# Single View Correspondence Matching for Non-Coplanar Circles Using Euclidean Invariants

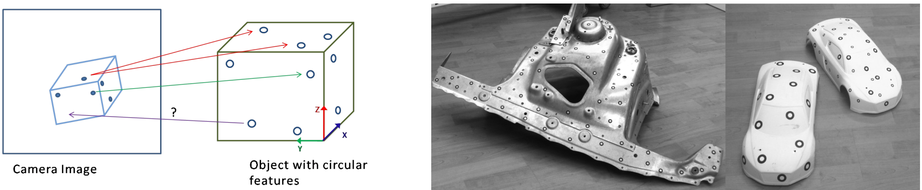
BMVC 2014 Submission # 123


**Abstract**

In this work we introduce a method to determine 2D-3D correspondence for non-coplanar circles using a single image, given that the 3D information is known. The core idea of our method is to compute 3D information from 2D features, thereby transforming a 2D-3D problem to a 3D-3D problem. Earlier researchers suggested that a pair of non-coplanar circles preserves Euclidean invariants under perspective projection. These invariants can be extracted from their image projections, but with a two fold ambiguity. In this paper, we propose *Conic pair descriptor* based on the Euclidean invariants. The proposed descriptor computes unique Euclidean invariants from known 3D model and Euclidean invariants with two fold ambiguity from its image projections. The proposed matching approach follows three steps to obtain correspondences between the circular features against the ambiguity. In this paper, we have included a detailed account of factors affecting the computation of invariants from conic projections. We have conducted experiments on real and synthetic models, in order to evaluate the proposed method. The experiment with synthetic images focuses on showing the impact of the size and plane orientation of the circles on the success of descriptor matching. We prepared 3D models with artificial circular features and obtained the ground truth 3D data with a Photogrammetric measurement system. The results of the correspondence matching algorithm are evaluated against the ground truth. We also show that our method is robust against false positives and capable of supporting real-time applications.


# 1 Introduction

Correspondence matching is one of the key problems in computer vision. Pose estimation and object recognition applications require accurate knowledge of correspondence relation [Yuji: $(m_i \leftrightarrow M_i)$]$m_i \leftrightarrow M_i$ between 3D model features [Yuji: $(M_i)$]$M_i$ and 2D image features [Yuji: $(m_i)$]$m_i$. In case of monocular systems the [Yuji: $m_i \leftrightarrow M_i$]2D-3D correspondence problem becomes more challenging as the depth information is lost. Various Computer Vision and Augmented Reality applications demand correspondence matching from a single view. Popular approach to achieve this is to compute projective invariant from features [Yuji: like]such as points, lines and conics [8]. Such features are [Yuji: selected]preferred because they are easier to detect from the images. Invariants are extensively studied topic in vision community, Forsyth et al. [6] and Gros [7] covered a detailed study on [Yuji: projective invariant descriptors]projective invariants and their stability under projective transformation.




(a) A primitive example explaining the correspondence problem when circular features exist on different planes of the model

(b) Example of industrial objects with circular markers used in close range Photogrammetry

Figure 1: Introduction to correspondence matching problem

In this paper we will focus on a specific class of features, that is *circles*. Circles are one of the most primitive features[Yuji: ,]. In case of industrial scenario circles are widely present on the model as natural features or circular fiducial are used for Photogrammetric measurements [16] ([Yuji: ]see Fig. 1). In such cases it is favourable to build tracking approach depending on such features. A single view correspondence matching problem with coplanar circles and ellipses is widely studied [13][Yuji: ,]. However matching problem with non-coplanar circles has not been addressed. Figure 1 shows a basic example of the problem, where multiple identical circular features exist on different planes of a 3D model. In this case features existing on the model is an advantage, but coplanar invariants can not be used for correspondence matching. In out work we focus on solving this problem of single view matching of multiple non coplanar conics.

3D objects retain Euclidean invariants rather than projective invariants [6]. In case of 3D objects Euclidean invariants are difficult to extract from images due to perspective mapping. Circles have a special property to retain depth information under projective transformation. A world circle always produces an elliptical curve on the image plane. If size of circle is known orientation of circle plane can be defined in 3D (camera coordinates) with a two fold ambiguity [6] [25]. Further, Forsyth et al. [6] proposed that up to three projective invariant can be computed from a non-coplanar pair of circles. They explain that angle between circle planes (angle between surface normals) and distance between centre of the circles are invariant quantities. The concept was proposed in early 90s, however these invariants have remained unexplored. We propose using these invariants to solve correspondence problem when multiple 3D circular features exist on a model. In this approach we bring problem from 2D to 3D by computing 3D invariants from image features, then solve 3D-3D matching problem with model. We introduce *ConicPairDescriptor*, which encapsulate invariants computed from elliptical image features to solve the conic ambiguity and provide accurate matching with 3D features. The proposed method is first attempt to use these invariants, therefore we also carried out simulations to show stability of invariants against change of perspective.

Our contribution is a new method to accurately identify image correspondences when multiple identical 3D circular features exist in the scene. We assume that calibration of camera is known and 3D information of features is available. Often in Industry based model tracking applications 3D-CAD data is known. Our matching method is suitable for tracking any 3D objects having known circles on different planes. In close range photogrammetry multiple circular markers (Figure. 1) are placed on 3D models for surface measurements [16]. These measurements include computation of surface normal and 3D position of each marker. This process involves taking multiple images of the model with additional presence of encoded markers in the scene to solve correspondence. Once 3D measurements are done our method can be extremely useful to support tracking application without using coded patterns. Similarly various industrial parts having natural circles can be identified and tracked with this method. We prepared two car models with circular markers for evaluating our matching method. The proposed method can find corresponding circular marker from a single image, with high accuracy. We also show that our method is stable against false positives detected from the scene. Our method is fast enough to support real-time tracking applications.

## 2 Related Work

Object detection and pose estimation from conic features is widely studied in 3D vision literature [2][25] [29] [22]. Circular shape is also a popular choice for designing artificial fiducial. Detection of contour points from image and fitting ellipse is a well studied topic [5]. Quan [23] proposed a two view approach for finding correspondence and 3D reconstruction with conic section. Authors have proposed methods to compute invariants for coplanar conics [6] [4]. A 3D problem is simplified to 2D when coplanar features are recovered and used for correspondence. Ying et al. [30] use a coplanar pair for camera calibration. Uchimaya et al. [28] developed invariant descriptors from multiple coplanar circles, and extended the work for deformable model [26]. Work of [14] [18] [21] propose a circular marker for 6D pose estimation, no invariants are computed as correspondence is solved by using unique coded pattern around the circle. [14][21] use circular shape to define circle plane in 3D, Additionally use a coded pattern is used encode 6D pose without ambiguity.(Fig). Luhmann [16] provides detailed account of methods using point circular fiducials in close range Photogrammetry. The current state of the art methods coded patterns are introduced to simplify correspondence problem. [GOM][AICON] are one of the industrial supplies for close range photogrammetry measurement equipments.

[Refer Thesis: Comment on catalogue based methods, ]

Literature study suggests that, existing methods either provide a solution for coplanar circular features or non-coplanar coded circular features. The novelty of our method is that it addresses 2D-3D matching problem for non planar circles present in the scene.

## 3 New introduction

[Paragraph: what is correspondence matching

- correspondence matching is ...
- 2 solutions.

] Correspondence matching aims to find correspondences between images or ones between object model and its image projection, thus is one of the key problems in computer vision. Since perspective transformation differs appearance of a single object on images, correspondence matching needs feature descriptors, which can compute invariant quantities against some transformation, and establishes correspondence by matching the invariant quantities obtained by the descriptors.

[Paragraph: Projective invariants on planar objects or primitive features] [Yuji: ask Hemal whether this paragraph is right or not.] One of the popular approaches use invariant quantities on planar object against perspective transformation. This type of approaches are based on invariants theory, which has been well-studied in 90s. There exist several works that covered a detailed study on projective invariants and their stability under projective transformation [6, 7] The invariants theory revealed that primitive features such as points, lines, and conics have invariant quantities against perspective transformation [8] and the idea is to use these invariant quantities for matching. Especially, planar projective descriptors, which compute invariant quantities against perspective transformation from planar object, have been investigated. Quan used **WHAT** invariant quantity from conic section for finding correspondence between two images and 3D reconstruction [23]. Coplanar conics was used to compute **WHAT** invariant quantity [4, 6]. [Yuji: any more projective invariants on primitive features?] [Yuji: will mention why these invariants have used less in this decade.] [Yuji: is here the best place to mention triangulation based matching such as ART?]

[Paragraph: some invariants on fiducial markers] Another approach also using primitive features embeds binary patterns into a planar object, called fiducial markers. The embedded patterns are extracted from images by adaptive thresholding for matching the correspondence between the marker model and its image projection and then the extracted patterns are decoded for identifying the extracted marker. The original work by Kato and Billinghurst [9] uses the contour of a rectangle but does not use any invariant quantity. Matsunaga et al. made a special chessboard with different pattern sizes and use cross ratio, which is a projective invariant quantity, defined by intersection of the patterns [17]. Recent works compute perspective [19], affine [20], rotation invariants [27] on key-points distribution defined by a set of key-points' location. [Yuji: any disadvantage?]

[Paragraph: local texture descriptors] Recent trend computes some invariant quantities defined by local texture in image such as SIFT descriptor [15]. Contrast to the above approaches, this type of approaches does not assume any primitive features in target scene. Alternatively, the approaches assume textured scene so that we can use the approaches for mobile applications, in which primitive shapes cannot be assumed. The approaches focus on improving their invariance as well as efficient memory usage so that the matching can run even on powerless processors in mobile phones [1, 10]. One of the disadvantages of these approaches is that their invariance are up to 2D transformation such as affine and rotation transformation. To handle view change caused by perspective transformation, we must explicitly learn how invariant quantities are affected by perspective transformation [11, 12]

## 3.1 Our contribution

[Paragraph: Our contribution] [Yuji: will touch later]
[Comments: h: our work based on conic invariants] We propose using these invariants to solve correspondence problem when multiple 3D circular features exist on a model. In this approach we bring problem from 2D to 3D by computing 3D invariants from image features, then solve 3D-3D matching problem with model. We introduce *ConicPairDescriptor*, which encapsulate invariants computed from elliptical image features to solve the conic ambiguity and provide accurate matching with 3D features.The proposed method is first attempt to use these invariants, therefore we also carried out simulations to show stability of invariants against change of perspective.

[Comments: i: our contribution] Our contribution is a new method to accurately identify image correspondences when multiple identical 3D circular features exist in the scene. We assume that calibration of camera is known and 3D information of features is available. Often in Industry based model tracking applications 3D-CAD data is known. Our matching method is suitable for tracking any 3D objects having known circles on different planes. In close range photogrammetry multiple circular markers (Figure. 1) are placed on 3D models for surface measurements [16]. These measurements include computation of surface normal and 3D position of each marker. This process involves taking multiple images of the model with additional presence of encoded markers in the scene to solve correspondence. Once 3D measurements are done our method can be extremely useful to support tracking application without using coded patterns. Similarly various industrial parts having natural circles can be identified and tracked with this method. We prepared two car models with circular markers for evaluating our matching method. The proposed method can find corresponding circular marker from a single image, with high accuracy. We also show that our method is stable against false positives detected from the scene. Our method is fast enough to support real-time tracking applications.

# 4 Conic Invariants : Theory and Computation

This section explains the theory of conic invariants, which was derived in [6]. From here, variables written in capital letters denote variables related to conics in 3D model and ones written in lowercase letters denote variables related to conics in images.

## 4.1 Conic invariants in a 3D coordinate

Suppose there exists a set of non-coplanar ~~conics~~ circles in a 3D coordinate 3D space?/system? and each of which is described by two variables the normal vector $N_i \in \mathbb{R}^3$ and the centre position $M_i \in \mathbb{R}^3$. plane and a normal defines a plane rather than a conic ~~Forsyth et al. [6] explained that in case of three dimensional objects invariant descriptors consists of Euclidean invariants rather than projective invariants. Various invariants can be computed from a pair of two non-coplanar circles based on Ellipse backprojection.~~ better to move this striked out sentence to other places : Introduction? Hemal Forsyth et al. pointed that a pair of two non-coplanar conics $(N_i, M_i)$ and $(N_j, M_j)$ gives us the following Euclidean invariant quantities under perspective transformation: angle between the conics and the distance between their centre positions. The angle between the conics $\Theta$ is equivalent to the angle between the surface normals, thus is obtained as

$$\Theta_{i,j} = \angle(N_i, N_j), \quad i \neq j, \tag{1}$$

where $i$ and $j$ denote the index of the conics. The distance between the conic centres $D$ is computed as

$$D_{i,j} = \mathrm{dist}(M_i, M_j), \quad i \neq j. \tag{2}$$

## 4.2 Conic invariants in an image coordinate

Same as the 3D case, we can compute the angle $\theta$ and the distance $d$ between two non-coplanar conics appearing on an image given their radius $r_s$ and $r_t$ in a 3D coordinate. It is assumed that the conic parameters $C$ and its centre position $p$ of each conic in the image are already known. Following a conic back projection method described in [6], we can compute the normal vector $n_s, n_t \in \mathbb{R}^3$ and the centre position $m_s, m_t \in \mathbb{R}^3$ up to a two fold ambiguity as

$$\{(n_s^a, m_s^a)\} = \text{EllipseBackprojection}(p_s, C_s), \quad a = \{1,2\}, \tag{3}$$

$$\{(n_t^b, m_t^b)\} = \text{EllipseBackprojection}(p_t, C_t), \quad b = \{1,2\}, \tag{4}$$

where the superscripts $a, b$ represent the conic ambiguity. The two fold ambiguity is called ~~backprojection ambiguity~~ or conic ambiguity?.

Same as Eqs. 1 and 2, we can compute the angle $\theta_{s,t}$ and the distance $d_{s,t}$ between the two conics $\{(n_s^a, m_s^a)\}, \{(n_t^b, m_t^b)\}$ as

$$\begin{aligned} d_{s,t}^{a,b} &= \text{dist}(m_s^a, m_t^b), \\ \theta_{s,t}^{a,b} &= \angle(n_s^a, n_t^b), \end{aligned} \qquad s \neq t, \quad a, b = \{1,2\}. \tag{5}$$

Forsyth et al. mentioned that $d_{s,t}^{a,b}$ is consistent while $\theta_{s,t}^{a,b}$ varies in [6]. Following this, we regard that a pair of two non-coplanar conics in an image derives a unique $d_{s,t}$ and four $\theta_{s,t}^{a,b}$.

# 5 Method

This section first introduces the proposed descriptor, called *Conic pair descriptor*, in Sec. 5.1 and further describes a matching method using the descriptor in Sec. 5.2. We assume that both the 2D and 3D data are already available and focus on the proposed descriptor and the matching method in detail. The 3D data includes the surface normal $N_i$, centre position $M_i$ and size $R_i$ (diameter) of each circle on the model. The 2D data includes circle centre $m_i$ and conic matrix $C_i$. The 2D data is further used to recover the surface normals ${Nc_i}^a$ and centre positions ${Mc_i}^a$ from the image conics, where $a = 1, 2$ denotes the index of ambiguous solution. The 2D data is extracted from an input image by the following procedure: ellipse detection given an input image [18]; conic parameters estimation from the detected ellipse [3]; and the normal vector and the centre position recovery from the conic parameters [14].

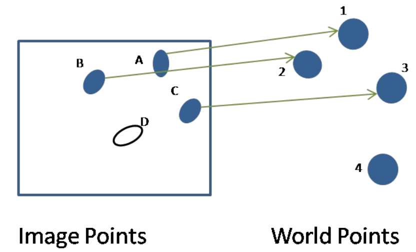
## 5.1 Descriptor Generation

This part mainly discusses generation of *Conic pair descriptor* from Euclidean invariants. The invariants for 3-D model are computed from available 3D data $(M_i, N_i)$ without any ambiguity (Eq. 6) using Eq. 1, 2. The same set of invariants can be computed from corresponding image features using Eq. 5, where the recovered $d_c$ component is unique and $\theta$ component has 4 solutions (Eq. 7). In our approach we pursue the idea that the existence of multiple features on the model can be used to overcome the *Conic ambiguity* problem. The principle idea is to generate descriptors from conic invariants to perform a descriptor matching to obtain $m_i \leftrightarrow M_i$ correspondence. The proposed *Conic pair descriptor* structure,

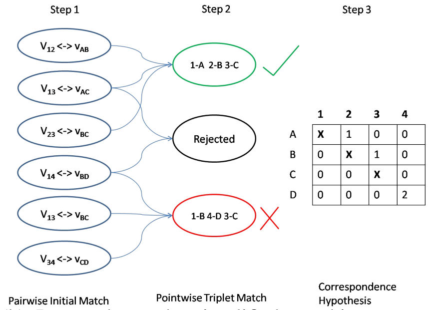
$$\textit{Conic pair descriptor}_{model} = V_p \equiv \langle d_c, \theta \rangle_{i,j} \tag{6}$$

$$\textit{Conic pair descriptor}_{image} = v_q \langle d_c, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22} \rangle_{i,j} \tag{7}$$

where $V_p$ represents world circles $i, j$ and $v_q$ represents image conic pair $i, j$. The reader should note that given a set of points and their corresponding normals in 3D camera space,




(a) A problem showing four image points, and their correspondence with world points (1-A,2-B,C-3) and one false positive (D)




| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| A | X | 1 | 0 | 0 |
| B | 0 | X | 1 | 0 |
| C | 0 | 0 | X | 0 |
| D | 0 | 0 | 0 | 2 |

(b) Image shows the simplified matching process for the problem suggested in (a), partial results of all 3 steps are shown.

Figure 2: Matching problem and the overview of the method to generate correspondence hypothesis

PFH descriptor [24] also computes similar invariants. In this case, the concept can not be applied to the results of *Ellipse backprojection* as the *Conic ambiguity* restricts us from computing a unique set of invariants. Unlike popular methods, a *Conic pair descriptor* represents two features at same time. In order to uniquely represent a single conic using Euclidean invariants, at least more than two conic features are required. Addition of each conic feature adds 3 wrong solutions of $\theta$ in the descriptor, additionally the matching must rely on detection of all the conics used for descriptor computation. We propose matching $v \leftrightarrow V$ first, thereby finding a corresponding conic pair, further we solve individual correspondence $m_i \leftrightarrow M_i$ problem. Descriptor $v_{\{1..q\}}$ are computed among each pair of detected $n$ image conics, where $q = \binom{n}{2}$. $V_{\{1...p\}}$ are computed off-line as the 3D data is already available. In this case for $l$ world circles $p \leq \binom{l}{2}$, as pairs not likely to appear in same image can be rejected. After computing the *Conic pair descriptors* the following 3 step matching approach is used to achieve $m_i \leftrightarrow M_i$ correspondences.

## 5.2 Descriptor matching

### 5.2.1 Step 1: Pairwise Initial Matching

In the first stage of the matching process we compare the *Conic pair descriptors*. The objective is to reduce complexity of the problem by finding the possible pair correspondences ($v \leftrightarrow V$). First the unique component($d_c$) between the descriptors is compared, if matched then ($\theta$) component of $V$ is compared with all 4 values of $\theta$ component in $v$ (Ref. Algorithm 1). $T_{d_c}$ and $T_\theta$ are the threshold values used to compare the respective components. The stage may result in *one to many* type of relation between descriptors. This can be either due to similar feature orientation on object or due to presence of the *Conic ambiguity*. The reader should note that a descriptor represents a pair of conics, therefore the stage is called pairwise matching. The example given in Figure 2 shows possible outcome of step 1 with respect to the given example problem.

```
Goal : Find all possible V_p similar to v_q ;
Initialisation : T_{d_c} = 10 , T_θ = 5 ;
forall the 3D Feature Descriptors (V), p ← 0 to n do
    forall the 2D Feature Descriptors (v), q ← 0 to l do
        if compared_c (V_p,v_q) < T_{d_c} then // compares d_c component
            if compareθ (V_p,v_q) < T_θ then // compares θ component
                // All 4 solutions of θ in v_q are checked
                ;
                SavePairResult(p,q) // Save matching descriptor
                   pair
                ;
            end
        end
    end
end
```

**Algorithm 1:** Pairwise Initial Matching algorithm

### 5.2.2 Step 2: Pointwise Triplet Matching

In this stage we simplify the problem further and obtain hypothesis on point wise matching ($m_i \leftrightarrow M_i$) by performing a verification on $v \leftrightarrow V$ matching results. The objective is to compare the results of step 1 to identify and reject false descriptor matches. We seek three $v \leftrightarrow V$ results, such that they complement each other to form a unique three points $m_i \leftrightarrow M_i$ hypothesis. A simple two stage approach is proposed to generate a triplet matching hypothesis,

1 Find any two results of Pairwise Initial Matching in which both the image and the model descriptors represent one and only one common conic. If such results exist then an initial triplet matching hypothesis can be proposed.

$$V_{12} \leftrightarrow v_{AB}, V_{13} \leftrightarrow v_{AC} \xrightarrow{\text{Triplet Hypothesis}} [1\ 2\ 3] \leftrightarrow [A\ B\ C]$$

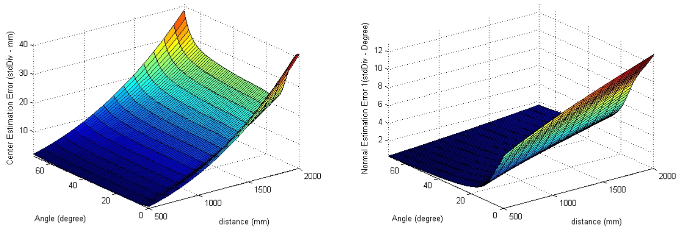
In the example above we can see that world conic 1 and image conic A is common among the two solutions. We form a 3 point matching hypothesis with these results.

2 Find a new descriptor matching pair which can verify the triplet matching hypothesis formed in the previous stage (e.g. $V_{23} \leftrightarrow v_{BC}$).

The verified triplets are saved and others are rejected. The results may also contain false triplet matches (Fig. 2). If $x$ number of results are generated in step 1, the number of pairs compared in this stage is $\binom{x}{3}$.

### 5.2.3 Step 3: Correspondence Hypothesis

In this final step results of triplet matching are combined and a voting matrix is generated (Fig. 2). A $m_i \leftrightarrow M_i$ pair having maximum votes in the matrix is proposed as a correspondence hypothesis. In case of conflicting votes the respective $m_i \leftrightarrow M_i$ relation is not considered. A minimum of 3 correspondence are required to compute the pose of the object




(a) Centre estimation error $Mc_i$ (b) Normal estimation error $Nc_i$

Figure 3: Ellipse backprojection results,Image noise = $\sigma = 0.3$, $R_i = 12mm$

[13], if camera intrinsics are known. A pose of the object can be computed by selecting top 3 correspondence results and verify other correspondences obtained from the matrix. If only 3 out of $n$ conics are detected in the image, verification with pose is not possible and results may not be reliable.

# 6 Evaluation

In this section we will cover experiments carried out to comment on accuracy and robustness of the algorithm. The reader should note that the problem of achieving single image 2D-3D correspondences for non-coplanar circles has not been addressed earlier. Therefore, alternative methods for comparison are not available. We prepared test models by attaching circular markers 1, Model 1 with $R_i$ = 12 mm ($p = 20$) and Model 2 with $R_i$ 5mm ($p = 26$). These markers used are widely accepted and used in the Industrial domain for Photogrammetric measurements. 3D measurements for the markers are done with state of the art metrology system, and the ground truth is established by giving each model point a unique ID in the database. The markers are attached randomly and coplanar placement is avoided. A high resolution (2560x1920) camera is used for the experiments and MATLAB is used for synthetic experiments.

## 6.1 Preliminary experiment

The quality of recovered plane from *Ellipse backprojection* depends on distance from the camera $r$ and the viewing angle $\eta$(angle between the image plane and the circle plane)[29]. We performed simulations to understand behaviour of Ellipse Backprojection with respect to both $r$ and $\eta$. The parameter $r$ is varied from 0.5 to 2 m and $\eta$ from 0-70° in step wise manner, 100 iterations are performed at each position (Fig. 3). The results (fig. 3) suggest that, at low viewing angles $\eta$ 0-10° both normal and centre estimation errors are higher, at any given distance. At lower values of $\eta$ the image projection of a circle is more circular than elliptical, therefore recovery of ellipse parameters may have errors. The estimation error grows at higher camera distance, however the error in normal recovery appears less sensitive to increase in camera distance than the error in centre recovery. The results obtained with $R_i = 5$ mm show similar pattern, although the magnitude of error is higher as smaller image projections reduce the accuracy of computation of ellipse parameters. We also compared the ambiguous results of estimated centres $Mc_i{}^1$ and $Mc_i{}^2$. The maximum distance recorded between the two is $\leq 0.1$ mm for $R_i$ = 12 mm, this suggests that ambiguity can be neglected

Table 1: Descriptor Matching Analysis

| $R_i$ | $\theta$ | $d_c$ | Camera Distance (mm) | Min-Max Success(%) |
|---|---|---|---|---|
| 5 | 10-40 | 10-150 | 500-2000 | 58-82 |
| | 40-80 | 10-150 | 500-2000 | 35-65 |
| 12 | 10-40 | 10-150 | 500-2000 | 64-86 |
| | 40-80 | 10-150 | 500-2000 | 40-69 |
| 20 | 10-40 | 10-150 | 500-2000 | 74-92 |
| 20 | 40-80 | 10-150 | 500-2000 | 48-77 |

for the recovered centre position. This behaviour also explains consistency of invariant $d_c$ between two image conics.

## 6.2 Experiment 1 : Descriptor Matching vs Marker Orientation

This experiment is carried out synthetically to observe the effect of orientation of the circles on the descriptor matching results. Two circle planes are placed in different orientations and images are captured (noise $\sigma = 0.3$) from 1000 different camera positions for each orientation. The control parameters, the distance between circle centres $d_c$ is varied from 10 - 150 (mm) and the angle between the planes $\theta$ is varied from 10 - 80°. The objective is to recover the descriptor components $d_c$ and $\theta$ from the images, compare them with ground truth and measure the success rate. Realistic values are used for camera intrinsics, $T_{d_c} = 10$ and $T_\theta = 5$ are kept constant. Camera positions are chosen at random, x-y-z rotation range is $\pm 70°$, x-y translation range is $\pm 500$ mm, z-translation (Camera distance) range is 500 to 2000 mm.

The table 1 provides summary of key observations made during the experiment. The matching success shows inversely proportional relation with $\theta$, independent of $d_c$. We learn that descriptor matching is influenced more by angle between planes than distance between the circle centres. It is also seen that success of matching can be improved by increasing the size of the circles. This experiments suggest that when features placement is possible, it is advised to choose larger circles or surfaces with lower plane angles for improved matching results.

## 6.3 Experiment 2 : Correspondence Matching vs Threshold Settings

In case of model with natural features choice over size or placement is not available. The aim of this experiment is to understand the role of threshold values $T_{d_c}$ and $T_\theta$ in $m_i \leftrightarrow M_i$ matching results. In order to perform this experiment we took 75 images of car model with 12 mm markers (Distance Range 500-2000 mm). As suggested in Sec. 5.2.3, each image has at least 4 detected conics. The results are considered *Not Converged* (NC) in case of less than 3 $m_i \leftrightarrow M_i$ results.

In Sec. 4, we learned that $d_c$ recovery is weak and therefore a flexible threshold may be appropriate for matching. The results ( Table. 2) show that higher flexibility in $T_{d_c}$ impacts both precision and recall values in negative manner. On the other hand, very stringent thresholds lead to lower recall values. Therefore, a right balance of threshold can be selected to achieve higher precision and recall rates. Our preferred settings for experiments is $T_{d_c} = 10$ and $T_\theta = 5$. The selection may require change based on density of the features.

Table 2: Correspondence matching with varying threshold settings

| $T_\theta$ | $T_{d_c}$ | NC | Positive | FP | Precision | Recall |
|---|---|---|---|---|---|---|
| 5 | 5 | 13 | 62 | 0 | 100 | 82.6 |
| 5 | 10 | 8 | 67 | 0 | 100 | 89.33 |
| 5 | 15 | 4 | 67 | 4 | 94.36 | 89.33 |
| 3 | 5 | 28 | 47 | 0 | 100 | 62.66 |
| 3 | 10 | 21 | 54 | 0 | 100 | 72 |
| 3 | 15 | 19 | 53 | 3 | 94.64 | 70.66 |

Table 3: Robustness against False Positive

| Model | Images | NC | Positive | FP | Precision | Recall |
|---|---|---|---|---|---|---|
| Model 1 | 50 | 1 | 49 | 0 | 100 | 98 |
| Model 2 | 50 | 4 | 31 | 15 | 67.39 | 62 |

## 6.4 Robustness against false positives

This experiment aims to show robustness of the matching method in presence of false positives in the scene. In order to introduce false positives in the scene, Model 1 and Model 2 are placed in the same scene and images are captured from different positions. The matching method is provided 3D information of one model at a time, which in turn make the markers present on the other model act as false positives in the image. The same set of images are used for the two experiments and the results of the experiment are shown in Table 3. We learn that the false positives are completely rejected when matching is focussed on Model 1, On the contrary in case of Model 2 precision and recall values suffer due to presence of false positives from Model 1. In case of Model 1, markers have bigger size and therefore invariant recovery is strong. This can explain higher robustness of Model 1 against false positives. [Exp with False positive on Industrial Model : Data Yet to be collected from Office]

## 6.5 Time Analysis

In this experiment we focus on analysing time consumed by the matching method when introduced into a tracking application. Two cameras CAM 1 (2560 x 1920) and CAM 2 (640x480) are used for tracking, the results presented are averaged over 100 frames. The results show that our method takes $\leq 1\%$ time from in the tracking pipeline. In terms of frame rates we achieve 2-3 FPS with CAM 1 and 7-8 FPS with CAM 2. Additionally, an exhaustive experiment with 90-140 false positives in the scene shows that the matching method consumes maximum time in the pipeline (0.7 FPS). Limited tracking range ($< 500$ mm) of CAM 2 does not allow experiment with such large number of false positives.

Table 4: Time Analysis

| Algorithm Stage | CAM 1 | | CAM 2 |
|---|---|---|---|
| | Model | Model + FP | Model |
| Image Undistortion | 39.53% | 12 | 11.79 |
| Marker Detection | 38.51 | 34 | 30.75 |
| Correspondence Matching | 0.35 | 44.2 | 1.34 |
| Pose Estimation | 21.61 | 9.8 | 56.12 |

# 7 Conclusion & Future work

In this paper we have demonstrated a successful approach for solving 2D-3D correspondence matching problem for non-coplanar circular features from a single image. We propose a new *Conic descriptor* which represents euclidean invariants generated by a pair of non-coplanar circles. Our method can successfully define correspondences for more than 3 circular features are present in the scene. The proposed method is the first to address the correspondence matching using these invariants since its introduction in the 90s. Our contribution also includes providing detailed understanding of behaviour of invariants with respect to orientation and size of the circles. The major factors affecting matching are also discussed to optimize the method for best possible matching results based on application (threshold settings, circle size, camera distance). The results of the experiments support our claim, that the method is fast, reliable and robust against false positives. Our method can be used for object tracking or object identification in Industrial scenario, where natural or artificial circular features exist on the models. However, the method is generic and can be used for any application dealing with non coplanar circles. The 3D information of the features on the model and camera calibration are the only prerequisites for matching. The reader should also note that algorithm may not perform well with symmetric or coplanar arrangement of circular features.

In context of future work, we would like to improve the method to be able to handle features of different sizes simultaneously. Also a faster matching strategy is required to handle large number of feature points and false positives. We would like to use the same invariants to compute 2D-2D correspondence matching between two images in order to generate the 3D data which is a prerequisite now. We also consider using such matching algorithm to support creating 3D markers for monocular Augmented Reality applications. This can be a cheap alternative to conventionally used 3D spherical markers.

# References

[1] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: binary robust independent elementary features. In Proceedings of the 11th European conference on Computer vision: Part IV, ECCV'10, pages 778–792, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-15560-X, 978-3-642-15560-4. URL http://dl.acm.org/citation.cfm?id=1888089.1888148.

[2] Michel Dhome, Jean-Thierry Lapresté, Gérard Rives, and Marc Richetin. Spatial localization of modelled objects of revolution in monocular perspective vision. In Computer Vision—ECCV 90, pages 475–485. Springer, 1990. URL http://link.springer.com/chapter/10.1007/BFb0014897.

[3] Gerald E Farin and Dianne Hansford. The geometry toolbox for graphics and modeling. A.K. Peters, Natick, Mass., 1998. ISBN 1568810741 9781568810744.

[4] M. Ferri, F. Mangili, and G. Viano. Projective pose estimation of linear and quadratic primitives in monocular computer vision. CVGIP: Image Underst., 58(1):66–84, July 1993. ISSN 1049-9660.

[5] Andrew Fitzgibbon, Maurizio Pilu, and Robert B. Fisher. Direct least square fitting of ellipses. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21(5): 476–480, 1999.

[6] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell. Invariant descriptors for 3d object recognition and pose. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 13(10):971–991, 1991.

[7] Patrick Gros and Long Quan. Projective invariants for vision. Technical Report RT 90 IMAG - 15 LIFIA, MOVI - IMAG-INRIA Rhône-Alpes / GRAVIR , Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle - LIFIA, 1992.

[8] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge University Press, Cambridge, UK; New York, 2003.

[9] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR), pages 85–94, Washington, DC, USA, 1999. IEEE Computer Society. ISBN 0-7695-0359-4. URL http://dl.acm.org/citation.cfm?id=857202.858134.

[10] Yan Ke and Rahul Sukthankar. Pca-sift: a more distinctive representation for local image descriptors. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 2, pages II–506–II–513 Vol.2, 2004. doi: 10.1109/CVPR.2004.1315206.

[11] Daniel Kurz, Thomas Olszamowski, and Selim Benhimane. Representative feature descriptor sets for robust handheld camera localization. In IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pages 65–70, 2012. doi: 10.1109/ISMAR.2012.6402540.

[12] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(9):1465–1479, 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.188.

[13] Vincent Lepetit and Pascal Fua. Monocular model-based 3d tracking of rigid objects: A survey. Now Publishers Inc, 2005.

[14] Diego Lo\' pez de Ipin a, Paulo RS Mendon\cca, Andy Hopper, and Andy Hopper. TRIP: a low-cost vision-based location system for ubiquitous computing. Personal and Ubiquitous Computing, 6(3):206–219, 2002.

[15] D.G. Lowe. Object recognition from local scale-invariant features. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pages 1150–1157 vol.2, 1999. doi: 10.1109/ICCV.1999.790410.

[16] Thomas Luhmann, Stuart Robson, and Stephen Kyle: Ian Harley. Close Range Photogrammetry: Principles, Techniques and Applications: Principles, Methods and Applications. Whittles Publishing, revised edition edition, October 2006. ISBN 1870325508.

[17] Chikara Matsunaga, Yasushi Kanazawa, and Kenichi Kanatani. Optimal grid pattern for automated camera calibration using cross ratio. IEICE Transactions on Fundamentals, E83–A:1921–1928, 2000.

[18] Leonid Naimark and Eric Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In Proceedings of the 1st International Symposium on Mixed and Augmented Reality, page 27, 2002.

[19] Tomohiro Nakai, Koichi Kise, and Masakazu Iwamura. Camera-based document image retrieval as voting for partial signatures of projective invariants. In International Conference on Document Analysis and Recognition (ICDAR), pages 379–383. IEEE Computer Society, 2005. ISBN 0-7695-2420-6. doi: 10.1109/ICDAR.2005.64. URL http://dx.doi.org/10.1109/ICDAR.2005.64.

[20] Tomohiro Nakai, Koichi Kise, and Masakazu Iwamura. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. In Proceedings of the 7th international conference on Document Analysis Systems, DAS'06, pages 541–552, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-32140-3, 978-3-540-32140-8. doi: 10.1007/11669487_48. URL http://dx.doi.org/10.1007/11669487_48.

[21] Alain Pagani, Johannes Köhler, and Didier Stricker. Circular markers for camera pose estimation. In 12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2011), 2011.

[22] Long Quan. Invariant of a pair of non-coplanar conics in space: Definition, geometric interpretation and computation. In Computer Vision, 1995. Proceedings., Fifth International Conference on, pages 926–931, 1995. 00018.

[23] Long Quan. Conic reconstruction and correspondence from two views. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18(2):151–160, 1996.

[24] Radu Bogdan Rusu. Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany, October 2009.

[25] Reza Safaee-Rad, Ivo Tchoukanov, Kenneth Carless Smith, and Bensiyon Benhabib. Three-dimensional location estimation of circular features for machine vision. Robotics and Automation, IEEE Transactions on, 8(5):624–640, 1992. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=163786.

[26] Hideaki Uchiyama and Eric Marchand. Deformable random dot markers. In Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on, pages 237–238, 2011.

[27] Hideaki Uchiyama and Hideo Saito. Random dot markers. In IEEE Virtual Reality Conference, 2011. URL http://hvrl.ics.keio.ac.jp/uchiyama/me/code/UCHIYAMARKERS/index.html.

[28] Hideaki Uchiyama and Hideo Saito. Random dot markers. In Virtual Reality Conference (VR), IEEE, pages 35–38, 2011.

[29] Naoufel Werghi, Christophe Doignon, and Gabriel Abba. Pose estimation of objects based on circular patterns in monocular computer vision. In SPIE's 1996 International Symposium on Optical Science, Engineering, and Instrumentation, pages 254–257. International Society for Optics and Photonics, 1996. 00002.

[30] Xianghua Ying and Hongbin Zha. Camera calibration using principal-axes aligned conics. In Asian Converence on Computer Vision, pages 138–148, 2007.