Multimed Tools Appl
DOI 10.1007/s11042-014-2134-8

Bayesian perspective-plane (BPP) with maximum likelihood searching for visual localization

Zhaozheng Hu · Takashi Matsuyama

Received: 8 November 2013 / Revised: 15 April 2014 / Accepted: 26 May 2014
© Springer Science+Business Media New York 2014

Abstract The proposed "Perspective-Plane" in this paper is similar to the well-known "Perspective-n-Point (PnP)" or "Perspective-n-Line (PnL)" problems in computer vision. However, it has broader applications and potentials, because planar scenes are more widely available than control points or lines in daily life. We address this problem in the Bayesian framework and propose the "Bayesian Perspective-Plane (BPP)" algorithm, which can deal with more generalized constraints rather than type-specific ones. The BPP algorithm consists of three steps: 1) plane normal computation by maximum likelihood searching from Bayesian formulation; 2) plane distance computation; and 3) visual localization. In the first step, computation of the plane normal is formulated within the Bayesian framework, and is solved by using the proposed Maximum Likelihood Searching Model (MLS-M). Two searching modes of 2D and 1D are discussed. MLS-M can incorporate generalized planar and out-of-plane deterministic constraints. With the computed normal, the plane distance is recovered from a reference length or distance. The positions of the object or the camera can be determined afterwards. Extensions of the proposed BPP algorithm to deal with un-calibrated images and for camera calibration are discussed. The BPP algorithm has been tested with both simulation and real image data. In the experiments, the algorithm was applied to recover planar structure and localize objects by using different types of constraints. The 2D and 1D searching modes were illustrated for plane normal computation. The results demonstrate that the algorithm is accurate and generalized for object localization. Extensions of the proposed model for camera calibration were also illustrated in the experiment. The potential of the proposed algorithm was further demonstrated to solve the classic Perspective-Three-Point (P3P) problem and classify the solutions in the experiment. The proposed BPP algorithm suggests a practical and effective approach for visual localization.

Keywords Visual localization · Bayesian Perspective-Plane (BPP) · Generalized constraints · Maximum likelihood searching, Perspective-Three-Point (P3P)

Z. Hu (✉)
ITS Research Center, Wuhan University of Technology, Wuhan 430063, People Republic of China
e-mail: zhaozheng.hu@gmail.com

Z. Hu · T. Matsuyama
Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

Published online: 12 June 2014

1 Introduction

Visual localization has important applications in computer vision, virtual/augmented reality (VR/AR), multimedia modeling, human-computer interface (HCI), motion estimation, navigation, and photogrammetric communities, etc. This topic has been intensively investigated for the past decades [2,4,5,8,9,12]. The basic objective is to compute the position of an object or the camera in a reference coordinate system by using either multiple (two or more) or single image (s). For example, stereo vision is a commonly used approach for localization from two images, which can retrieve the 3D coordinates or position of an object in the scene from image disparity, baseline width, and camera focal length [8]. Compared to the multi-view methods, localization from a single view usually relies on some "prior knowledge" of the scene. Based on different scene constraints, there are many algorithms developed for visual localization. However, these algorithms are usually type-specific ones and not generalized enough.

In the literatures, there are many localization methods that try to exploit different scene constraints. For example, model-based localization is a very popular localization approach, which tries to exploit prior knowledge of a "well-known model" to estimate the depth, distance, or relative pose between object and camera. Actually, the model can be represented in different forms, such as a complete 3D map, e.g., a CAD model of the environment, a set of points or lines, which have known coordinates in a reference coordinate system (also known as control points or control lines), a well-defined landmark including structured light patterns, a natural or artificial object, a specific shape, etc. The researches on localization from a set of control points or lines have been formulated as the well-known Perspective-n-Point (PnP) and Perspective-n-Line (PnL) problems in the field of computer vision [6,10,11,15,18,25]. For the past decades, a lot of algorithms and computation models have been successfully developed to address the P3P (n=3), P4P (n=4), P5P (n=5), and P3L (n=3), P4L (n=4), P5L (n=5) problems. The PnP and PnL researches try to address two problems: 1) how to solve the PnP problem; 2) how to classify the multiple solutions. As for $n \geq 6$ cases, PnP and PnL problems have unique solutions. Existing approaches, such as Direct Linear Transform (DLT) [8], have been successfully developed to address these problems. Landmark-based algorithms are another category for visual localization, which have been intensively investigated, especially for mobile robot navigation. For the purpose of easy recognition and position computation, landmarks are usually professionally designed with good features, such as dominant colors, textures, etc., and pre-defined shapes [1,29]. Moreover, some natural or artificial objects in the scene, such as human body, face, eye, window in the street, etc., can also be used for localization. For example, we can estimate the relative position between a camera and a human face from a single image, by using a calibrated 3-D face model [21], which has a lot of application for human-computer interface (HCI). A single image of human eye balls can help to localize the eyes and compute the gaze directions, which have important applications for human attention analysis [19]. Zhu and Ramanan investigated the localization problem from landmark in the wild [30]. Camera pose can be computed from 3D corner, which is widely available in artificial buildings [20]. Urban building and roads can provide rich information for visual localization of mobile devices, which has important applications for navigation [27]. Symmetry is also important clue for localization [23]. More generally, if a planar structure is determined, it is possible to localize object and camera in between. Actually, planar scenes are commonly found in daily life. Hence, localization based on planar structure is very practical and flexible. In existing algorithms, planar structure, such as the plane normal or the distance, can be computed from circular points, parallel lines, homography, conics, control points, deterministic or statistical texture, etc. For example, Hu et al. determine the support plane from the geometric constraints of three control points [10]. If there are four

control points or more on the plane, we can compute the homography and determine the position of the plane by decomposition [28]. It is also feasible to compute the plane normal from vanishing line, if two sets of parallel lines are available in the scene [13]. Planar conics, such as circles, give other important clues for plane computation, which can be interpreted by the circular points [26]. Planar structure can be calculated from video sequences using specific constraints [7]. Pretto et al. investigated the 3D localization of planar objects for industry bin-picking [17]. Textures were investigated for plane shape recovery, which usually assume prior knowledge of statistical characteristics, such as homogeneity, or repetition of specific shapes [24].

From the literatures, it can be found that existing single view algorithms for localization or planar structure computation algorithms are mostly based on some specific scene constraints. They are not generalized or practical enough because of two reasons. On the one hand, some strict scene constraints, which are required by the type-specific algorithm, are sometimes difficult to satisfy in the scene. On the other hand, some constraints, which are available in the scene, are difficulty to be utilized by the type-specific algorithm. As a result, these algorithms have limitations in practice, especially for the occlusion and partially visible situations. For example, the rectangle based method is not feasible if the one edge of the rectangle target is occluded. However, with the proposed method, we can still make use of the three angles between the three remaining edges for visual localization, as the proposed method can exploit generalized scene constraints rather than specific shapes. Hence, the motivation of this paper is to propose a new model that could incorporate generalized scene constraints rather than type-specific ones for visual localization. The proposed is expected to replace or complement existing localization methods by proposing a unified framework to cope with different constraints.

In this paper, we propose a novel localization problem of "Perspective-Plane" to address the issues arisen in the constraint-specific methods in the literatures. "Perspective-Plane" is similar to PnP or PnL problems in computer vision, but with broader applications and potentials, since planar scenes are more widely available in daily life. The "Perspective-Plane" problem deals with the planar object instead of particular control points or lines cases for visual localization. And we propose "Bayesian Perspective-Plane" (BPP) algorithm to solve the "Perspective-Plane" problem. The BPP algorithm can incorporate both planar and out-of-plane deterministic constraints to compute the planar structure and localize the object. By using the Bayesian formula, determination of the planar structure is formulated as a maximum likelihood problem, which is solved by the maximum likelihood searching model (MLS-M) on Gaussian hemisphere. We also present 1D searching mode to simplify and enhance the searching. Localization is accomplished from the recovered planar structure afterwards. The proposed model can also be extended to deal with un-calibrated images and for camera calibration. Since planar scenes are commonly found in daily life, the proposed BPP algorithm is expected to be practical and flexible in many computer vision applications.

The contributions of this paper are summarized as follows: 1) propose the conception of "Perspective-Plane", and address it within the Bayesian framework by proposing the Bayesian Perspective-Plane (BPP) algorithm. The BPP algorithm can deal with more generalized constraints rather than specific ones for visual localization, compared to the traditional PnP and PnL problems; 2) model the likelihood with normalized Gaussian functions and proposed Maximum Likelihood Searching Model (MLS-M) to solve for the plane normal by searching on Gaussian hemisphere. Moreover, we proposed two searching modes of MLS-M, 1D and 2D searching modes, for efficient computation; 3) extend the proposed BPP method to un-calibrated images and camera calibration.

The structure of the paper is organized as follows. Section 2 introduces "Perspective-Plane" geometry and the formulation of planar and out-of-plane geometric constraints from scene prior

knowledge; Section 3 proposes the algorithm of Bayesian Perspective-Plane (BPP) by using generalized constraints and the extensions of the model; Section 4 presents the experimental results with both simulation and real image data. The conclusions are drawn in the final section.

2 Formulation of geometric constraints

2.1 Geometry from a reference plane

Under a pin-hole camera model, a physical plane and its image are mapped by a 3×3 homography [1]

$$x \cong HX \tag{1}$$

where X is for the coordinates of the point on the physical plane, x is for the image coordinates, and ≅ means equal up to a scale. Given the camera calibration matrix and the plane structure, i.e., the plane normal and distance, there are two approaches to compute the homography. One is based on the stratified reconstruction by determining the vanishing line and the circular points [16]. The other is to assume a world coordinate system (WCS) from the physical plane [28]. For example, let the X-O-Y plane in WCS coincides with the physical plane, we try to derive two unit orthogonal directions N_1 and N_2, as the directions of x- and y- axes, from plane normal (i, j=1, 2 and i≠j)

$$\begin{cases} N^T N_{\mathrm{i}} = 0 \\ N_i^T N_{\mathrm{j}} = 0 \\ N_i^T N_i = 1 \end{cases} \tag{2}$$

The third constraint in Eq. (2) forces unit N_1 and N_2. Actually, we cannot uniquely determine N_1 and N_2 from Eq. (2). In practice, we can set N_1 as

$$N_1 \cong \begin{bmatrix} N_z & 0 & -N_x \end{bmatrix}^T \tag{3}$$

where $N = \begin{bmatrix} N_x & N_y & N_z \end{bmatrix}^T$. With N_1, the direction N_2 is computed from Eq. (2) readily. As a result, we can compute the homography as [28]

$$H = \begin{bmatrix} K N_1 & K N_2 & Kt \end{bmatrix} \tag{4}$$

where K is the calibration matrix, t is the translation vector between the WCS and camera coordinate system, and Kt is the image of the origin of the WCS.

With the homography, we can compute the coordinates X on the physical plane from its image point x as $X \cong H^{-1}x$. Hence, a Euclidean reconstruction of the plane is possible. And more geometric attributes, such as distance, angle, length ratio, curvature, shape area, etc., are calculated readily. Note that a metric reconstruction of the plane is also feasible with known plane normal and unknown distance. However, all the absolute geometric attributes, such as coordinates, distance, etc., are computed up to a common scale to the actual ones. And some non-absolute geometric attributes, such as length ratio, angle, etc., are equal to the actual ones. This is because the metric reconstruction is up to a similarity transform with the Euclidean one.

It is also feasible to compute the geometric attributes associated with some specific out-of-plane features by referring to a known plane. For example, Wang et al. show that an orthogonal plane can be measured [22], if a reference plane is well determined. Criminisi et al. show that if the vanishing line of a reference plane and one vanishing point along a reference direction are

known, object height, distance, length ratio, angle, and shape area on the parallel planes etc., are readily computed [3]. From projective geometry, both the vanishing line of the plane and the vanishing point along the vertical direction can be computed from the plane normal and the camera calibration matrix. Hence, the out-of-plane geometric attributes, as discussed above, can be computed from a determined planar structure and camera calibration matrix.

2.2 Geometric constraint formulation

Prior knowledge of the scenes can generate geometric constraints on the planar structure. We classify the scene constraints into two categories: planar and out-of-plane constraints. Planar constraints are derived from prior knowledge of the geometric attributes of planar features on the plane, while the out-of-plane constraints are from the out-of-plane features. The planar features are those that lie on the plane, such as the point, line, angle, curve, shape, etc. (see Fig. 1 (a)), while the other are defined as out-of-plane features, such as an orthogonal line, an angle on parallel plane, etc. (see Fig. 1 (b)). Formulation of both planar and out-of-plane constraints is introduced as follow.

Assume a calibrated camera so that the computation of the geometric attributes relies on the planar structure only. Actually, the plane normal can determine the plane up to a scale, which allows the computation of a lot of non-absolute geometric attributes, such as the relative length, angle, etc., without knowing the plane distance. Hence, we only consider the non-absolute geometric attributes to calculate the plane normal. The plane distance is computed readily from reference length, once the plane normal is determined, as will be discussed in section 3.4.

Let $Q_i(N)$ be the geometric attribute computed from a given plane normal N. Let u_i be the deterministic value associated with such geometric attribute that we know a prior. The difference is defined as the measurement error associated with the normal N

$$d_i(N) = Q_i(N) - u_i \tag{5}$$

By forcing zero measurement error, we can derive one constraint C_i as

$$C_i : d_i(N) = Q_i(N) - u_i = 0 \tag{6}$$

Note that Eq. (6) holds for both planar and out-of-plane constraints. Similarly, we can derive a set of geometric constraints on plane normal

$$C = \left\{ C_i \middle| d_i(N) = Q_i(N) - u_i = 0 \right\} \quad (i = 1, 2, \cdots M) \tag{7}$$

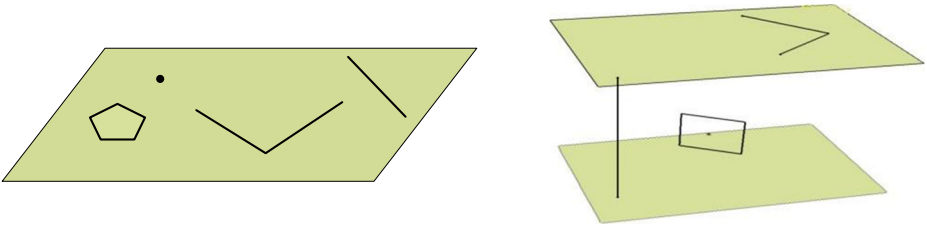

Fig. 1 a) planar features, e.g., control point, control *line*, known *angle*, specific shape, etc., on the plane; **b**) Out-of-plane features that lie out of the plane, e.g., orthogonal *line*, known angle on a *parallel* plane, point on an orthogonal plane, etc

The plane normal can be computed by solving Eq. (7), which is, however, not an easy task. Existing algorithms are usually based on some specific constraints to solve Eq. (7), and not generalized or practical enough. A generalized algorithm, "Bayesian Perspective-Plane (BPP)" is discussed as follows.

3 Bayesian perspective-plane (BPP)

3.1 Localization from planar structure

Localization is possible from single image of a planar structure, i.e., with known plane normal and distance, as illustrated in Fig. 2. The point on the plane can be localized by calculating the intersection of the back-projection ray and the plane as follows

$$\begin{cases} X = \lambda K^{-1} x \\ X^T N = d \end{cases} \tag{8}$$

where the first equation in Eq. (8) defines the back-projection ray by using the camera calibration matrix and the image (see the dash line in Fig. 2), while the second one is the equation for a known plane (N, d). In a similar way, we can localize all the points on the plane. If we use the plane to expand a reference coordinate system, the position of the camera can be computed as well [28]. Hence, we can localize the object and camera in-between. And the key is to determine the plane normal and distance.

3.2 Formulation of Bayesian perspective-plane (BPP)

We assume the plane normal N with certain distributions in the searching space. For example, if no geometric constraints are given, N is uniformly distributed, as shown in Fig. 3 (a). Given one geometric constraint, the distribution of the normal ($P(N|C_i)$) is updated from uniform to non-uniform, as shown in Fig. 3 (b). Given a set of geometric constraints, the plane normal has a new distribution ($P(N|C_1, C_2, \cdots C_M)$) with a dominant peak (if we have sufficient constraints). And the plane normal with the highest probability

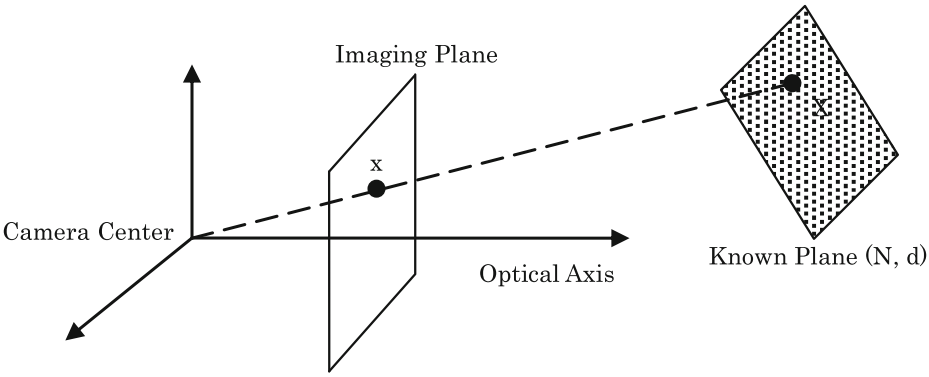

Fig. 2 Localize a 3D point X in the camera coordinate system from single view of a known plane by computing the intersection with the back-projection ray

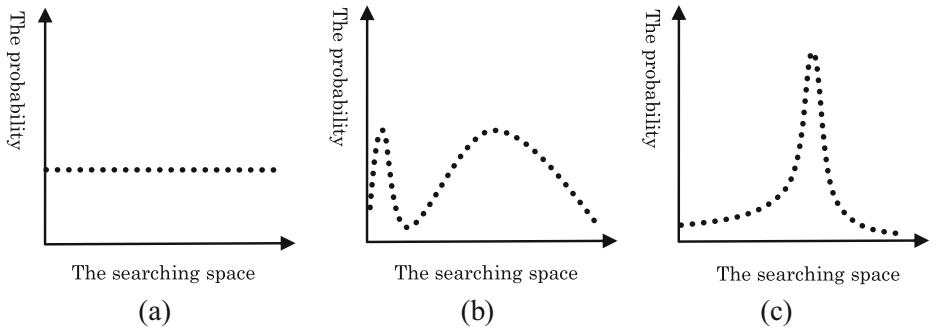

Fig. 3 Distributions of plane normal in different conditions: **1**) Uniform distribution ($P(N)$) with no constraints; **2**) Non-uniform distribution ($P(N|C_i)$) from one geometric constraint; **3**) Distribution ($P(N|C_1, C_2, \cdots C_M)$) with a dominant peak from a set of geometric constraints

gives the solution (see Fig. 3 (c)). Note that we use probability instead of probability dense function model (P.D.F.) throughout the paper, because we need to partition Gaussian hemisphere to discrete the space for maximum likelihood searching.

Based on the above consideration, computation of the plane normal from a set of geometric constraints is formulated as to maximize the following conditional probability

$$N^* = \arg\max_N P\left(N \middle| C_1, C_2, \cdots C_M\right) \tag{9}$$

However, it is difficult to solve Eq. (9) directly. By using Bayesian formula, we can transform Eq. (9) and derive the following equation

$$P\left(N \middle| C_1, C_2, \cdots C_M\right) = \frac{P\left(C_1, C_2, \cdots C_M \middle| N\right) P(N)}{P(C_1, C_2, \cdots C_M)} \tag{10}$$

where $P(C_1, C_2, \cdots C_M|N)$ is the likelihood. $P(N)$ and $P(C_1, C_2, \cdots C_M)$ are the prior probabilities for the plane normal and geometric constraints, respectively. Assume that the constraint $C_i (i = 1,2,\ldots M)$ is conditionally independent to each other, so that

$$P(C_1, C_2 \ldots C_M|N) = \prod_{i=1}^{M} P(C_i|N) \tag{11}$$

Assume that the plane normal is uniform distribution so that $P(N)/P(C_1, C_2, \cdots C_M)$ is a constant. We can substitute Eq. (11) into Eq. (10) to derive the relative probability

$$P\left(N \middle| C_1, C_2, \cdots C_M\right) \propto P\left(C_1, C_2, \cdots C_M \middle| N\right) = \prod_{i=1}^{M} P\left(C_i \middle| N\right) \tag{12}$$

Substituting Eq. (12) into Eq. (9) yields

$$N^* = \arg\max_N \; P\left(N \middle| C_1, C_2, \cdots C_M\right) \propto \arg\max_N \prod_{i=1}^{M} P\left(C_i \middle| N\right) \tag{13}$$

Therefore, to solve Eq. (13) is equivalent to compute the normal, which yields the maximum joint likelihood. In order to solve Eq. (13), we need to model $P(C_i|N)$,which is defined as the likelihood that the i^{th} constraint is satisfied, given

the plane normal N. In order to develop a reasonable and accurate model, we develop the following rules

1) The likelihood is determined by the measurement error. More the absolute measurement error yields lower the likelihood is. The maximum likelihood is reached when the measurement error is zero;
2) The maximum likelihoods (for zero measurement error) for different constraints should be equal so that all constraints contribute equally to solve Eq. (13);
3) The measurement error should be normalized to deal with the geometric attributes with different forms, units, and scales, etc.

To serve these purposes, we proposed the normalized Gaussian function to model the likelihood as

$$G\left(d_i(N)\middle|u_i,\sigma_i\right)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(Q_i(N)-u_i)^2}{2u_i^2\sigma_i^2}\right)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{d_i^2(N)}{2u_i^2\sigma_i^2}\right) \tag{14}$$

where u_i from Eq. (5) is used to normalize the measurement error. Note that Eq. (14) expects non-zero u_i. Otherwise, the measure error is not normalized. It can be observed from Eq. (14) that the first rule is well satisfied. The likelihood decreases with the absolute measurement error. And the maximum likelihood is reached for zero measurement error. In order to satisfy the second rule, the standard deviations for all Gaussian models should be equal ($\sigma_i=\sigma_j=\sigma$). In practice, we can choose appropriate σ. As a result, the likelihood model for each constraint is derived as

$$P\left(C_i\middle|N\right)=G\left(d_i(N)\middle|u_i,\sigma\right)\propto\exp\left(-\frac{d_i^2(N)}{2u_i^2\sigma^2}\right) \tag{15}$$

Once we model the likelihood function, we can re-organize Eq. (9) by substituting Eq. (15) into Eq. (12) and derive the following equation

$$P(N|C_1,C_2,\cdots C_M)\propto\prod_{i=1}^{M}G(d_i(N))\propto\prod_{i=1}^{M}\exp\left(-\frac{d_i^2(N)}{2u_i^2\sigma^2}\right) \tag{16}$$

Finally, we can derive the following equation to compute the plane normal

$$N^*=\arg\max_N P(N|C_1,C_2,\cdots C_M)\propto\arg\max_N\exp\left(-\sum_{i=1}^{M}\frac{d_i^2(N)}{2u_i^2\sigma^2}\right) \tag{17}$$

3.3 Maximum likelihood searching model (MLS-M)

A maximum likelihood searching model (MLS-M) is proposed to solve Eq. (17). And two searching modes of 2D and 1D are proposed in the following paragraph.

a. 2D searching mode

A unit plane normal corresponds to a point on the Gaussian sphere surface. Hence, the Gaussian sphere defines the searching space. In practice, we search on Gaussian hemisphere to reduce the searching space into half, because only the planes in front of the camera are visible

in the image. Once we define the searching space, we need to partition the space for plane normal sampling. A uniform sampling can be realized by evenly partition Gaussian hemisphere into a number of cells. Some algorithms are developed to serve this purpose. And in this paper, we used the recursive zonal equal area sphere partitioning algorithm in [14]. After partition, the center of each cell represents a sampled normal. The likelihood for each sampled normal is computed by using Eq. (17). Once the likelihoods for all the samples are calculated, the maximum likelihood is computed by sorting. The corresponding plane normal is the solution. Since Gaussian sphere in 3D space is two dimensional (2D), it is called the 2D searching mode.

b. 1D searching mode

A linear constraint can reduce 2D searching into 1D to enhance the searching efficiency. In practice, some linear constraints are widely available in the scene, such as parallel lines, orthogonal lines or planes, etc. For example, a linear constraint on the plane normal can be formulated from an orthogonal line. Let L be the orthogonal line, with its image l. Because L is parallel to the plane normal, the vanishing point V_{normal} along the normal direction lies on l and satisfies the following equation

$$l^T V_{normal} = l^T (KN) = \left(K^T l\right)^T N = 0 \tag{18}$$

The term $K^T l$ represents a re-projection plane passing through the camera center and the image l, on which the normal vector lies. Since N also lies on the Gaussian sphere, we can derive

$$\begin{cases} \left(K^T l\right)^T N = 0 \\ N^T N = 1 \end{cases} \tag{19}$$

Eq. (19) above defines a circle in 3D space, which is the intersection between the plane and the Gaussian sphere (see Fig. 4).

We can thus search on a 1D circle (actually half of the circle) instead of 2D Gaussian hemisphere. A uniform sampling is implemented to discrete the space (see APPENDIX A). Similarly, we compute the likelihood for each sample normal and search for the maximum likelihood to compute the plane normal. Other linear constraints can also help reduce 2D into

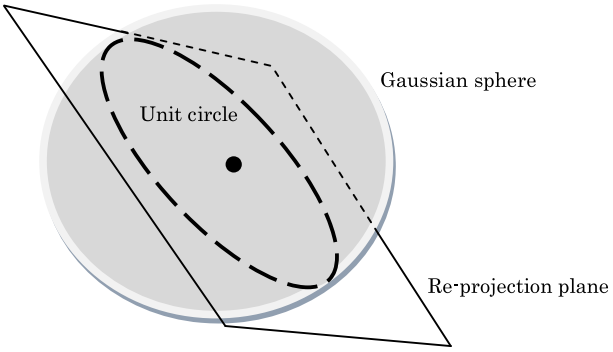

Fig. 4 1D searching on a unit *circle* in 3D space, which is the intersection of Gaussian sphere and a plane

1D searching by following the similar steps, such as a vanishing point on the plane, an orthogonal plane, etc.

3.4 Plane distance determination and localization

Assume unit plane distance and re-write Eq. (8) as follows

$$\begin{cases} X^{'} = \lambda K^{-1} x \\ X^{'T} N = 1 \end{cases} \tag{20}$$

By comparing Eq. (8) and Eq. (20), the computed coordinates are equal to the actual ones up to a scale: $X' \cong X$. As a result, we can compute the 3D coordinates for all the points on the plane, and localize them in the camera coordinate system, all up to a common scale.

In order to get the actual coordinates and distances, we need a reference length or distance. Assume the actual length L_{AB} between two points A,B is known, the scale factor can be determined by

$$\beta = \frac{L_{AB}}{\left\| X^{'}_{A} - X^{'}_{B} \right\|} \tag{21}$$

where $X^{'}_{A}$ and $X^{'}_{B}$ are the 3D coordinates of A,B from Eq. (20) by assuming unit plane distance and $\| X^{'}_{A} - X^{'}_{B} \|$ is the distance. Once the scale factor is determined, the actual plane distance is $d=\beta$. We can compute the actual coordinate for localization as

$$X = \beta X^{'} \tag{22}$$

3.5 Extensions of the BPP algorithm

So far we assume calibrated camera so that the geometry computation relies on the planar structure only. However, the BPP algorithm can be extended to deal with un-calibrated images by assuming minimum camera calibration, i.e., with zero skew, unit aspect ratio, central principle point. The geometry computation is determined by both the plane normal and the focal length. Hence, the measurement error is given by

$$d_i(N, f) = Q_i(N, f) - u_i \tag{23}$$

As a result, we re-formulate Eq. (9) by incorporating both plane normal and focal length:

$$(N^*,\ f^*) = \arg\max_{N,f} P\left(N, f \middle| C_1, C_2, \cdots C_M\right) \tag{24}$$

The focal length is assumed independence to the normal with uniform distribution. We can use the Bayesian formula to transform Eq. (24) by applying the Gaussian likelihood model as follows

$$(N^*,\ f^*) = \arg\max_{N,f} P\left(N, f \middle| C_1, C_2, \cdots C_M\right) \propto \arg\max_{N,f}\ \exp\left(-\sum_{i=1}^{M} \frac{d_i^2(N, f)}{2u_i^2 \sigma^2} \right) \tag{25}$$

In order to compute the maximum likelihood, we need to search for both the plane normal and the focal length. Hence, a 3D searching is required. However, as we discussed in the above

paragraphs, if there is one linear constraint on the plane normal available, the 3D searching can be reduced into 2D.

A further extension of the proposed model is for camera calibration. We start from the assumption that the plane vanishing line is known, the same assumption as used in [7,13]. We can compute the plane normal from the vanishing line, given the camera focal length, and compute the geometric attributes. As a result, we formulate the camera calibration as to maximize the following conditional probability

$$f^* = \arg\max_f P\left(f \middle| C_1, C_2, \cdots C_M\right) \tag{26}$$

Once again, we can apply the Gaussian likelihood model and derive the following equation for camera calibration (focal length computation)

$$f^* = \arg\max_f P\left(f \middle| C_1, C_2, \cdots C_M\right) \propto \arg\max_f \exp\left(-\sum_{i=1}^{M} \frac{d_i^2(f)}{2u_i^2\sigma^2}\right) \tag{27}$$

where the measurement error for the corresponding focal length is given by

$$d_i(f) = Q_i(f) - u_i = 0 \tag{28}$$

Note that Eq. (27) can also incorporate both planar and out-of-plane constraints for camera calibration, and is more generalized than the algorithms proposed in [7,16]. The proposed camera calibration method is illustrated in the following experiment.

4 Experimental results

The proposed BPP algorithm was tested with both simulation and real image data by using planar and out-of-plane constraints. Extension of the proposed model for camera calibration was also tested. Furthermore, the proposed model was applied to solve the well-known Perspective-Three-Point (P3P) problem in the real image test. Although the proposed algorithm is designed to deal with generalized deterministic constraints, it is not feasible to discuss all types of constraints in the experiments. Without loss of generality, we chose two types of constraints: 1) known length ratio; 2) known angle, which are also two fundamental geometric attributes in measurement.

4.1 Results with simulation data

A simulated pin-hole camera was used to generate the image data. The camera has the focal length of 1800 pixels, unit aspect ratio, zero skew, and principle point at [800, 1000] ^{T} (in pixel). Random Gaussian noises with zero mean and standard deviations of half pixels were added to the image coordinates to model the practical image noises.

The BPP algorithm was first tested by using planar geometric constraints. Three geometric constraints were used: 1) two constraints from two known angles of 10 and 20°; 2) one constraint from known length ratio (valued 2) of two segments. A physical plane with the normal [−0.5612, 0.3333, 0.7576] ^{T} and a distance of 72 cm was projected onto the imaging plane by the simulation camera. The MLS-M model was applied to compute the plane normal. The 2D searching mode was applied and the Gaussian hemisphere was partitioned into 200× 400 cells, with each cell center representing one sampled plane normal. The likelihood for each sampled normal was computed with Eq. (15) by using the three geometric constraints. Due to

the Gaussian hemisphere sampling, we may not derive the exact normal. Hence, we define a "best" normal that has minimum angle with the actual one. In this test, the "best" normal is $[-0.5641\ 0.3336\ 0.7553]^T$, which has the minimum angle of 0.21° with the actual one.

Figures 5 (a)~(c) are for the likelihoods from the three constraints respectively, where the image intensity represents the likelihood, with high intensity for high likelihood. It can be observed that the likelihood of the plane normal is non-uniform due to the different geometric constraints. The joint likelihood from the three constraints is presented in Fig. 5 (d), from which we can observe a dominant area with high likelihoods (see the bright area on the bottom right). Fig. 5 (e) shows the positions of the 30 normal with the highest likelihoods. And the maximum likelihood was searched in the joint likelihood map, with the position marked by '+' (see Fig. 5 (f)). The corresponding plane normal was computed as $[-0.5590\ 0.3306\ 0.7604]^T$. In Fig. 5 (f), the position of the best normal was also marked by 'o'. From the computation results, the computed normal from the proposed model has 0.25 and 0.45° with the actual and the best normal, respectively. The results show that MLS-M is accurate.

The second simulation experiment was performed to test the proposed model by using both planar and out-of-plane constraints. The same simulation camera was used. The physical plane has the normal direction $[-0.2374,\ 0.9102,\ 0.3322]^T$ with 167 cm distance to the camera center. Three geometric constraints were used for plane structure computation and localization. Among them, one planar constraint is from a known angle (45°). Two out-of-plane constraints are derived from: 1) one line orthogonal to the plane; 2) a known length ratio (0.5) between two segments on a parallel plane.

The 2D searching mode was first applied to compute the likelihood and calculate the plane normal by searching for the maximum likelihood. The Gaussian hemisphere was also partition into 200×400 cells. Figure 7 shows the likelihoods for each sampled normal from different constraints. Among them, Fig. 7a and c are for the likelihoods generated from the three individual constraints, which clearly show the non-uniform distributions for the plane normal. The joint likelihood from the three constraints is presented in Fig. 7 (d), from which we can clearly observe a dominant bright area on the upper right corner. It shows that the plane normal has dominant peak distribution due to three constraints. The positions of the 30 plane normal with the highest likelihoods are shown in Fig. 7 (e) to highlight the area. From the joint likelihood map in Fig. 7 (f), the maximum likelihood was computed with the corresponding

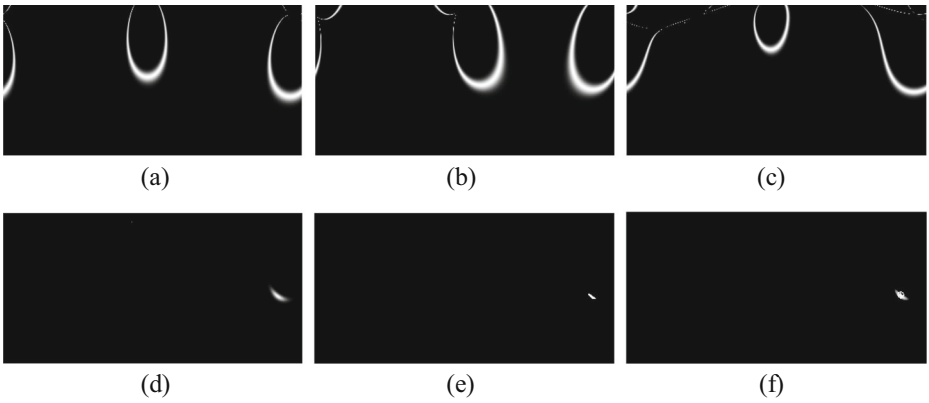

Fig. 5 (**a**) (**b**) (**c**) the likelihoods from three geometric constraints, respectively, (**d**) the joint likelihood, (**e**) the positions of the 30 plane normal with the highest likelihood; (**f**) positions of the computed and the best normal marked by cross '+' and circle 'o', respectively

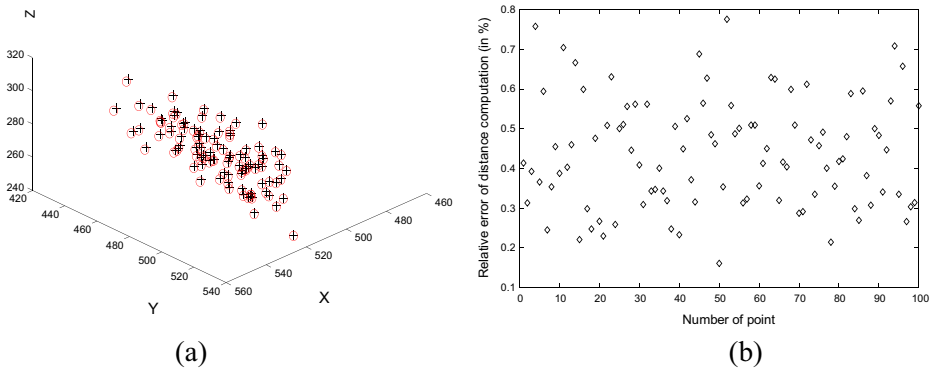

Fig. 6 Localization results: **a**) 3D positions (*marked by '+'*) of 100 points on the plane in the camera coordinate system (in cm), with the actual positions marked by 'o'; **b**) Relative errors for point-to-camera distance computation

normal position marked by '+'. We also marked the position of the "best" normal by circle. It can be observed that they are in the same position. Hence, the proposed model calculated the "best" normal in this test.

From the orthogonal line constraint, 1D search mode is feasible to search for the maximum likelihood. We first defined the circle to search by using Eq. (17). Afterwards, we sampled 10,000 points within the searching space (half of the circle in 3D space). The likelihoods were computed for each sampled normal by using Eq. (15) and Eq. (17). Fig. 8a and b illustrate the likelihoods computed from the planar constraints of known angle, and out-of-plane constraints of known length ratio on the parallel plane. We can observe multiple peaks in both likelihood curves, which indicate multiple solutions from the individual constraint. More knowledge on multiple solutions is referred to [26]. Fig. 8 (c) shows the joint likelihood, from which we can observe a unique peak. The plane normal was then calculated as $[-0.2474\ 0.9101\ 0.3313]^{T}$, which has 0.03 and 0.01° with the actual and "best" normal, respectively.

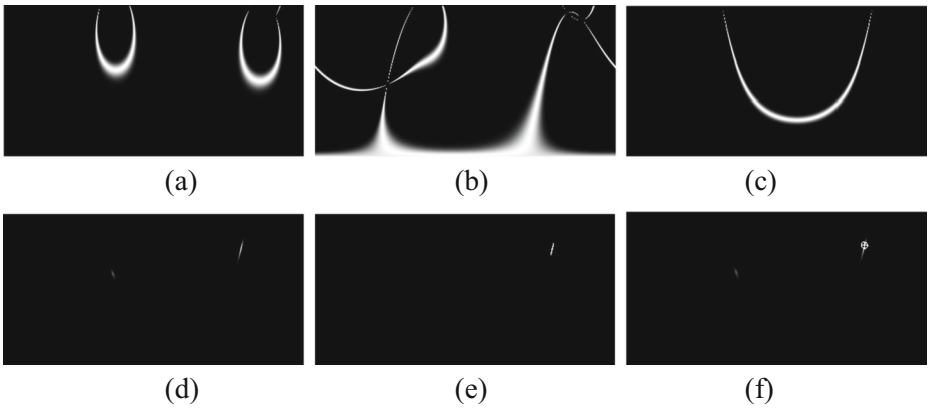

Fig. 7 (**a**) (**b**) (**c**) likelihoods from three geometric constraints of known *angle*, known *angle*, and known length ratio, (**d**) joint likelihood, (**e**) positions of the 30 plane normal with the highest likelihood; (**f**) positions of the computed and the best normal marked by cross '+' and circle 'o', respectively

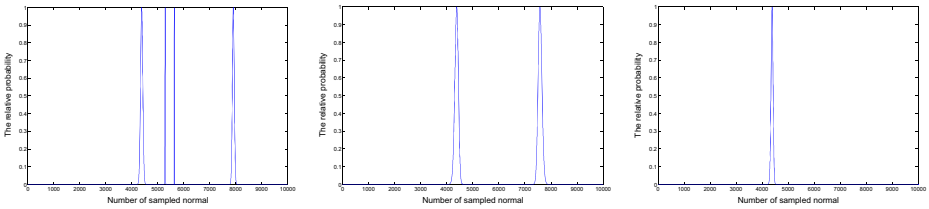

Fig. 8 1D searching mode with the likelihoods from the two constraints (**a**-**b**), and the joint constraints (**c**)

Once the plane normal was calculated, we computed the plane distance by referring to a known length on the plane. And the points on the plane were localized afterwards. Figure 9 (a) shows the reconstructed 3D positions of one hundred points. Figure 9 (b) shows the relative errors for the point-to-camera distance computation. It can be observed that relative distance computation errors are less than 0.8 %. And the mean of the relative errors is 0.7 %. Compared to results in Fig. 6, the mean of the relative error increases by 0.3 %. This is because the plane normal has more tilted angle with the camera optical axis and the sampled points are further from the camera.

Finally, the proposed BPP algorithm was compared to the existing localization method with simulation data. We typically chose the classic homography-based method for comparison [28], which is also the foundation of many other localization methods. For the purpose of fair judgment, we used the same constraints for both methods, which are four control points with known coordinates on a reference plane. The homography-based method was realized as follows. First, the homography between the reference plane and its image was computed from the four control points linearly by using the DLT method [8,28]. Second, the rotation and translation were calculated from the computed homography and camera calibration matrix. And finally the normal of the reference plane was calculated from the rotation. In order to robust estimate the homography matrix, the coordinates of the control points were normalized. As for the proposed BPP method, we derived six length ratios and four angles from the control points as the constraints. Moreover, we partitioned the Gaussian hemisphere into 250×500 cells. In the test, the images of the control points were noised with random zero-mean Gaussian

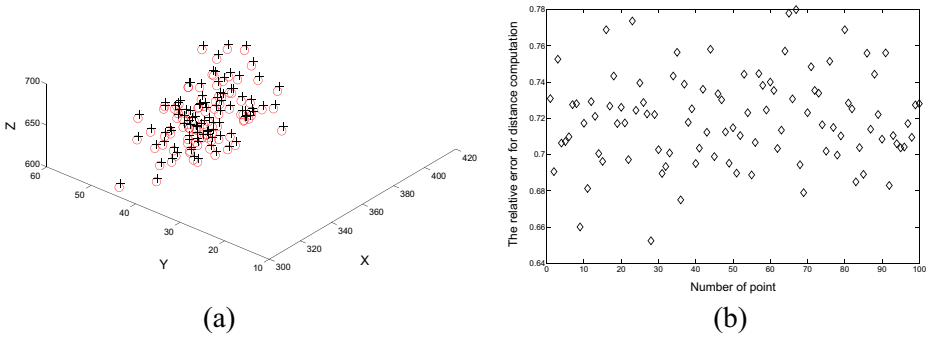

Fig. 9 localization results: **a**) 3D positions (*marked by '+'*) of 100 points in the camera coordinate system (in cm), with the actual positions marked by 'o'; **b**) Relative errors for point-to-camera distance computation

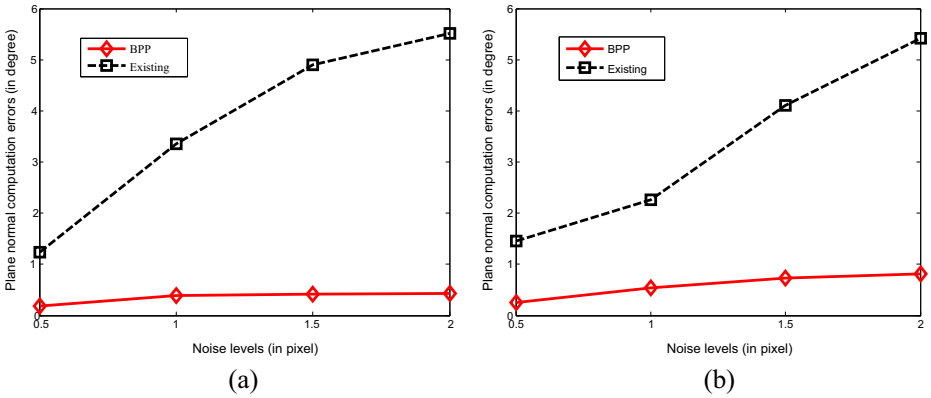

Fig. 10 comparison of the proposed BPP with the homography-based method: **a**) the mean computation errors; **b**) standard deviations of the computation errors

noises. And four noise levels (or the standard deviation of the noises) were set as 0.5, 1.0, 1.5, and 2.0 (in pixel). Both the proposed BPP and homography-based methods were tested against different noise levels. In our test, we used the angle between the computed plane normal and the actual normal as the error to evaluate the performance of the methods. For each noise level, we ran 20 trials for both methods and computed the means and standard deviations. The results are illustrated in Fig. 10, where Fig. 10 (a) is for the means and Fig. 10 (b) is for the standard deviations for both methods.

As can be observed from Fig. 10, the proposed method performs better than existing homography-based method. Especially, with the noise level increasing, the proposed BPP method is more accurate and robust than the homography-based one. For example, when the noise level is as high as 2.0 pixels, the mean of BPP has about 0.5° angle, while the existing homography-based one has 5.5 mean degrees. The corresponding standard deviation of BPP is less than 1.0°, while the homography-based one has standard deviation of 5.5°. The results demonstrate the proposed BPP is more accurate and robust than existing homography-based method.

Fig. 11 Localization from planar constraints: **a**) original image; **b**) three constrains (two known angles and one length ratio) from extracted lines and points

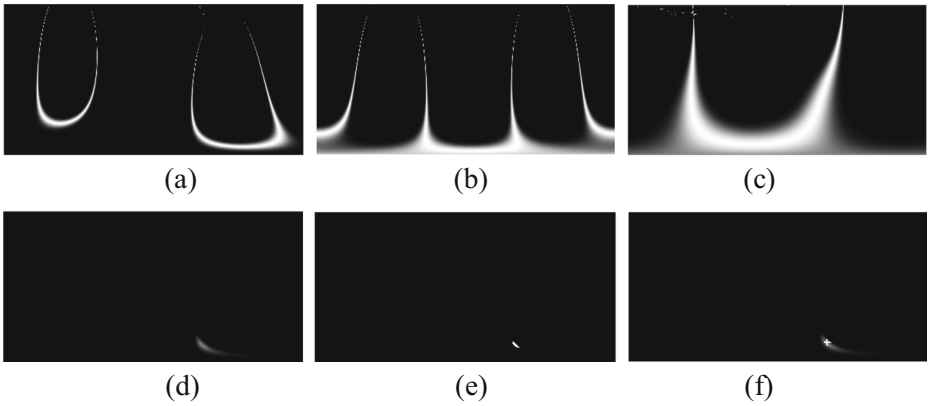

Fig. 12 (**a**) (**b**) (**c**) likelihoods from three geometric constraints, (**d**) the joint likelihood from the three constraints, (**e**) positions of the 30 normal with the highest likelihood; (**f**) position of the maximum likelihood marked by '+'

4.2 Results with real image data

A Nikon700 digital camera was used to generate the real image data in the first three real image experiments. The images have the resolution of 2218×1416 (in pixel). The camera was calibrated with Zhang's calibration algorithm. The calibration results show the camera has 1369.2 (in pixel) focal length, 1.001 aspect ratio, zero skew, and principle point at [1079, 720] ^{T} (in pixel). The calibration results were used for the planar structure recovery and localization. The focal length was also used as ground truth to validate the proposed calibration algorithm. Note that we assume that the image features such as lines, points, corners, etc., were well extracted using image processing techniques in the following tests.

The first real image experiment was performed to compute the plane structure and localize the points using planar constraints. Figure 11 (a) shows an original image of a white A4-size paper, attached on a wall. Six black line segments were drawn on the paper, from which we

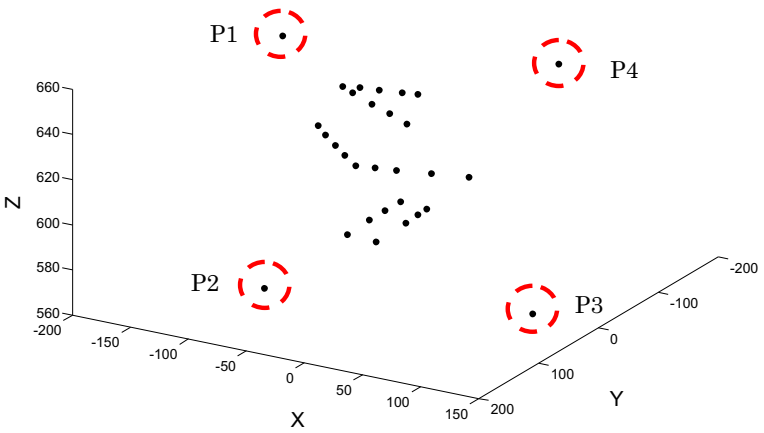

Fig. 13 localize *points* and *lines* in the camera coordinate system (unit in mm)

Fig. 14 **a**) original image of a book for localization; **b**) three lines (L1, L2, L3) expand three planes

derived two angles from the lines L1-L2 and L3-L4, and one length ratio between the two segments of M1-M2 and M3-M4 (see Fig. 11 (b)). They were formulated into three planar constraints.

The MLS-M model was then applied to compute the plane normal. The likelihood map from each individual constraint is presented in Fig. 12 (a)-(c), which clearly shows the non-uniform distribution of plane normal due to different constraints. The joint likelihood map from the three constraints is given in Fig. 12 (d), in which a bright area on the bottom can be observed. The bright area is further highlighted in Fig. 12 (e) by showing the positions of the plane normal with the highest likelihoods. The position of the maximum likelihood in the joint likelihood map is marked by a cross '+' (see Fig. 12 (f)). The corresponding plane normal is $[0.0993\ 0.1679\ 0.9808]^{T}$.

The plane distance was computed by referring the length of P1-P2 (297 mm) in Fig. 13). All the points were localized in the camera coordinate system. Figure 13 shows the localization of the points on the six line segments, and the four corners points of the paper (P1, P2, P3, and P4, marked with circle 'o'). The computed results were validated as follow. We used the calculated 3D coordinates to compute the length of P2-P3, P3-P4, and P1-P4, which are 206.7 mm, 294.6 mm, and 211.8 mm, respectively. According to the ground truth (the standard A4-size paper of 210 mm×297 mm), the absolute errors are 3.3 mm, 2.4 mm, and 1.8 mm, corresponding to 1.6 %, 0.8 %, and 0.9 % relative errors. It demonstrates that the computation results of plane normal computation and localization are accurate.

A second real image experiment was performed to test the proposed model by using both planar and out-of-plane constraints with 1D searching mode. As shown in Fig. 14 (a) and

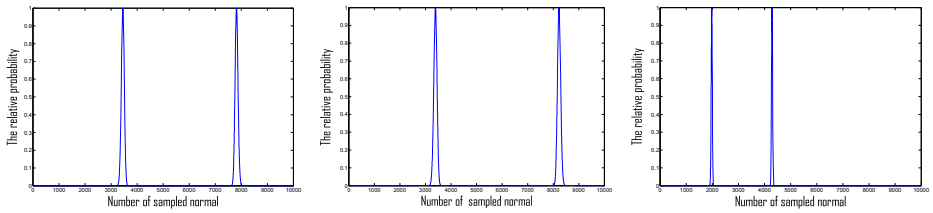

Fig. 15 The likelihoods for the normal of the three planes defined by: **a**) L1-L2; **b**) L1-L3; **c**) L2-L3

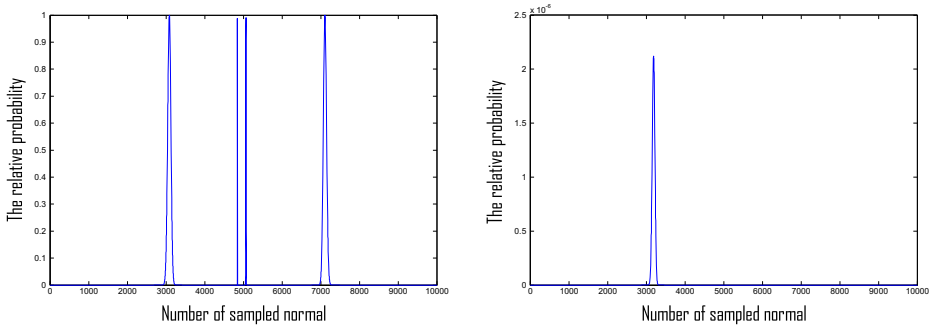

Fig. 16 Unique solution by adding one constraint from length ratio: **a**) the likelihood from the length ratio constraint; **b**) the joint likelihood from two constraints with one peak point

Fig. 14 (b), there is an orthogonal corner structure on the book. And every two orthogonal lines define a plane and we totally have three orthogonal planes. And for each plane, there are two constraints: 1) an orthogonal line to the plane; 2) a right angle on the plane. The orthogonal line constraint allows the 1D searching mode to compute the plane normal.

We applied the proposed MLS-M model to compute the normal directions for these three planes from two constraints for each plane. The plane normal was computed by 1D searching mode. We uniformly sampled 10,000 points within the searching space (half of the circle in 3D space), with each point representing a sampled normal. Hence, the angle between two neighboring sampled normal is 0.0018°. The likelihood for each sample normal was computed with Eq. (17). The results for the normal computation of the three planes are shown in Fig. 15.

One important phenomenon, the multiple solutions to each plane, can be observed in Fig. 15 (a)~(c). In each likelihood curve, there are two peaks to indicate two solutions to the plane normal. More about the multiple solutions for pose computation can be found in [20]. And the results with the proposed model comply with conclusions in the literature [20]. Hence, we totally have eight combinations for the three orthogonal planes, with two of them satisfying the mutually orthogonal constraints. And the proposed algorithm can classify and yield all the

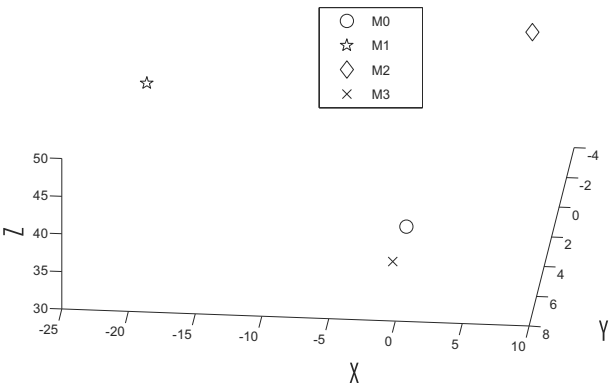

Fig. 17 Localize the points in the camera coordinate system (unit in cm)

Fig. 18 Camera calibration from generalized constraints. *Left*: original image. *Right*: two constraints generated from extracted *line* (L1) and points (M1, M2, and M3)

possible solutions. The algorithm can complement existing non-linear iteration methods by providing good initial guess for result refinement.

To remove the ambiguity, we added one more constraint to uniquely define the normal of the plane L1-L2, which is the length ratio (valued 1.5) of the length (M0-M1) and width (M0-M2) of the book. The computation results are illustrated in Fig. 16 (a). The joint likelihood (see Fig. 16 (b)) was then computed from the two constraints by combing the results from Fig. 15 (a) ad Fig. 16 (a). From the joint likelihood, we searched for the maximum likelihood and derived a unique normal to the plane defined by L1 and L2, which is $[-0.0022\ 0.8425\ 0.5387]^T$. Once the L1-L2 plane normal is determined, the normal for the other two planes can be uniquely computed. We used one actual length to compute the plane distance and localize the corners of the book in the camera coordinate system. The positions are shown in Fig. 17.

We validated the computation results based on the fact that the three lines, defined by M0-M1, M0-M2, and M0-M3, are mutually orthogonal. We calculated the three angles by using the calculated 3D coordinates of the points in Fig. 17. The computed angles are 89.4, 88.1, and 94.1°, respectively, with the absolute errors of 0.6, 1.9, and 4.1°. The relative errors are 0.7 %, 2.1 %, and 4.6 %, with the average 2.4 %. The results demonstrate that the BPP algorithm is accurate for visual localization. We also localized the camera in the reference coordinate system, defined by the corner structure, i.e., with M0 the origin and L1, L2, L3 as the three axes, by simple coordinate system transformation. The result is $[-21.64, -23.82, -12.26]^T$ (in cm). Hence, the object and the camera can be localized from each other.

Another experiment was performed to extend the BPP algorithm to camera calibration. As shown in Fig. 18, there is a rectangle-shape mouse pad on the desktop, from which the

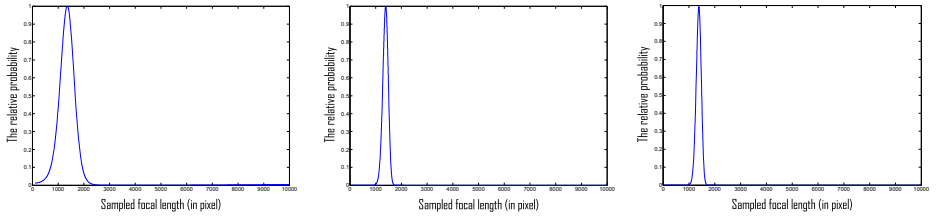

Fig. 19 Likelihoods of focal length from: **a**) length ratio; **b**) orthogonal line constraint; **c**) joint constraints

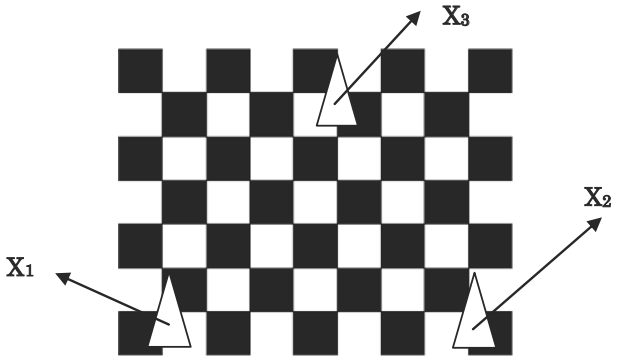

Fig. 20 Three corner points (non co-linear) randomly chosen from a chessboard pattern as the three control points

vanishing line of the desktop plane was computed as $[-0.0006\ 0.0015\ -1.0000]^{T}$. In order to test the calibration algorithm with generalized constraints, we chose two representative constraints: 1) one out-of-plane constraint that L1 is orthogonal to the desktop plane; 2) one planar constraint from the length ratio (valued 1.2) between the two segments of M2-M3 and M1-M2.

The searching space for focal length was set from 100 to 10000 pixels. The corresponding horizontal view angles range from 169 to 13° (See APPENDIX B), which covers a big range from the wide angle to high resolution lens. The partition resolution is one pixel. The likelihood for each sampled focal length was computed by using Eq. (27). Fig. 19 shows the likelihoods, generated from the individual and joint constraints. It can be observed that there is one peak in each curve, which demonstrates that focal length can be computed from one constraint. The peak point was searched in the joint likelihood and the corresponding focal length is 1388 pixels. Compared with the calibration results, the absolute error is 18.8 pixels, corresponding to 1.4 % relative error. The result demonstrates the proposed model can be extended for camera calibration by exploiting different constraints.

In the last experiment, the proposed BPP algorithm was applied to solve the well-known P3P problem. We define a support plane from the three control points (see Fig. 20). It can be proved that the determination of the support plane is the necessary and sufficient condition to solve the P3P problem (see APPENDIX C). The three point-to-point distances from the control points allow the computation of three angles and three length ratios as the geometric constraints, with which we applied MLS-M model to compute the support plane, and localize the three control points.

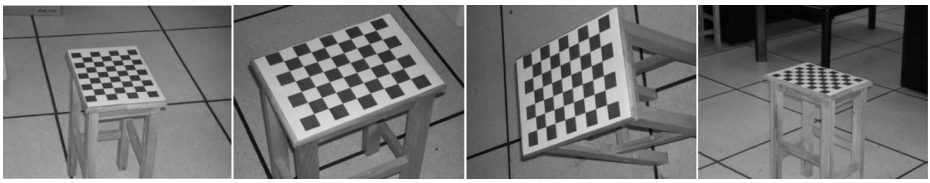

Fig. 21 Images of chessboard pattern for camera calibration and for plane normal recovery with the proposed algorithm, from *left* to *right* numbered as 1, 2, 3, 4

Table 1 The normal computation results from three control points

	Img1	Img2	Img3	Img4
Computed normal	[0.078 –0.823 –0.562]	[0.034 –0.634 –0.773]	[−0.700 –0.134 –0.701]	[0.029 –0.935 –0.354]
Ground truth normal	[0.076 –0.825 –0.560]	[0.030 –0.631 –0.776]	[−0.697 –0.133 –0.705]	[0.027 –0.934 –0.356]
Angle error (in ^{0})	0.20	0.33	0.26	0.20

In the experiment, a different camera, Nikon Cool-Pix 4100, was used to take the images. The images have the resolution of 1024×768 (in pixel) (see some of the images in Fig. 21). The three control points (non co-linear) were chosen randomly from grid points on a chessboard pattern, as shown in Fig. 19, so that the support plane coincide with the pattern plane. The chessboard pattern was used for two purposes: 1) for accurate camera calibration with Zhang's approach [28]. The calibrated focal length is 4175.5 pixels; 2) to obtain the ground truth data for the plane normal and point-to-camera distances for the three control points. In the experiments, the ground truth data were obtained from the calibrated camera exterior parameters, such as the rotation matrix and the translation vector.

The computed plane normal from the four images are presented in Table 1 above. As can be observed from Table 1, the second and the third rows are for the results of the computed and ground truth plane normal. The angle between them is defined as computation errors, which are presented in the last row (unit in degree). It can be observed that the maximum and the average angles are 0.33 and 0.25°, respectively, which demonstrate that the computation model is accurate.

The distances of the control points were computed afterwards. Table 2 presents the computed distances from four images, where $\|\widetilde{X}_i\|$ and $\|X_i\|$ (i=1,2,3) are for the computed and ground truth distances, respectively. The difference $\|\widetilde{X}_i - X_i\|$ defines the distance computation error. From Table 2, the maximum and average distance computation errors are 0.8 cm and 0.4 cm, respectively, within 1 m to 2.2 m distances, which demonstrates the BPP algorithm is accurate.

The multiple solutions to P3P were also illustrated with the proposed BPP algorithm. The multiple solutions to P3P indicate multiple support planes. Since P3P gives two solutions most of the time, we demonstrate a typical two-solution P3P problem in the test. The classification of two solutions for P3P is referred to [11,25]. As can be observed in Fig. 22, there are two peak points in the likelihood map computed from the original image on the left. The likelihood map is highlighted to show the positions of the 30 plane normal with the highest likelihoods by

Table 2 The computed distances between the control points and the camera (unit in cm)

	$\|\widetilde{X}_1\|$	$\|X_1\|$	$\|\widetilde{X}_1 - X_1\|$	$\|\widetilde{X}_2\|$	$\|X_2\|$	$\|\widetilde{X}_2 - X_2\|$	$\|\widetilde{X}_3\|$	$\|X_3\|$	$\|\widetilde{X}_3 - X_3\|$
Img1	164.3	163.5	0.7	160.6	159.8	0.8	176.6	175.8	0.8
Img2	109.3	109.0	0.3	120.4	120.2	0.2	114.8	114.5	0.3
Img3	114.0	114.6	0.5	113.1	113.6	0.6	126.1	126.5	0.5
Img4	199.8	199.9	0.1	204.7	204.6	0.1	216.9	216.9	0.0

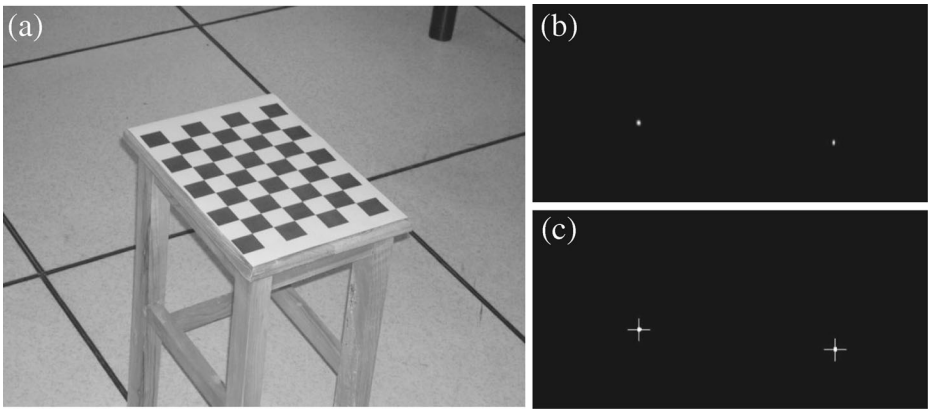

Fig. 22 Illustration of the two solutions to a practical P3P problem: **a**) *Left*: original image; **b**) *Top right*: likelihood map with two dominant local maximums; **c**) Bottom right: positions of the 30 plane normal with the highest likelihoods and the two computed normal (marked with '+')

thresholding. We can observe that there are two dominant areas segmented from the map. For each area, we computed the local maximum likelihood to derive two plane normal, which are $[0.0699–0.8883\ 0.4540]^T$ and $[-0.0917\ 0.8019\ 0.5904]^T$ (see the two positions marked with '+' in Fig. 22). Hence, the proposed BPP algorithm can not only solve a P3P problem but also classify the solutions.

5 Conclusions and recommendations

Planar scenes are commonly found in daily life and provide rich information for visual localization. In this paper, we proposed the "Perspective-Plane" problem, a similar but more general localization problem, compared to the well-known PnP and PnL problem. We addressed the problem within the Bayesian framework and proposed the "Bayesian Perspective-Plane (BPP)" algorithm. The core conception is the computation of the plane normal with a Maximum Likelihood Searching Model (MLS-M) by using generalized planar and out-of-plane constraints. The likelihood for each constraint is modeled with a normalized Gaussian function. The 2D and 1D searching modes were proposed to find the maximum likelihood and compute the plane normal. The proposed BPP algorithm has been tested with both simulation and real image data, which show that it is generalized to utilize different types of constraints for accurate object localization. Extensions of the proposed BPP algorithm for camera calibration were illustrated in the experiment with good results reported. Moreover, the BPP algorithm was successfully applied to solve the well-known Perspective-Three-Point (P3P) problem and classify the solutions. The results demonstrate that the proposed BPP algorithm is practical and flexible with good potentials for visual localization.

Future researches based on current work are recommended as follow: 1) incorporate the inequality constraint into the model, such as the first angle is larger than the second one; 2) improve the likelihood model to deal with different scales, units, etc.; 3) fast implementation of the maximum likelihood searching model (MLS-M), especially for high partition resolution of the Gaussian hemisphere. Some possible approaches can be coarse-to-fine searching strategy,

non-uniform distribution of plane normal on Gaussian sphere; 4) extend the proposed models and methods to more general scenes besides planar scenes, such as curve surface, sphere, etc.

Acknowledgments The work presented in this paper was sponsored by National Natural Science Foundation of China (NSFC) (No. 51208168), Tianjin Natural Science Foundation (No. 13JCYBJC37700), the Youth Top-Notch Talent Plan of Hebei Province, China, the Fundamental Research Funds for the Central Universities (WUT: 2014-IV-068), and the Grant-in-Aid for Scientific Research Program (No. 10049) from the Japan Society for the Promotion of Science (JSPS).

Appendixes

A. Uniformly sampling on the circle in 3D space

We can uniform sample on the circle in Eq. (19) by first sampling on a standard circle, and then mapping the sampled points via a rotation transform. This is implemented by following the three steps:

Step 1 Uniformly sample the standard circle, which has the following equation:

$$\begin{cases} X^2 + Y^2 = 1 \\ Z = 0 \end{cases} \tag{29}$$

This can be done by uniformly sampling within the angle space, and a sampled point is represented by $[\cos(\theta_i) \quad \sin(\theta_i) \quad 0]^T$, with $\theta_i \in [0\ 2\pi)$.

Step 2 Compute the rotation transform. The two circles defined by Eq. (19) and (29) are mapped via a rotation matrix, which satisfies

$$R[0 \quad 0 \quad 1]^T = N \tag{30}$$

The rotation matrix is not uniquely determined, because a circle is invariant to the rotation around the normal. We practically can use two arbitrary orthogonal vectors that satisfy Eq. (3), as the first and second column vectors.

Step 3 Transform the sampled points by rotation. Each sampled normal on the standard circle is transformed with the rotation matrix

$$R\begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \\ 0 \end{bmatrix} = \begin{bmatrix} r_1\cos(\theta_i) + r_4\sin(\theta_i) \\ r_2\cos(\theta_i) + r_5\sin(\theta_i) \\ r_3\cos(\theta_i) + r_6\sin(\theta_i) \end{bmatrix} \tag{31}$$

Hence, a uniform sampling on the circle from Eq. (19) is accomplished. We can prove that the angles between two neighboring sampled points before and after transformation are identical, because

$$\left(R\begin{bmatrix} \cos(\theta_{\mathrm{i}}) \\ \sin(\theta_{\mathrm{i}}) \\ 0 \end{bmatrix}\right)^T \left(R\begin{bmatrix} \cos(\theta_{\mathrm{i\text{-}1}}) \\ \sin(\theta_{\mathrm{i\text{-}1}}) \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(\theta_{\mathrm{i}}) \\ \sin(\theta_{\mathrm{i}}) \\ 0 \end{bmatrix}^T \begin{bmatrix} \cos(\theta_{\mathrm{i\text{-}1}}) \\ \sin(\theta_{\mathrm{i\text{-}1}}) \\ 0 \end{bmatrix} \tag{32}$$

B. Define the searching space for focal length

It is difficult to define the searching space of the focal length directly, since images are

in different resolutions in practice. Instead, we define the searching space from the camera viewing angles. There are three view angles defined for a camera: horizontal, vertical, and diagonal. In this paper, we use the horizontal one. It is defined as

$$\theta_h = 2\tan^{-1}\left(\frac{h}{2f}\right) \tag{33}$$

where h is the horizontal resolution of the image, and $\boldsymbol{f}$ is the camera focal length. As a result, we can estimate the focal length from the view angle and the image resolution:

$$f = \frac{h}{2\tan\left(\frac{\theta_h}{2}\right)} \tag{34}$$

Usually, we have some prior knowledge of the common used lens. We can thus define the searching space of the focal length from the range of the view angle and the image resolution.

C. Lemma: Determining the support plane is the necessary and sufficient condition to solve the P3P problem

1. Proof: 1) Necessary condition

   The distances of the three control points to the camera are known from a solved P3P. Hence, we can determine the 3D coordinates of each control points by the following equations

$$\begin{cases} X_i = \lambda K^{-1} x_i \\ \|X_i\| = d_i \end{cases} \tag{35}$$

   With the recovered control points, the support plane is uniquely determined then.

1. 2) Sufficient condition

   The plane normal and distance are known from a determined support plane. As a result, we can use Eq. (4) to compute the 3D positions of each control points. The distances of the three points are computed readily from the 3D coordinates. Hence, the P3P problem is solved.

References

1. Adan A, Martin A, Valero E, Merchan P (2009) Landmark real-time recognition and positioning for pedestrian navigation. CIARP, Guadalajara, Mexico
2. Cham T, Arridhana C et al (2010) Estimating camera pose from a single urban ground-view omni-directional image and a 2D building outline map. CVPR, SF, CA
3. Criminisi A, Reid I, Zisserman A (2000) Single view metrology. International Journal of Computer Vision 40(2):123–148

4. Desouza GN, Kak AC (2002) Vision for mobile robot navigation: a survey. IEEE Trans on Pattern Analysis and Machine Intelligence 24(2):237–267
5. Durrant-Whyte H, Bailey T (2006) Simultaneous localization and mapping (SLAM): part I the essential algorithms. IEEE Robotics and Automation Magazine 13(2):99–110
6. Guan P, Weiss A, Balan A, Black M (2009) Estimating human shape and pose from a single image, international conference on computer vision. Kyoto, Japan
7. Guo F, Chellappa R (2010) Video metrology using a single camera. IEEE Trans on Pattern Analysis and Machine Intelligence 32(7):1329–1335
8. Hartley R, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
9. Horn BKP (1987) Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am A 4(4): 629–642
10. Hu Z, Matsuyama T (2012) Bayesian perspective-plane (BPP) for localization, international conference on computer vision theory and applications (VISAPP). Rome, Italy, pp 241–246
11. L. Kneip, D. Scaramuzza, R. Siegwart,, (2011) A novel parameterization of the perspective-three-point problem for a direct computation of absolute camera position and orientation, Proc. of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
12. H. Lategahn, C. Stiller, Vision-Only Localization, IEEE Intelligent (2014) Transportation Systems Magazine, DOI: 10.1109/TITS.2014.2298492
13. Lee DC, Hebert M, Kanade T (2009) Geometric reasoning for single image structure recovery. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)(June)
14. Leopardi P (2006) A partition of the unit sphere into regions of equal area and small diameter. Electronic Transactions on Numerical Analysis 25(12):309–327
15. Li S, Xu C (2011) A stable direct solution of perspective-three-point problem. International Journal of Pattern Recognition and Artificial Intelligence 25(11):627–642
16. Liebowitz D, Zisserman A (1998) Metric rectification for perspective images of planes. CVPR, Santa Barbara, CA
17. A. Pretto, S. Tonello, E. Menegatti, (2013) Flexible 3D localization of planar objects for industrial bin-picking with monocamera vision system, IEEE International Conference on Automation Science and Engineering, pp. 168–175
18. S.Q., Li, C. Xu, and M. Xie, A Robust (2012) O (n) Solution to the perspective-three-point problem, IEEE Transaction on Pattern Analysis and Machine Intelligence, 34 (7): 1444–1450
19. Schneider D, Fu X, Wong K (2010) Reconstruction of display and eyes from a single image. CVPR, SF, CA
20. Shi F, Zhang X, Liu Y (2004) A new method of camera pose estimation using 2D-3D corner correspondence. Pattern Recognition Letters 25(10):805–809
21. Sun Y, Yin L (2008) Automatic pose estimation of 3D facial models. ICPR, FL, US
22. Wang G, Hu Z, Wu F, Tsui H (2005) Single view metrology from scene constraints. Image and Vision Computing 23(9):831–840
23. Wang R, Jiang G, Quan L, Wu C (2012) Camera calibration using identical objects. Machine Vision and Applications 23(3):579–587
24. Witkin AP (1981) Recovering surface shape and orientation from texture. Artificial Intelligence 17(1–3):17–45
25. Wolfe W, Mathis D, Sklair C, Magee M (1991) The perspective view of 3 points. IEEE Transaction on Pattern Analysis and Machine Intelligence 13(1):66–73
26. Wu Y, Li X, Wu F, Hu Z (2006) Coplanar circles, quasi-affine invariance and calibration. Image and Vision Computing 24(4):319–326
27. A Zakhor, A Hallquist, Single view pose estimation of mobile devices in urban environments, Proceedings of the 2013 I.E. Workshop on Applications of Computer Vision (WACV), 2013, pp. 347–354
28. Zhang Z (2000) A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11):1330–1334
29. Zhang BW, Li Y (2008) Dynamic calibration of the relative pose and error analysis in a structured light system. J Opt Soc Am A 25(3):612–622
30. Zhu X, Ramanan D (2012) Face detection, pose estimation, and landmark localization in the wild. Computer Vision and Pattern Recognition (CVPR), Providence, RI

Zhaozheng Hu received the Bachelor and PhD degrees from Xi'an Jiaotong University, China, in 2002 and 2007, respectively, both in information and communication engineering. During his PhD period, he visited the computer vision lab at the Chinese University of Hong Kong from Nov., 2004 to Jun., 2005. From Jul., 2007 to Aug., 2009, he worked as a Post-doc Fellow in Georgia Institute of Technology, U.S.A. From Jul., 2010 to Jul., 2012, he was with the visual information processing lab in Kyoto University, Japan, under a JSPS Fellowship program. He is currently a professor in Wuhan University of Technology, Wuhan, China. His research topics mainly focus on visual geometry, stereo vision, intelligent transportation system, intelligent surveillance system, etc.

Takashi Matsuyama received the BEng, MEng, and DEng degrees in electrical engineering from Kyoto University, Japan, in 1974, 1976, and 1980, respectively. Currently, he is working as a professor in the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. His research interests include knowledge-based image understanding, computer vision, 3D video, human-computer interaction, and smart energy management. He has written more than 100 papers and books, including two research monographs: A Structural Analysis of Complex Aerial Photographs (Plenum, 1980) and SIGMA: A Knowledge-Based Aerial Image Understanding System (Plenum, 1990). He won 10 best paper awards from Japanese and international academic societies including the Marr Prize at ICCV'95. He is on the editorial board of the Pattern Recognition Journal. He was awarded fellowships from the International Association for Pattern Recognition, the Information Processing Society of Japan, and the Institute of Electronics, Information, and Communication Engineers Japan. He is a member of the IEEE.