because the error due to the center assumption increases with the size of the circle, and consequently the error of the computed pose grows as the camera gets closer to such a marker. This remark also holds for the methods proposed in [7] and [16]. In [7], a marker consists of four black dots, forming a square, located on a white background. In this approach, the marker detection is performed through machine learning techniques and it is to some extent capable of handling the presence of motion blur.

The fiducial markers proposed in [16] and [14] are the most similar to the one presented in this work. In [16], the authors propose a marker made up of concentric white rings on black background. However, even in this case, the image of the marker’s center is incorrectly assumed to be at the geometrical center of the marker’s image. Another drawback of this method is a very low information-coding capacity, which leads to a library size of only four different markers, thus dramatically limiting the number of possible applications. In [14], a fiducial system robust to motion blur is proposed. It relies on self-similar templates defined by a 2D rotationally invariant, bitonal intensity function. However, the self-similarity property *only* holds in theory for fronto-parallel acquisitions by calibrated cameras, even if it is empirically observed robustness to perspective distortions. While not discussed in [14], any identification task

should run the detection algorithm as many times as the number of IDs, drastically reducing the identification power and/or increasing the computational time.

The problem of strong occlusion is addressed in Rune-Tag [2], where a marker is composed of rings of circular dots. The authors have shown that such a marker allows an accurate estimation of the camera pose while being robust to severe occlusion. The PiTag fiducial markers [3], which are also made up of circular dots but arranged in rectangles, have shown a similar robustness to occlusion, although with an even smaller number of circular dots.

### 3. The Proposed Circular Fiducial System

The proposed circular fiducial is a planar pattern consisting of a set of concentric black circular rings on a white support. Every circular fiducial naturally encodes three points two of which being the circular points of its supporting plane, which are complex conjugate points at infinity [15]. The circular point-pair encodes the 2D Euclidean structure of this plane naturally, allowing to determine from its image a metric rectification that removes the projective distortion on the imaged marker. In the case of a calibrated camera, it also allows to compute the camera pose with respect to the supporting plane, up to an unknown planar rotation. Another benefit of circular fiducials is their known resilience to severe occlusions [19, 13] as well as linear motion blur [16]. In terms of their information-coding capacity, a circular fiducial can be seen as a circular bar-code which encodes data by varying the widths and spacings of circular rings. Consequently, one of its (relative) shortcomings is the limited amount of information that can be encoded in this bar-code compared to some other fiducials. This shortcoming can be overcome, for example, by reusing IDs but combining them in unique patterns. Overall, despite such an high potential for being the ‘ideal’ fiducial, the circular ring pattern has received relatively little attention in the literature. An explanation of this lack of popularity may be the absence of code released under a public license.

We have recently reported [4, 5] how to integrate circular fiducials within a unified Structure-from-Motion paradigm, where we describe how images of circular point-pairs can be combined with images of natural points.

We assume that all views capture a cluttered scene in which  $M$  circular fiducials with  $N$  rings are visible. The gray-scale intensity of one image is represented by a differentiable function  $I : [1, m] \times [1, n] \subset \mathbb{R}^2 \rightarrow [0, 1]$ , whose discretized representation is the sampling of  $I$  on the domain  $[1, m] \times [1, n] \subset \mathbb{N}^2$ . The image gradient field is denoted by  $\nabla_I : I \rightarrow \mathbb{R}^2$ . We assume to be given the set of edge points that correspond to maxima of the gradient magnitude (e.g., using Canny’s edge detector [6]).

We use the Matlab-like notation  $M_{(1:r,1:c)}$  for denoting the  $r \times c$ -submatrix of  $M$  selected by the row range  $[1, r]$  and

the column range  $[1, c]$ . Similarly  $M_{(:,1:c)}$  (resp.  $M_{(1:r,:)}$ ), selects the first  $c$  (resp.  $r$ ) columns (resp. rows) of  $M$ .

#### 3.1. Circular fiducial detection

The problem of detection is formulated as that of seeking regions in the image that can potentially support a portion of the external contour of the imaged circular fiducial, so-called *outer elliptical arc*. To be candidates, the regions must obey some loose geometric constraints, as convexity and smoothing, as well as more specific photometric constraints derived from differential properties of the gray-scale intensity, related to the gradient field  $\nabla_I$  within an imaged circular fiducial, as reported in the next paragraphs.

There are three different selection steps which can be summarized as follows.

**Pixel selection** (§3.1.2): We create paths in the image in the form of sequences of  $2N$  linked edge pixels (when a fiducial has  $N$  rings) such that the direction of the path segment starting at pixel  $p$  is given by the image gradient at  $p$ . Then, we select pixels which appear as path ending points the greatest number of times (through a vote procedure).

**Region-pair grouping**: We group into ‘inner’ regions the selected ending points to form polygonal approximations of convex arcs and we group into ‘outer’ regions the associated starting points.

**Region-pair selection** (§3.1.3): We select only region-pairs which satisfy the ‘conservative constraint’ which basically ensures that the ingoing amount of gradient magnitudes through the outer region must be equal to the outgoing amount of gradient magnitudes through the inner region.

##### 3.1.1 Theoretical foundations of detection

At this step, in order to provide the theoretical foundations of our approach in a simple form, we will only consider circular fiducials with *one* ring, delimited by its outer and inner circles. Both circles are centered at the origin with radii 1 and  $r < 1$ , and associated with closed and open disks,  $B_1[\mathbf{0}]$  and  $B_r(\mathbf{0})$  in  $\mathbb{R}^2$ . Thus,  $\Omega = B_1[\mathbf{0}] \setminus B_r(\mathbf{0})$ , defines the surface of the ring between the two circles, in the following referred to as the interior of the ring. Exceptionally, the ring is not painted black but (again for pedagogic purposes) with a continuous gradation of gray hues (black is hue 0 and white is hue 1) as defined by the function  $\alpha : [-1, 1]^2 \rightarrow [0, 1]$  we have  $\alpha(\mathbf{x}) = 1 - \|\mathbf{x}\|^2$  for all  $\mathbf{x} \in \Omega$  (interior of the ring),  $\alpha(\mathbf{x}) = 1$  for all  $\mathbf{x} \notin \Omega$ . This is seen in figure 3(b).

Consider a view of the scene under an imaging process  $H : [-1, 1]^2 \rightarrow [1, m] \times [1, n] \subset \mathbb{R}^2$  through some homographic mapping that restricts the central projection to the ring’s supporting plane [11]. Under the assumptions that (i) the scene is illuminated by a uniform parallel light beam, (ii) the surface supporting the fiducial is Lambertian, (iii) the chirality constraints are ensured, the intensity value for

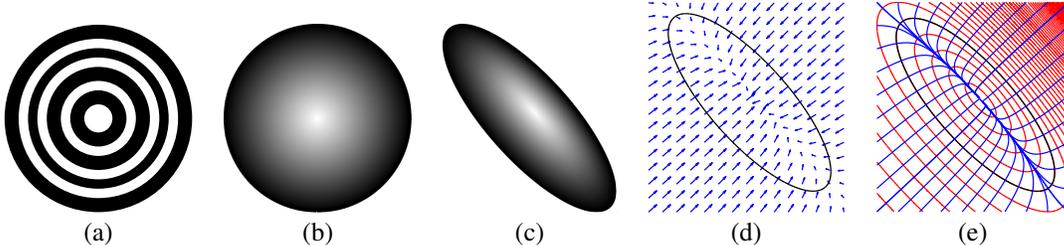


Figure 3. (a) A circular fiducial with four black rings. (b) A circular fiducial with one single ring painted with a continuous gradation of gray hues (the inner circle has a near infinitesimal radius). Note that the circles of the fiducial in (a) are equipotentials in (b). (c) An imaged pattern. (d) The gradient map of the imaged pattern. (e) The field lines of the gradient map.

an image point  $\mathbf{u} \in H([-1, 1]^2)$  is given by (see figure 3(c))

$$I(\mathbf{u}) = \alpha(H^{-1}(\mathbf{u})) \quad (1)$$

For  $\mathbf{u} \notin H([-1, 1]^2)$ ,  $I(\mathbf{u})$  is simply the luminance related to the corresponding projected scene point. Note that since the ring can be seen as an “infinite” set of concentric circles associated with the equipotentials of  $\alpha$ , their images under  $H$  are the equipotentials of (1), the outer and inner equipotentials coinciding with the images of the outer and inner circles (see figure 3(c)).

Assuming there are no edge points in  $H(\Omega)$  except a few outliers, we now shift our first problem of determining the points of the outer ellipse into the one of linking each point of the outer ellipse to some point of the inner ellipse. Our idea for achieving this is to follow –if possible– for each edge point  $\mathbf{u}_e$  the field line of the gradient field  $\nabla_I$  through  $\mathbf{u}_e$  and to stop as soon as another edge point is encountered (see figures 3(d-e)). This field line is the planar curve passing through  $\mathbf{u}_e$  and whose tangent at each of its points  $\mathbf{u}$  is collinear to the vector  $\nabla_I(\mathbf{u})$ , or more formally, the curve has parameterization  $\phi : U \subset \mathbb{R} \mapsto \mathbb{R}^2$ , solution of the differential system 
$$\begin{cases} \phi'(t) = \nabla_I(\phi(t)) \\ \phi(t_0) = \mathbf{u}_0 \end{cases}$$
 where  $(t_0, \mathbf{u}_0) \in \Omega \times \mathbb{R}^2$  defines an initial condition.

Two key results are now given (cf. figure 3(e)):

**Proposition 1.** *In the continuous image,*

- *Through any point in  $H(\Omega)$ , except the image of the circle centre, passes one and only one field line of  $\nabla_I$ ;*
- *Any field line necessarily converges to one point on the inner ellipse.*

The proofs are omitted here due to lack of space (the proof of the latter result relies on the fact that, when the inner radius of the ring is 0, any field line converges to the image of the circle center).

### 3.1.2 Pixel selection

Here, circular fiducials can have  $N$  rings, that is  $2N$  concentric circles. The idea is to choose an adequate number  $N$

of rings in order to guarantee a good approximation of the field lines in the discrete image.

We call *linked sequence* (or *path*) a polygonal line with  $2N$  edge points  $\{\mathbf{u}_{i_1}, \dots, \mathbf{u}_{i_P}\}$  as vertices, when all pairs  $(\mathbf{u}_{i_j}, \mathbf{u}_{i_{j+1}})$ ,  $j \in [1, 2N - 1]$  are pairs such that  $\mathbf{u}_{i_j}$  is an edge point and  $\mathbf{u}_{i_{j+1}}$  is the closest edge point to  $\mathbf{u}_{i_j}$ , outside the neighbourhood of  $\mathbf{u}_{i_j}$ , that lies on the line passing through  $\mathbf{u}_{i_j}$  with  $(-1)^j \nabla_I(\mathbf{u}_{i_j})$  as direction. Assuming  $K$  linked sequences, for any  $k \in K$ , let  $\mathcal{S}$  be the  $K \times 2N$ -matrix where the row  $\mathcal{S}_{(k,:)}$  concatenates the indices of edge points for the sequence number  $k$  and  $V_i = \{l \in K \mid \mathcal{S}_{(l,2N)} = i\}$  is the set of indices of all linked sequences in which  $\mathbf{u}_i$  is the  $2N$ -th (i.e., last) point. Note that  $\text{card}(V_i)$  can be seen as the number of ‘votes’ for  $\mathbf{u}_i$  as last point in a sequence.

If  $E$  denotes the set of indices of all edge points, then we can state the following pixel selection rule. *The set  $F \subset E$  with  $\text{card}(F) = T$ , containing the indices of the  $T$  edge points with the  $T$  highest votes i.e., of the  $\mathbf{u}_i$  with  $i$  satisfying  $\min_{i \in F} \text{card}(V_i) \geq \max_{j \in E \setminus F} \text{card}(V_j)$ , is the set of indices of points selected as point candidates for the inner ellipses while  $\mathcal{S}(\cup_{i \in F} V_i, 1)$  yields the set of indices of points selected as point candidates for the outer ellipses.*

In practice, we have implemented this algorithm in the discrete image as follows. An edge pixel is first (arbitrarily) selected. We can see it as having a *voting intention*. It yields a *valid vote* only if it is possible to build from it a polygonal line  $\{\mathbf{u}_i\}_{i \in \mathcal{S}_{(k,:)}}$ , whose vertices are  $2N$  edge points satisfying some conditions (see below). If it is the case, the vote is confirmed. The purpose is to create polygonal lines that connect edge points of the outer ellipse to edge points of the inner ellipse. The two conditions for a polygonal line to approximate a field line arc are the following. For all  $j \in \{1, \dots, 2N - 1\}$ : (i)  $\mathbf{u}_{\mathcal{S}_{(k,j+1)}} = \mathbf{u}_{\mathcal{S}_{(k,j)}} + a_{\mathcal{S}_{(k,j)}} \nabla_I(\mathbf{u}_{\mathcal{S}_{(k,j)}})$ , where  $a_{\mathcal{S}_{(k,j)}} < 0$  if  $j$  is uneven and  $a_{\mathcal{S}_{(k,j)}} > 0$  otherwise; (ii) any segment with endpoints  $\mathbf{u}_{\mathcal{S}_{(k,j)}}$ ,  $\mathbf{u}_{\mathcal{S}_{(k,j+1)}}$  does not include edge points other than vertices. It is worth of mentioning that we empirically verified that the underlying assumption holds in practice: defining a pixel as a square whose sides have length 1, through the pixel  $\mathbf{u}_{\mathcal{S}_{(k,1)}}$  and

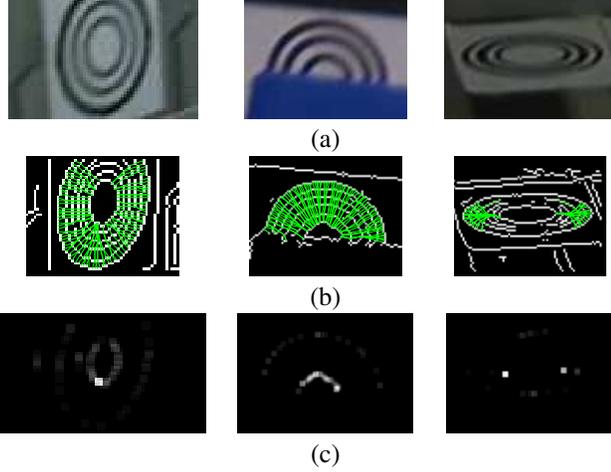


Figure 4. (a) Three images of circular fiducial. (b) Binary edge images returned by the edge detection on which has been plotted a set of linked sequences. (c) Images of resulting votes.

$\mathbf{u}_{S_{(k,2N)}}$  it passes at least one field line in the continuous image. The voting procedure applied on real images is illustrated in figure 4.

### 3.1.3 Region-pair selection

At this step, a candidate is any pair of regions  $(R_1, R_2)$  in the discrete image such that  $R_2$  is a convex chain of pixels  $q$  with a number of votes  $\text{card}(V(q)) \geq 2$ , while  $R_1$  is the set of pixels  $p$  such that  $p$  is the first pixel in the linked sequence of which  $q$  is the last pixel. The issue here is to validate  $R_2$  as a support of an arc of the inner ellipse by checking the average number of received votes.

It is assumed to have fitted all the pixels in  $R_i$  to the parameters of some elliptical arcs  $\mathcal{E}_i$  using some robust technique<sup>1</sup>, and, in particular, to have estimated  $L(\mathcal{E}_i)$ , the length of  $\mathcal{E}_i$ . Note that if any of the two regions does not ‘fit well’ with a portion of an ellipse, then the pair is eliminated. In the next paragraph, we show that *the total number card(V(R<sub>2</sub>)) of received votes by R<sub>2</sub>, i.e., in the region candidate to support an inner arc E<sub>2</sub>, must satisfy, for some threshold ε > 0,*

$$|\text{card}(V(R_2)) - L(\mathcal{E}_1)| < \epsilon \quad (2)$$

**Principle of flow conservation for validation.** For the moment, we deal again with a theoretical circular fiducial with one ring painted with a continuous gradation of gray hues, as detailed in previous paragraphs.

Let  $\mathcal{T}_1$  and  $\mathcal{T}_2$  be two line fields of the gradient map  $\nabla_I$  defined on  $H(\Omega)$  i.e., inside the imaged ring. Let  $\mathcal{A}_1$  and  $\mathcal{A}_2$  be the arcs of the outer and inner ellipses respectively, such that the curve delimited by  $\mathcal{A}_1$ ,  $\mathcal{T}_1$ ,  $\mathcal{A}_2$  and  $\mathcal{T}_2$

<sup>1</sup>The outer ellipse is estimated using the ‘ellipse growing’ technique of [13], where multiple arcs can be assembled.

is closed, as shown in figure 5. We denote by  $\mathcal{A}_i$ , with  $i \in \{1, 2\}$ , the arclength-parameterized curve modeling the arc  $\mathcal{A}_i$  with  $L(\mathcal{A}_i)$  denoting the length of  $\mathcal{A}_i$ .

The flux (or total flow) through the arc  $\mathcal{A}_i$  is defined as

$$\int_{\mathcal{A}_i} (\nabla_I \cdot \mathbf{n}_i) ds = \int_{\mathcal{A}_i} \|\nabla_I\| ds \quad (3)$$

since  $\nabla_I$  and  $\mathbf{n}_i$ , the unit normal to  $\mathcal{A}_i$ , both evaluated at  $\mathcal{A}_i(s)$ , are collinear vectors.

The proposed *conservative constraint* in our validation step relies on the principle of conservation.

$$\int_{\mathcal{A}_1} \|\nabla_I\| ds = \int_{\mathcal{A}_2} \|\nabla_I\| ds \quad (4)$$

which says that the total ingoing flow though arc  $\mathcal{A}_1$  is exactly the same than the total outgoing flow through arc  $\mathcal{A}_2$ . This follows from the fact that the gradient field is, by definition, a conservative vector field. In terms of vocabulary, we will say that  $(\mathcal{A}_1, \mathcal{A}_2)$  is a *conservative arc-pair*.

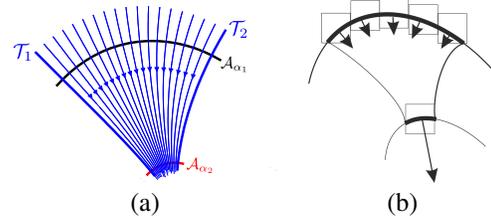


Figure 5. (a) Flow conservation in a continuous image.. (b) Flow conservation in a discrete image.

In the discrete image, under the hypothesis that  $R_1$  is a region (set of pixels) that supports an arc  $\mathcal{E}_1$  of the inner ellipse, one aims at determining whether the associated region  $R_2$  supports an arc  $\mathcal{E}_2$  of the outer ellipse such that the arc-pair  $(\mathcal{E}_1, \mathcal{E}_2)$  is a conservative arc-pair i.e., a pair like  $(\mathcal{A}_1, \mathcal{A}_2)$  which satisfies (4). This is achieved by checking the total number of received votes by  $R_2$ .

Let us deal now again with bitonal circular fiducials with multiple rings while assuming, that the principle of flow conservation holds for its image. The voting procedure creates paths (in the form of the so-called linked sequences) in the image gradient field whose endpoints are assumed to lie on a single field line. An important property of conservative vector fields is that the line integral is path-independent. Thus, the flows carried between the endpoints are the same for either any path or the line field. On the other hand, one can say that the paths created by linked sequences have the property that the incoming flow for the starting pixel of the path is 1 (a starting pixel initiates a vote). Hence, we can conclude that the outgoing flow of the inner arc is equal to  $\text{card}(V(R_2))$ , the total number of received votes by the pixels of  $R_2$ , cf. figure 5(b).

### 3.2. Circular fiducial validation

In the discrete image, let  $(R_1, R_2)$  be two regions supporting two elliptical arcs  $(\mathcal{A}_1, \mathcal{A}_2)$ , respectively outer and inner, which satisfy the necessary condition (2) of the conservative constraint. We seek the image  $\mathbf{c}$  of the circle centre (detected at sub-pixel accuracy) such that the image region delimited by the elliptical sector formed by  $\mathbf{c}$  and  $R_2$  can be validated as the image of a sector of a circular fiducial or not. For this purpose, our idea is to check the appearance similarity of ‘metric rectifications’ of different cross sections through  $\mathbf{c}$ .

We first describe the geometric component of our validation process. We call any homography  $G$  of the image onto itself a *canonical rectifying homography* such that, by applying  $G$  to the two image ellipses corresponding the projections of two concentric circles, we obtain two concentric circles centred at the origin with radius 1. In other words, we recover the original Euclidean geometry of the plane supporting the circles up to a 2D rotation and scaling. It is well-known [15] that the two image ellipses corresponding to the projections of two concentric circles (or, equivalently, one of these ellipses plus the image of the centre) are sufficient to determine of an exact solution for a canonic rectifying homography. In this work, we describe in proposition 2 a minimal parameterization of a canonical rectifying homography with only two degrees of freedom which are the coordinates of the image of the circle centre (the proof is omitted).

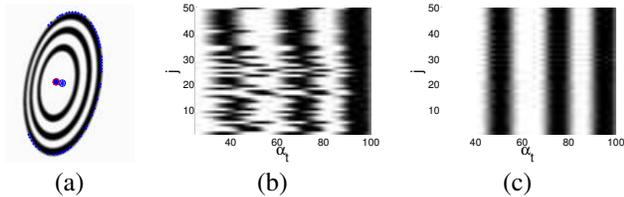


Figure 6. (a) Estimated image of the centre at each iteration (b) The signal collected on the cross sections before the optimization (c) Rectified signal after optimization.

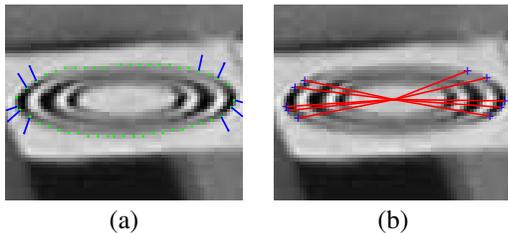


Figure 7. (a) Selected edge points with associated gradient directions. (b) Cross sections.

**Proposition 2.** Consider an adequate 2D affine representation of the image plane such that the (elliptical) image of

a circle has simplified equation  $Q_{11}u^2 + Q_{22}v^2 = 1$ . Then, for any variable point  $\mathbf{u} = (u, v)^\top$ , the matrix function<sup>2</sup>

$$\mathbf{G}(\mathbf{u}) = \left( \begin{bmatrix} -1 & Q_{22}uv & -u \\ 0 & -Q_{11}u^2 + 1 & -v \\ -Q_{11}u & Q_{22}v & 1 \end{bmatrix} \begin{bmatrix} r & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & s \end{bmatrix} \right)^{-1}$$

with  $r = \left( -\frac{Q_{22}}{Q_{11}}(Q_{11}u^2 + Q_{22}v^2 + Q_{33}) \right)^{1/2}$ , (5)

$$s = (-Q_{22}(1 - Q_{11}u^2))^{1/2}$$

defines an homography of the image plane onto itself such that  $\mathbf{G}(\mathbf{c})$  is a canonical rectifying homography, providing  $\mathbf{c} = (u_c, v_c)^\top$  represents the image of the circle centre.

To check the appearance similarity of the candidate, we proceed as follows. Let  $\mathbf{x}$  be a point located inside the inner ellipse and  $\mathbf{u}_j$  a pixel in  $R_1$ , the region supporting the outer ellipse. Let us designate by *cross section at  $\mathbf{u}_j$*  the line segment in the image joining  $\mathbf{u}_j$  and  $\mathbf{x}$ . Let  $\{\alpha_t \mathbf{m}_j \mid \alpha_t = \frac{t-1}{T-1}, t = 1..T\}$  be the discretization into  $T$  equally spaced points of the canonical metric rectification of the cross section at  $\mathbf{u}_j$ , i.e., where  $\mathbf{m}_j = g_{\mathbf{x}}(\mathbf{u}_j)$  is the image of  $\mathbf{u}_j$  by  $G(\mathbf{x})$ .

The optimization of the image of the circle centre consists in seeking a solution of :

$$\min_{\mathbf{c}} \sum_{\substack{(j_1, j_2) \in J \\ j_1 \neq j_2}} \sum_{t=1}^T \{I(g_{\mathbf{c}}^{-1}(\alpha_t \mathbf{m}_{j_1})) - I(g_{\mathbf{c}}^{-1}(\alpha_t \mathbf{m}_{j_2}))\}^2$$

(6)

where  $J$  indexes a set of sections and  $I(g_{\mathbf{c}}^{-1}(\alpha_t \mathbf{m}_{j_k}))$ ,  $k \in \{1, 2\}$ , is the intensity of the image of  $\alpha_t \mathbf{m}_{j_k}$  which is a point of a discretized rectified cross section.

This non-linear least-square optimization uses as initial solution for  $\mathbf{c}$  the centre of the outer ellipse (which is usually very close to the image of the centre). A solution is found using (few) quasi-Newton iterations. This is illustrated in figure 6. The outer ellipse is validated if the residual of optimization is below a given threshold.

## 4. Experiments

We conducted a large number of experiments with synthetic images to quantify the performance of the proposed fiducial system. We use a combination of several metrics. These are the (1) false positive rate, (2) misidentification rate and (3) false negative rate. The *false positive rate* reports the rate at which a marker is detected where none is present. The *misidentification rate* expresses the rate at which a marker is successfully detected, but misidentified as another marker with another ID. The *false negative rate* measures the rate at which no marker is detected although

<sup>2</sup> We write it as the inverse of some matrix form because we actually use  $(\mathbf{G}(\mathbf{u}))^{-1}$  in the optimization step described in the sequel..

one is present. Reducing the false negative rate increases the risk of false positives and misidentifications. Metrics (1) and (2) related to the proposed fiducial system are reported in Table 1 whereas the metric (3) is reported in Figure 8, from (a) to (c). All three metrics are influenced by the conditions under which the fiducial systems achieves a reported performance, and here in particular, the systems (i) scaling tolerance, (ii) occlusion tolerance, (iii) motion blur tolerance and (iv) depth of field blur.

Scaling tolerance is the ability to detect and identify markers at a very small or very large size in pixels. A wide scaling tolerance allows a larger usable range of cameras as well as a more opportunities for placing markers in a scene.

Figure 8 shows results for the tolerance from (i) to (iii) (Figure 8 from (a) to (c)) while Figure 8.(d) shows the obtained imaged center accuracy. The distance from the camera to the marker is used to evaluate the scaling tolerance (i) (Figure 8(a)). The percentage of occlusion represents the percentage of area of the fiducial which is not visible (*i.e.* occluded) over the total area of the fiducial (reported in Figure 8.(b)). Finally, the amount of blur can be expressed by two different quantities representing respectively the amount of blur due to depth of field and the amount of motion blur (supposed unidirectional in accordance with our assumptions)

The results delivered by the proposed solution has been compared to the ones delivered by the three following fiducial marker systems: i) ARTKPlus [12]: this system remains, yet not recent, a reference in the publicly available solution; ii) PRASAD [16]: this system is the closest to our proposal as using concentric rings in order to propose a fiducial detection resilient to motion blur; iii) RuneTag [2]: as we want to demonstrate the robustness of our system to severe occlusions, we also decided to conduct a comparison against this recent solution.

**Scene.** The following evaluation is conducted on a scene consisting of a single marker. The size of each “category” of marker in the scene is defined so that they all cover the same area of the supporting plane and that a circular marker is of unit radius. The supporting plane belongs to the  $xy$ -plane while the marker center corresponds to the scene origin. The marker center is located at  $(x_m, y_m, D)$  w.r.t. the camera coordinate system, where  $(x_m, y_m)$  is randomly distributed in  $[-0.5, 0.5]^2$  and  $D$  is the distance from the marker center to the camera optical center whose unit is the circular marker radius. If not specified, the angle between the optical axis and the supporting plane normal is randomly distributed in  $[0, 75]$  degrees.

**Camera.** The synthesized cameras obey a pinhole camera model with square pixels; the principal point is at the center of the image and the focal length is of 800 pixels while the image resolution is  $640 \times 360$  pixels.

**Image signal corruption.** As we are dealing with

binary patterns, pixels take their values in  $\{0, 255\}$ , resp. black or white. In order to vary both the contrast and the signal to noise ratio, once the original image is rendered, all its pixel values are divided by the scalar value  $c$  randomly distributed in  $[1, 6]$ , thus simulating different lighting conditions. A first blur component related to the depth of field is simulated by applying a gaussian filtering of standard deviation  $\sigma$  randomly distributed in  $[0, 2]$ . Then a unidirectional motion blur of length  $l$  (whose magnitude is 0 pixel if not specified) is applied in a random direction. Finally, a noise of  $n_{gray}$  pixels randomly distributed in  $[0, 5]$  is added on every pixel of the obtained image.

**Results.** All the results in Figure 8 are expressed in terms of *detection rate*, *i.e.*  $\tau_d = 1 - \tau_n$  where  $\tau_n$  is the false negative rate. We can see that even without any occlusion and motion blur, the proposed system already outperforms the other ones. This can be mainly justified by the following: extensive experiments have shown that our system perform quite well even in the case where the marker is strongly inclined relatively to the pixel plane. This comes mainly from the fact that there are always image cuts along which the concentric circles are “minimally distorted”, and identical information content is equally readable along any single direction. Markers that require the extraction of data from one or more dedicated directions to extract encoded information (including a classical barcode as well as 2D patterns) pay their much higher information-coding capacity with an increase in misidentifications with distortion increases. Tolerance to distortion must then be countered by mechanisms that avoid ID collisions, but these increase the Hamming distance and reduce their information-coding capacity. It is also important to mention that the RuneTag solution presents very poor performance as the solution relies on the performance of an ellipse detection algorithm used to detect the imaged circular dots whose performance drastically drop down while the size of the imaged marker is decreasing.

**Accuracy under very challenging shooting conditions.** In the last experiment, whose results are illustrated in Figure 8(d), the tests have been run in very challenging conditions in term of lighting, noise, distance and motion blur. In such conditions, no fiducial system used for comparison has been able to detect any marker out of 800 images apart from the proposed one, which has been able to detect 506 of them, (*i.e.* 63 %). This experiment has been conducted with  $c = 5$ , *i.e.* pixels values ranking between 0 and 51 pixels, a noise randomly distributed in  $[0, 10]$  pixels and with a camera to marker distance  $D = 30$ .

We want to emphasize two important points with these experiments. The first one is that, even in such very challenging conditions, we obtain a high detection rate, *i.e.* 94%, 80%, 57% and 22% for a magnitude of motion blur of 0, 5, 10 and 15 pixels respectively. The second point is

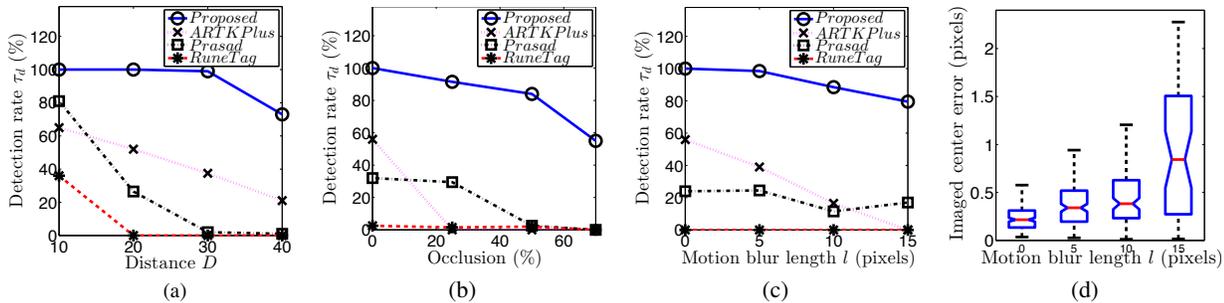


Figure 8. Detection rate (a) vs. the distance  $D$  (b) vs. the percentage of occlusion (c) vs. motion blur magnitude (d) Accuracy of the imaged center estimation vs. the motion blur magnitude. See text for details.

to emphasize that, even under these challenging conditions, with an increasing motion blur magnitude, the overall accuracy of the imaged center estimation is less than one pixel, while its median is less than 0.4 pixels for a motion blur magnitude less than  $l = 10$  pixels.

**False negative and confusion.** Lastly, Table. 1 shows the results of the evaluation of the false positive and the inter-confusion marker rate. In the current implementation, a ring (black or white) holds a 1-bit information, its width being 0.10 or 0.15 (w.r.t. the outer circle of unit radius). In such a way, a library of fiducial markers composed of  $N$  rings has  $2^{2N-1}$  unique markers. We use  $N = 3$  rings markers in our experiments, e.g. a total number of 32 IDs. Once a fiducial marker has been detected and its 1D signals along the cross sections have been rectified via the homography (5) obtained through the optimization (6), the latter are read in a robust manner via the distance proposed in [9] then delivering the marker ID associated to a marker profile. The false positive rate evaluation has been performed on five video sequences whose the image resolution is  $640 \times 360$  pixels. All the systems but PRASAD present a false positive rate of 0%. However, our system presents a misidentification rate of 2.8% which is more than the ones delivered by ARTK-Plus but still negligible for a large number of applications such as automatic 3D reconstruction via Structure-from-Motion techniques or also camera localization for which constraints brought by the geometry can be used for ID disambiguation (e.g. through guided matching techniques) as long as a subset of imaged fiducial are correctly detected and identified in the image.

**Time performance.** The current CPU implementation delivers a frame rate of 4fps on one CPU (i5-4590, 3Ghz) core on a  $1280 \times 720$  image (showing up to 5 markers), which has been raised to 11 fps using a preliminary GPU (Nvidia GTX 980 Ti, CUDA 7.0) implementation.

## 5. Conclusion

In this work we presented a new fiducial system based on concentric circles and we showed that our system

|                | Prop. | PRASAD | RuneTag | ARTKPlus | Nb. frames |
|----------------|-------|--------|---------|----------|------------|
| <b>Garden</b>  | 1     | 51     | 0       | 0        | 1268       |
| False pos.     | 0%    | 4%     | 0%      | 0%       |            |
| <b>Indoor</b>  | 0     | 21     | 0       | 0        | 1685       |
| False pos.     | 0%    | 1.3%   | 0%      | 0%       |            |
| <b>Street</b>  | 0     | 80     | 0       | 0        | 1153       |
| False pos.     | 0%    | 6.9%   | 0%      | 0%       |            |
| <b>Stripes</b> | 0     | 172    | 0       | 0        | 762        |
| False pos.     | 0%    | 22.6%  | 0%      | 0%       |            |
| <b>Text</b>    | 0     | 23     | 0       | 0        | 525        |
| False pos.     | 0%    | 4.4%   | 0%      | 0%       |            |
| <b>Total</b>   | 1     | 347    | 0       | 0        | 5393       |
| False pos.     | 0%    | 6.4%   | 0%      | 0%       |            |
| Nb. detect.    | 744   | 221    | 72      | 351      | 800        |
| Fig. 8 (a)     | 93%   | 27.6%  | 9%      | 44%      |            |
| Nb. conf.      | 21    | 71     | 0       | 0        | 800        |
| Fig. 8 (a)     | 2.8%  | 32%    | 0%      | 0%       |            |

Table 1. False positive and misidentification evaluation. See text for details.

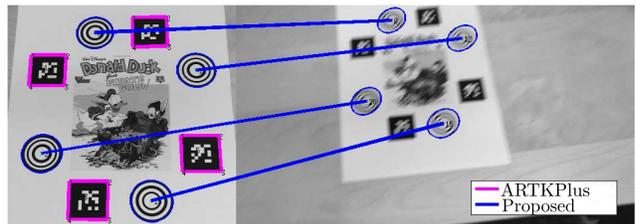


Figure 9. Detection and matching of fiducial across two images under challenging shooting conditions: the left image is correctly focused contrary to the right image which is acquired by a fast moving camera which generates motion blur. The ARTToolkitPlus system correctly detects and identifies its markers (surrounded in magenta) only in the sharp (left) image whereas no one is detected in the blurred view. The proposed system correctly detects and identifies the imaged markers in the two images.

provides an high detection accuracy as well as a good recognition rate in many challenging conditions, ranging from severe occlusions to motion blur to illumination changes. The code will be released in open source, available at <http://github.com/lcalvet/CCTag>.

**Acknowledgement.** This work has been supported by the POPART EU-H2020 project (number 644874) [1].

## References

- [1] POPART project. <http://www.popartproject.eu/>.
- [2] F. Bergamasco, A. Albarelli, E. Rodolà, and A. Torsello. Rune-tag: A high accuracy fiducial marker with strong occlusion resilience. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 113–120, 2011.
- [3] F. Bergamasco, A. Albarelli, and A. Torsello. Pi-tag: a fast image-space marker design based on projective invariants. *Mach. Vis. Appl.*, 24(6):1295–1310, 2013.
- [4] L. Calvet and P. Gurdjos. An enhanced structure-from-motion paradigm based on the absolute dual quadric and images of circular points. In *IEEE International Conference on Computer Vision (ICCV)*, pages 985–992, 2013.
- [5] L. Calvet, P. Gurdjos, and V. Charvillat. Camera tracking based on circular point factorization. In *International Conference on Pattern Recognition (ICPR)*, pages 2128—2131, 2012.
- [6] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.
- [7] D. Claus and A. W. Fitzgibbon. Reliable Fiducial Detection in Natural Scenes. In T. Pajdla and J. Matas, editors, *European Conference on Computer Vision (ECCV)*, volume 3024, pages 469–480, Berlin, Heidelberg, 2004.
- [8] M. Fiala. ARTag, a Fiducial Marker System Using Digital Techniques. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 590–596, 2005.
- [9] O. Gallo and R. Manduchi. Reading 1d barcodes with mobile phones using deformable templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(9):1834–1843, 2011.
- [10] L. B. Gatrell, W. A. Hoff, and C. W. Sklair. Robust image features: Concentric contrasting circles and their image extraction. In *Robotics-DL tentative*, pages 235–244. International Society for Optics and Photonics, 1992.
- [11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.
- [12] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *IEEE and ACM International Workshop on Augmented Reality (IWAR)*, pages 85–94. IEEE Comput. Soc, 1999.
- [13] K. Ken-ichi and O. Naoya. Automatic detection of circular objects by ellipse growing. *Int. J. Image Graphics*, 4(1):35–50, 2004.
- [14] A. Nègre, C. Pradalier, and M. Dunbabin. Robust vision-based underwater homing using self-similar landmarks. *J. Field Robotics*, 25(6-7):360–377, 2008.
- [15] Y. W. P. Gurdjos, P. Sturm. Euclidean structure from  $n \geq 2$  parallel circles: Theory and algorithms. In *European Conference on Computer Vision (ECCV)*, 2006.
- [16] M. G. Prasad, S. Chandran, and M. S. Brown. A motion blur resilient fiducial for quadcopter imaging. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 254–261, Washington, DC, USA, 2015.
- [17] M. Toyoura, H. Aruga, M. Turk, and X. Mao. Monospectrum marker: an AR marker robust to image blur and defocus. *The Visual Computer*, 30(9):1035–1044, Dec. 2013.
- [18] D. Wagner and D. Schmalstieg. ARToolKitPlus for Pose Tracking on Mobile Devices. In *Computer Vision Winter Workshop (CVWW)*, pages 139–146, 2007.
- [19] X. Ying and H. Zha. Efficient detection of projected concentric circles using four intersection points on a secant line. In *International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.