Correspondence matching of non-coplanar circles from a single image

BMVC 2014 Submission # 666

Abstract

In this work we have proposed a method to determine 2D-3D correspondence between non-coplanar circles from a single image. Our method uses image conics to compute circle plane orientation in camera coordinate system, thus bringing the problem from 2D to 3D domain. This information is used to generate projective invariant descriptors which can be used directly for correspondence matching, given that the 3D information about circles are known. Additionally, the evaluation also covers study stability of the projective invariants and the factors affecting their computation. One of the intended applications is for tracking industrial objects with circles on it. In our approach we use conic properties of circles to compute projective invariant descriptors. These descriptors are matched with known 3D information to establish correspondences. We also demonstrate stability of the invariants used to generate descriptors both in practice and simulation. In our approach we compute 3D plane orientation of each circle from its image contour, and compute projective invariants between each pair of circles. We propose a new descriptor

1 Introduction

goal : Explain the problem in question, the motivation for the work and the proposed application in mind. The assumptions made in our work. We have the camera calibrated and we use 3D points and their normal positions for the descriptor measurement.

Correspondence matching is one of the key problems in computer vision. Applications related to pose estimation or object detection require accurate knowledge of model features and their corresponding image features. The problem becomes more challenging for monocular systems as the depth information is lost. Popular monocular methods involve learning object with few initial frames or require initialisation to facilitate matching [N3M so on]. Many authors have proposed using natural features like points, lines and conics for solving correspondence problem [6]. Such features are easier to extract from images and can be used to compute reliable projective invariants. Invariants are extensively studied topic in early vision community, Forsyth *et al.* [5] provided a detailed account on 3D invariant descriptors and their stability under projective motion. In learning based model tracking methods initial frames of the camera are used to learn model and compute unique descriptors from dense or sparse set of natural model features [1] [7] [10]. Other methods involve using a selective set features (edges, contours, etc.) to compute invariants from a single image [8]. In our method we demonstrate how non-planar circular features can be used to compute invariant descriptors and solve matching from single image.

© 2014. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

In single images projective transformation of features make it difficult to compute invariant features. Coplanar conics, lines and points can be used to compute planar invariants [5]. 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 [5] [17]. Further, Forsyth *et al*. [5] 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. (Fig of decsriptor match) We use invariant descriptors (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 (fig) are placed on 3D models for surface measurements [11]. 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.

[REFER : Comment on descriptor being propose of 3D point an normal based info.]

2 Related Work

Object detection and pose estimation from conic features is widely studied in 3D vision literature [2][17] [20] [14]. 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 [4]. Quan [15] proposed a two view approach for finding correspondence and 3D reconstruction with conic section. Authors have proposed methods to compute invariants for coplanar conics [5] [3]. A 3D problem is simplified to 2D when coplanar features are recovered and used for correspondence. Ying *et al*. [21] use a coplanar pair for camera calibration. Uchimaya *et al*. [19] developed invariant descriptors from multiple coplanar circles, and extended the work for deformable model [18]. Work of [9] [12] [13] propose a circular marker for 6D pose estimation, no invariants are computed as correspondence is solved by using unique coded pattern around the circle. [9][13] use circular shape to define circle plane in 3D, Additionally use a coded pattern is used encode 6D pose without ambiguity.(Fig). Luhmann

[11] 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 Method

We assume that both 2D and 3D data is already available, and will focus on the matching method in detail. 3D data includes the surface normal (N_i), centre position (M_i) and size (R_i) of the circles on the model. 2D data includes centre points (m_i) and conic matrices (C_i) recovered from the undistorted image. The camera intrinsics (K) and distortion parameters are known. We have followed approach of Naimark [12] and Fitzgibbon [4] for ellipse detection and fitting.

3.1 Conic Invariants : Theory and Computation

In this part we will explain the theory and computational aspect euclidean invariants computed from the image. If camera's projection centre is assumed as the vertex of a cone which has the world circle is at its base. The image plane can be considered as a cutting plane π, which always creates an elliptical cross section. A new plane $\pi' = T * \pi$ can be computed such that intersection of plane π' with the cone is circular. Plane (π') is parallel to the base of the cone hence a plane normal Nc_i can be computed from T. Additionally, 3D position of circle centre Mc_i can be computed in camera coordinate system if original radius is known. A normal Nc_i and a point on plane Mc_i are sufficient to define the plane of circle in camera coordinate system. T is a combination of two rotations [5][9], one of the two has $\pm\phi$ rotation angle which introduces the ambiguity (two solutions) in plane recovery. This implies that every image conic results in to two possible plane orientations, we call this method as Ellipse Backprojection.

$$\text{Ellipse Backprojection}(m_i, C_i) \rightarrow {Nc_i}^1, {Nc_i}^2, {Mc_i}^1, {Mc_i}^2 \quad (1)$$

Forsyth *et al*. [5] explained that for three dimensional objects descriptors consist of euclidean invariants rather than projective invariants. Three type of *Conic invariants* can be computed from a pair of non-coplanar circles. We use following invariants for our method,

1. **Angle between planes** (θ) : It is same as angle between their surface normals (i.e. $\angle(Nc_i, Nc_j)$). θ can be recovered from conic image without knowledge of circle size in real world.

2. **Distance between circle centres** : This vector $V(Mc_i, Mc_j)$ has three degrees of freedom. The length of the vector d_c is invariant (object scale should be known) and consistent despite of the ambiguity.

It should be noted that ambiguity of Ellipse Backprojection produces 4 solutions for θ, only one of which is correct. Other invariants can be computed from recovered normal and centre values. These invariants are unstable [5] as the error from both recovered components influences the computation. The quality of Ellipse Backprojection depends on distance from the camera and the viewing angle (angle between the image plane and the circle plane)[20]. We performed simulations for circles of diameter (ϕ) of 5,8,12 mm to understand behaviour of Ellipse Backprojection on small features. We varied the camera distance (0.5 to 2 m) and viewing angle (0-70°) in step wise manner, while recording 100 iterations at each step. At low viewing angles 0-10° both angle and centre recovery has higher errors, at any fixed distance. The shape of ellipse is almost circular at small viewing angles, this factor can explain high errors. Error in estimation grows with camera distance, however normal recovery appears less sensitive to camera distance than centre recovery. We can conclude that normal recovery has less errors but computation of θ will result in 4 solutions. The comparison of the ambiguity explains that estimated centres are very close to each other ($d({Mc_i}^1, {Mc_i}^2) \leq$ 0.1 mm for ϕ= 12 mm), therefore any one of the solutions can be chosen. The centre recovery is prone to higher errors, however has a unique solution (all 4 solutions are consistent).

3.2 Descriptor Generation

This part mainly discusses generation of *Conic descriptor* from computed invariants. Invariants for model points can be computed from available 3D data (M_i, N_i) without any ambiguity. The same set of invariants can be computed a single 2D image using Ellipse backprojection with Conic ambiguity. In our approach we pursue the idea that existence of multiple features can be used to overcome the Conic ambiguity problem. The principle idea is to generate unique descriptors from Conic invariants to perform a direct matching. We propose *Conic Descriptor* which also encapsulates the conic ambiguity,

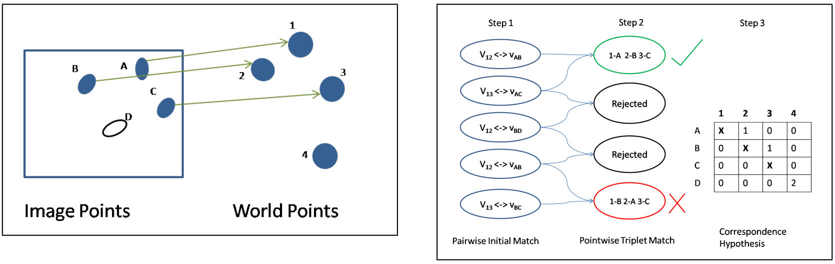
$$Conic\ Descriptor_{image} = v_q \langle d_c, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22} \rangle_{i,j} \tag{2}$$

$$Conic\ Descriptor_{model} = V_p \langle d_c, \theta \rangle_{i,j} \tag{3}$$

where v_q represents image conic pair i, j and V_p represents world circles i, j. PFH descriptors [16] show similar descriptor structure, input 3D point and normal information is without an ambiguity. Unlike popular methods *Conic Descriptor* is designed to represent two features at same time. A descriptor to represent a single conic requires using at least more than two conic features. Each conic adds 3 wrong solutions of θ, additionally matching must rely on detection of all conics used for descriptor computation. As correspondence is not available $v_{\{1..q\}}$ are computed for all 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 Conic descriptors from image following 3 step matching process is used to achieve 2D-3D correspondences.

3.3 Step 1: Pairwise Initial Matching

In the first stage of descriptor matching we find all possible matching descriptor pairs ($V \leftrightarrow v$). The strategy is to first compare unique component(d_c) of descriptors, if positive then check for ambiguous component (θ) for a possible match against all 4 values. T_{d_c} and T_θ are the thresholds used to compare the components. The reader should note that a descriptor



(a) (b)

Figure 1: Matching problem and the overview of the method to generate correspondence hypothesis ; (a) A problem showing four image points, and their correspondence with world points (1-A,2-B,C-3) and one false positive (D) (b) Image shows the simplified matching process for the problem suggested in (a), partial results of all 3 steps are shown.

represents a pair, therefore we will have hypothesis of a possible matching pair. Individual correspondence among the pair will not be solved in this stage.

Goal : Find all possible V_p similar to v_q ;
Initialisation : T_{d_c} = 10 , T_θ = 5 ;
forall the *3D Feature Descriptors (V), $p \leftarrow 0$* **to** *n* **do**
  forall the *2D Feature Descriptors (v), $q \leftarrow 0$* **to** *l* **do**
    if `compared`$_c$ `(`V_p,v_q`)` $< T_{d_c}$ **then** `// compares` d_c `component`
      if `compare`θ `(`V_p,v_q`)` $< T_\theta$ **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: Initial pair matching algorithm

3.4 Step 2: Pointwise Triplet Matching

In this stage we simplify the problem further and obtain point wise matching ($m_i \leftrightarrow M_i$) by performing a verification. After Initial pairwise matching we have a hypothesis about a possible conic pair match. However, since each descriptor represents two conics we can not verify exact correspondence by a single descriptor matching result. In order to solve this ambiguity we use two results (Fig 1)

Why? : because of ambiguity lot of results are false, by adding another constraint we restrict the probability of false matching. Also to make it point wise problem from a pair wise problem. - Matching steps on voting - Solving problem from pairwise matching to point wise matching

3.5 Step 3: Correspondence Hypothesis

Filter the voting matrix to generate point wise matching hypothesis. Comment on how many would be enough to check other hypothesis. (3 are enough to predict solutions) since we know the calibration. One can select only top voted 3 pairs (since they are 3D points) and then verify the whole choice, or go for 6 point matching and achieve the same.

4 Evaluation

Page 5: Explain the experiment model preparation. The selection of the markers as per standard. Good detection range with such sizes to give reader idea of tracking range. Model 1 : 12mm Model 2 : 8mm-5mm Ground truth : GOM data for 3D information and ground truth for 3D and 2D correspondences.

Descriptor matching success prediction: Synthetic experiments that mimic real world moving planes, with small and big markers to understand the expected outcome. With the given choice of thresholds for descriptor matching.

4.1 Matching comparison

5 Conclusion

Page 7 : 1. Reliable method for correspondence matching for non coplananr circular features. 2. Accuracy is improved with the size and method works better if the detection range is selected appropriate to the conic size. 3. Comment on the stability of the invariants 4. Comment on arrangement of the circles. (Random and non symmetric)

6 Future Applications

Future work : 1. 2D-2D correspondence to triangulate from multiple images to get rid of the initial condition. 2. Improving the matching method for higher runtime performance. 3. Using normals obtained for solving p3p problem.

Each model has its own orientation and random arrangement gives us choice between tracking same models with individual identity.

References

[1] D. Chekhlov, A.P. Gee, A. Calway, and W. Mayol-Cuevas. Ninja on a plane: Automatic discovery of physical planes for augmented reality using visual SLAM. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007. ISMAR 2007*, pages 153–156, 2007.

[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*, page 475–485. Springer, 1990. URL http://link.springer.com/chapter/10.1007/BFb0014897.

[3] 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.

[4] 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.

[5] 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.

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

[7] Stefan Hinterstoisser, Selim Benhimane, and Nassir Navab. N3m: Natural 3d markers for real-time object detection and pose estimation. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, page 1–7, 2007.

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

[9] 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.

[10] David G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, page 1150–1157, 1999.

[11] 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.

[12] 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.

[13] Alain Pagani, Johannes Koehler, Didier Stricker, and Augmented Vision-DFKI. Circular markers for camera pose estimation. 2011.

[14] 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*, page 926–931, 1995. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=466836. 00018.

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

[16] 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.

[17] 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.

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

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

[20] 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*, page 254–257. International Society for Optics and Photonics, 1996. 00002.

[21] Xianghua Ying and Hongbin Zha. Camera calibration using principal-axes aligned conics. *Computer Vision–ACCV 2007*, page 138–148, 2007.