Automatic Recognition of Spurious Surface in Building Exterior Survey

Yan Lu, Dezhen Song, Haifeng Li, and Jingtai Liu

Abstract— **Buildings consume around 40% of overall energy in the world. Planar mirror detection problem (PMDP) arises when surveying reflective building surface for building energy retrofit. PMDP is also important for collision avoidance when robots navigate close to highly reflective glassy walls. Our approach uses two views from an on-board camera. First, we derive geometric constraints for corresponding real-virtual features across two views. The constraints include 1) the mirror normal as a function of vanishing points of lines connecting the real-virtual feature point pairs and 2) the mirror depth in a closed form format derived from a mirror plane induced homography. Based on the geometric constraints, we employ a random sample consensus framework and an affine scale-invariant feature transform to develop a robust mirror detection algorithm. We have implemented the algorithm and tested it under both in-lab and field settings. The algorithm has achieved an overall detection accuracy rate of 91.0%.**

I. Introduction

The fast development of service robots has advanced robot work space from factory floors to our daily life. One important new task is to employ robots to perform building survey to assist building energy retrofit because buildings account for around 40% energy usage [1]. In the survey, robots need to recognize reflective surface to provide a proper estimation of building thermal load. Unfortunately, highly reflective surfaces, such as glassy building exterior and mirrored walls, challenge almost every type of sensors. Laser range finders, sonar arrays, and cameras are no exceptions because light and sound signals simply bounce off the surfaces, which become invisible to the sensors. Detecting these surfaces is also necessary to avoid collisions in robot navigation.

We report a method for this new planar mirror detection problem (PMDP) using two views from an on-board camera. First, we derive geometric constraints for corresponding real-virtual features across two views where virtual features refer to the mirror reflection of real features. The constraints include 1) the mirror normal as a function of vanishing points of lines connecting the real-virtual feature point pairs and 2) the mirror depth in a closed form format derived from a mirror plane-induced homography. We also address the issue that popular feature detectors, such as scale-invariant feature transform (SIFT), are not reflection invariant by combining a secondary reflection with an affine scale-invariant feature

transform (ASIFT). Based on the results, we employ a random sample consensus (RANSAC) framework to develop a robust mirror detection algorithm. We have implemented the algorithm and tested it under both in-lab and field settings. The algorithm has achieved an overall accuracy rate of 91.0%.

II. Related Work

PMDP is not a simple plane reconstruction problem using 3D vision. It relates to many areas including intelligence level tests in artificial intelligence (AI) community, planar catadioptric stereo (PCS) systems, construction of specular surfaces, and reflection invariant feature extractions.

In AI and animal behavior communities, researchers often assess intelligence levels based on the subject's ability of detecting a mirror or its own reflection [2], [3]. In the well known mirror and mark test, a subject has a mark that cannot be directly seen but is visible in the mirror. If the subject increases the exploration and self-direction actions towards the mark, it means that the subject recognizes the mirror image as self. Existing results show that chimpanzees [3], gorillas [4], dolphins [5], and magpies [6] have evident self-recognition in front of mirrors except monkeys [7]. We do not have mirror and mark tests for robots yet. It is clearly not a trivial problem. Initial related results focus on robot self recognition [8], [9] using motion and appearance, which is not as difficult as recognizing a mirror when a robot cannot see its own reflection. Such cases are not unusual because the robot cannot see itself when approaching a mirror from side. Our approach addresses this problem by exploring symmetricity in the scene.

Mirror detection is also related to PCS systems in computer vision. A PCS system usually consists of a static camera and one or more planar mirrors with the aim of achieving stereo or structure from motion (SFM) [10], [11]. Since detecting mirror pose is just a calibration problem in PCS systems, in-lab settings and calibration patterns (e.g. checkerboard) can be used here. However, this is not a viable approach when robots need to detect mirror surfaces *in situ*.

In a way, planar mirror detection can be viewed as a special case of specular surface construction. Existing approaches rely on active sensing by changing lighting [12], [13] and polarity [14], [15], or assuming curvature of the mirror [16]. These approaches have difficulty to be adapted for robots because natural lighting can easily overwhelm the setup. To avoid the issue, we use features from images.

SIFT [17] is well known for its invariance to image scaling and translation, and partial invariance to affine distortion. However, it is not reflection invariant and thus cannot be

Y. Lu and D. Song are with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843, United States. Emails: {*ylu, dzsong*}@*cse.tamu.edu*.

H. Li is with the College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, P. R. China. Email: *li-haifeng666@gmail.com*.

J. Liu is with the Institute of Robotics and Automatic Information System, Nankai University, Tianjin 300071, P. R. China. Email: *li-ujt@robot.nankai.edu.cn*.

applied to our problem. As extensions of SIFT, descriptors invariant to mirror reflection have recently been designed by modifying the SIFT descriptor structure at the expense of distinctiveness, such as MI-SIFT [18] and FIND [19]. They still cannot fit our need because our feature correspondence involves not only a reflection difference but also a significant projective distortion induced by perspective changes. On the other hand, descriptors invariant to affine transforms can handle large perspective changes (e.g., [20], [21]). Among these affine invariant descriptors, ASIFT [22] shows promising performance and becomes our choice of feature transformation. Later we will show how to make ASIFT reflection-invariant.

In a previous work [23], our group has investigated the problem of estimating the orientation of a mirror plane using a single view. However, the depth information cannot be extracted from a single view and it limits the detection capability. We also work on survey of building exterior surfaces [24] where we employ a multi-layer feature graph to describe build exterior.

III. PROBLEM DEFINITION

To define our problem and focus on the most relevant issues, we have the following assumptions.

a.1 Each view captures a real scene and its mirror reflection, and the scene is feature-rich.

a.2 The camera calibration matrix is known to be K.

a.3 The baseline distance $|\mathbf{t}|$ between two views is known. The distance is usually short and can be measured by on-board sensors like inertial measurement unit (IMU). If $|\mathbf{t}|$ is unknown, our method still applies but the depth result is measured in ratio instead of absolute value.

We also have the following conventions in notation. Let I and $\{I\}$ be the image and the image coordinate system (ICS) for the first view, respectively. I' and $\{I'\}$ are defined similarly for the second view. The camera coordinate system (CCS) is right-handed, with the origin $\mathbf{C}$ at the camera center, and Z-axis along the principal axis. With respect to the CCS of the first view, we define,

- $\pi_m = (\mathbf{n}_m^\mathsf{T}, d_m)^\mathsf{T}$ as the mirror plane where $\mathbf{n}_m$ is a 3×1 unit vector indicating the normal of π_m, and d_m is the plane depth (i.e., the distance from $\mathbf{C}$ to π_m),
- $\mathbf{X}_{\mathbf{r}i}$ as the i-th real 3D point and $\mathbf{X}_{\mathbf{v}i}$ as its mirror reflection (a virtual point),
- $\mathbf{x}_{\mathbf{r}i}$ and $\mathbf{x}_{\mathbf{v}i}$ as the projections of $\mathbf{X}_{\mathbf{r}i}$ and $\mathbf{X}_{\mathbf{v}i}$ in $\{I\}$, respectively, and
- $\mathbf{X}_{\mathbf{r}i} \leftrightarrow \mathbf{X}_{\mathbf{v}i}$ as a 3D real-virtual (R-V) pair and $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}$ as a 2D R-V pair.

In the CCS of the second view, notations differ from their counterparts in the CCS of the first view by adding a superscript $'$, e.g. $\mathbf{n}'_m$, $\mathbf{x}'_{\mathbf{r}i}$ and $\mathbf{x}'_{\mathbf{v}i}$. It is worth noting that there is a new type of correspondence between the 2D R-V pairs in both views, which is denoted in a quadruple format: $\mathbf{Q}_i = \{\mathbf{x}_{\mathbf{r}i}, \mathbf{x}_{\mathbf{v}i}, \mathbf{x}'_{\mathbf{r}i}, \mathbf{x}'_{\mathbf{v}i}\}$.

Also, all above notations about points are represented in homogeneous coordinates while their inhomogeneous counterparts are denoted by adding a tilde on their top, e.g., $\widetilde{\mathbf{x}}_{\mathbf{r}i}$.

With assumptions and notations defined, our PMDP is,

Definition 1: Given two views I and I', the camera calibration matrix K and the camera translation distance $|\mathbf{t}|$, determine if there is a mirror. If so, estimate π_m.

IV. MODELING

We begin with analyzing the geometric relationship between noise-free feature points. The geometric relationship will be used in a RANSAC framework to filter noisy inputs later. The noise-free feature inputs here are a set of quadruples $\{\mathbf{Q}_i\}$. The geometric relationship is constraints on quadruples induced by 3D reflection and the imaging process. As the result, π_m will be derived as a function of quadruples in two stages: orientation and depth. First, we solve the mirror orientation using quadruples.

Lemma 1: Given two quadruples $\mathbf{Q}_i$ and $\mathbf{Q}_j$, the mirror normal with respect to both CCSs can be obtained as follows,

$$\mathbf{n}_m = \mathrm{K}^{-1}(\mathbf{x}_{\mathbf{r}i} \times \mathbf{x}_{\mathbf{v}i}) \times (\mathbf{x}_{\mathbf{r}j} \times \mathbf{x}_{\mathbf{v}j}),$$
$$\mathbf{n}'_m = \mathrm{K}^{-1}(\mathbf{x}'_{\mathbf{r}i} \times \mathbf{x}'_{\mathbf{v}i}) \times (\mathbf{x}'_{\mathbf{r}j} \times \mathbf{x}'_{\mathbf{v}j}), \quad (1)$$

where symbol '$\times$' represents the cross product.

Proof: Consider the geometry relationship in Fig. 1. As a convention, we define $\overleftrightarrow{AB}$ as the line passing through
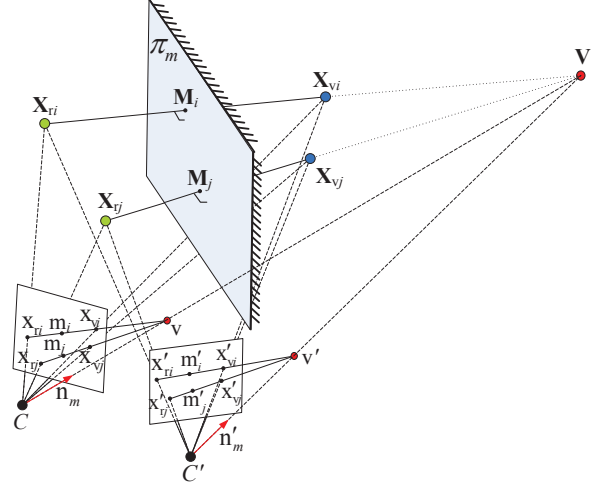

Fig. 1. A perspective illustration of the geometry relationship between real-virtual pairs across two views. Note that the parallel lines are not parallel in the illustrate because we want to bring vanishing point from infinity to the figure.

points A and B. From the property of planar mirror reflection, we have $\overleftrightarrow{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}} \perp \pi_m$, $\overleftrightarrow{\mathbf{X}_{\mathbf{r}j}\mathbf{X}_{\mathbf{v}j}} \perp \pi_m$, and thus $\overleftrightarrow{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}} // \overleftrightarrow{\mathbf{X}_{\mathbf{r}j}\mathbf{X}_{\mathbf{v}j}}$. After a projective transformation, the projections of $\overleftrightarrow{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}}$ and $\overleftrightarrow{\mathbf{X}_{\mathbf{r}j}\mathbf{X}_{\mathbf{v}j}}$ in $\{I\}$ (or $\{I'\}$) would intersect at a vanishing point $\mathbf{v}$ (or $\mathbf{v}'$) in the corresponding ICS,

$$(\mathbf{x}_{\mathbf{r}i} \times \mathbf{x}_{\mathbf{v}i}) \times (\mathbf{x}_{\mathbf{r}j} \times \mathbf{x}_{\mathbf{v}j}) = \mathbf{v}$$
$$(\mathbf{x}'_{\mathbf{r}i} \times \mathbf{x}'_{\mathbf{v}i}) \times (\mathbf{x}'_{\mathbf{r}j} \times \mathbf{x}'_{\mathbf{v}j}) = \mathbf{v}'. \quad (2)$$

On the other hand, $\mathbf{v}$ can be viewed as the projection of $\mathbf{n}_m$ in $\{I\}$

$$\mathbf{v} = \mathrm{K}\mathbf{n}_m, \text{ and similarly, } \mathbf{v}' = \mathrm{K}\mathbf{n}'_m. \quad (3)$$

Combining (2) and (3), we obtain (1). ∎

The second step is to derive mirror depth d_m. From epipolar geometry, we can obtain the camera rotation matrix R and translation vector $\mathbf{t}$ by decomposing the essential matrix [25]. A straightforward way of computing the equation of π_m is by reconstructing 3D points via triangulation. However, we will show a homography based method which avoids the triangulation process.

Our method involves the homography between the corresponding middle points of R-V pairs in two views. Let $\overline{AB}$ denote the line segment defined by points A and B in the rest of the paper. Denote the midpoint of $\overline{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}}$ by $\mathbf{M}_i$, and its projection in $\{I\}$ by $\mathbf{m}_i$ (see Fig. 1 for examples). $\mathbf{m}_i$ can be obtained using a cross ratio which is detailed in the following lemma:

Lemma 2: Given quadruple $\mathbf{Q}_i$, the projection $\mathbf{m}_i$ of the midpoint $\mathbf{M}_i$ of $\overline{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}}$ is determined as follows,

$$\widetilde{\mathbf{m}}_i = (1-a)\widetilde{\mathbf{x}}_{\mathbf{r}i}+a\widetilde{\mathbf{x}}_{\mathbf{v}i}, \quad \text{and} \quad a = \frac{|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{v}}|}{2|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{v}}| - |\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{x}_{\mathbf{v}i}}|}, \quad (4)$$

where $|\cdot|$ denotes the length of the line segment.

Proof: Consider the projection from $\overrightarrow{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}}$ to $\overleftrightarrow{\mathbf{x}_{\mathbf{r}i}\mathbf{x}_{\mathbf{v}i}}$. A basic invariant in this projection is the cross ratio of the four collinear points $\mathbf{X}_{\mathbf{r}i}$, $\mathbf{M}_i$, $\mathbf{X}_{\mathbf{v}i}$, and $\mathbf{V}$,

$$\frac{|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{m}_i}||\overline{\mathbf{x}_{\mathbf{v}i}\mathbf{v}}|}{|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{x}_{\mathbf{v}i}}||\overline{\mathbf{m}_i\mathbf{v}}|} = \frac{|\overline{\mathbf{X}_{\mathbf{r}i}\mathbf{M}_i}||\overline{\mathbf{X}_{\mathbf{v}i}\mathbf{V}}|}{|\overline{\mathbf{X}_{\mathbf{r}i}\mathbf{X}_{\mathbf{v}i}}||\overline{\mathbf{M}_i\mathbf{V}}|} = \frac{1}{2}. \quad (5)$$

Representing $\mathbf{m}_i$ as $\widetilde{\mathbf{m}}_i = (1 - a)\widetilde{\mathbf{x}}_{\mathbf{r}i} + a\widetilde{\mathbf{x}}_{\mathbf{v}i}$, $0 \le a \le 1$, in the inhomogeneous coordinate, we have

$$|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{m}_i}| = a|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{x}_{\mathbf{v}i}}|,$$
$$|\overline{\mathbf{m}_i\mathbf{v}}| = |\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{v}}| - a|\overline{\mathbf{x}_{\mathbf{r}i}\mathbf{x}_{\mathbf{v}i}}|. \quad (6)$$

Substituting (6) into (5) gives the final result in (4). ∎

We now can derive the mirror depth with $\mathbf{m}_i$.

Lemma 3: Given quadruple $\mathbf{Q}_i$ and mirror normal $\mathbf{n}_m$, the mirror depth is

$$d_m = ([\mathbf{m}_i']_\times \mathrm{K R K}^{-1}\mathbf{m}_i)^\dagger[\mathbf{m}_i']_\times \mathrm{K}\mathbf{t}\mathbf{n}_m^\mathsf{T}\mathrm{K}^{-1}\mathbf{m}_i \quad (7)$$

where $(\cdot)^\dagger$ denotes the pseudoinverse operation, and $[\mathbf{m}_i']_\times$ is a skew-symmetric matrix,

$$\begin{bmatrix} 0 & -m_{i3}' & m_{i2}' \\ m_{i3}' & 0 & -m_{i1}' \\ -m_{i2}' & m_{i1}' & 0 \end{bmatrix}. \quad (8)$$

Proof: Observe that $\mathbf{M}_i$ lies on the plane π_m. Then $\mathbf{m}_i$ and $\mathbf{m}_i'$ must obey a homography $\mathbf{m}_i' = \mathrm{H}\mathbf{m}_i$ induced by π_m, where H can be expressed as [25]

$$\mathrm{H} = \mathrm{K}(\mathrm{R} - \frac{1}{d_m}\mathbf{t}\mathbf{n}_m^\mathsf{T})\mathrm{K}^{-1} \quad (9)$$

H has 1 degree of freedom (DOF) since only d_m is unknown.

$\mathbf{m}_i$ and $\mathbf{m}_i'$ can be computed from $\mathbf{Q}_i$ using (4). Since $\mathbf{m}_i' = \mathrm{H}\mathbf{m}_i = \mathrm{K}(\mathrm{R} - \frac{1}{d_m}\mathbf{t}\mathbf{n}_m^\mathsf{T})\mathrm{K}^{-1}\mathbf{m}_i$, we have

$$\mathbf{m}_i' \times \mathrm{K}(\mathrm{R} - \frac{1}{d_m}\mathbf{t}\mathbf{n}_m^\mathsf{T})\mathrm{K}^{-1}\mathbf{m}_i$$
$$= [\mathbf{m}_i']_\times \mathrm{K R K}^{-1}\mathbf{m}_i - [\mathbf{m}_i']_\times \mathrm{K}\frac{1}{d_m}\mathbf{t}\mathbf{n}_m^\mathsf{T}\mathrm{K}^{-1}\mathbf{m}_i = \mathbf{0}$$

Then we have

$$[\mathbf{m}_i']_\times \mathrm{K R K}^{-1}\mathbf{m}_i d_m = [\mathbf{m}_i']_\times \mathrm{K}\mathbf{t}\mathbf{n}_m^\mathsf{T}\mathrm{K}^{-1}\mathbf{m}_i$$

The above system of equations is over-determined since the rank of $[\mathbf{m}_i']_\times$ is 2. Thus, the least-square solution of d_m is given by (7), which is also an exact solution when the system is noise-free. ∎

V. ALGORITHM

Section IV provides geometric relationship for noise-free quadruples. To complete the algorithm, we need to select correct feature transformation and verify the geometric relationship with respect to noisy features using the well-accepted RANSAC framework. First, let us detail the feature selection in quadruple extraction.

A. Quadruple Extraction

To form a quadruple, we need two kinds of point correspondences: cross-view correspondence, e.g. $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{r}i}'$, and R-V pair correspondence, e.g. $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}$. The former can be handled by standard feature extraction methods such as SIFT, since the change of camera viewpoint is usually not too large. However, the latter is nontrivial because $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}$ involves an improper transformation in 3D (between $\mathbf{X}_{\mathbf{r}i}$ and $\mathbf{X}_{\mathbf{v}i}$).

Therefore, the key to this problem is to find features and their correspondence under the improper transformation. We want to convert the 3D reflection to a rigid body transformation such that existing feature extraction and matching algorithms can be applied. Our solution is to introduce a second mirror π_s hypothetically, inspired by the fact that two consecutive reflections lead to a rigid body transformation regardless of the mirror configuration [10]. The simplest scenario happens when two mirrors are placed in the same position but facing oppositional directions. Denote the reflection of $\mathbf{X}_{\mathbf{v}i}$ about the second mirror by $\mathbf{X}_{\mathbf{s}i}$. In this special case, $\mathbf{X}_{\mathbf{s}i}$ would be the same as $\mathbf{X}_{\mathbf{r}i}$.

However, the projection $\mathbf{x}_{\mathbf{s}i}$ of $\mathbf{X}_{\mathbf{s}i}$ in the image is hard to predict if the position of π_s is arbitrarily chosen. In a previous work [23], we have found that if π_s passes the camera principal axis, then for any $\mathbf{X}_{\mathbf{s}i}$, its projection $\mathbf{x}_{\mathbf{s}i}$ can be obtained by flipping $\mathbf{x}_{\mathbf{v}i}$ about an axis in the image plane. In another words, the 3D reflection by π_s effectively causes a 2D flipping for every point in the image plane. Moreover, the 2D flipping axis is even unimportant for the point matching result as long as a 2D-rotation-invariant descriptor (e.g. SIFT descriptor) is used. As a result, we can flip the image around an arbitrary axis in practice.

Although image flipping allows us to find correspondences between $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{s}i}$ instead of dealing with $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}$ directly, it comes at a price. A new issue is that the rotation angle θ of the rigid body motion (between $\mathbf{X}_{\mathbf{r}i}$ and $\mathbf{X}_{\mathbf{s}i}$) is twice as large as the angle α between the principal axis and π_m [10] (see Fig. 2). As the value of α varies in different situations, it can easily lead to a significant perspective change that fails the standard SIFT algorithm.
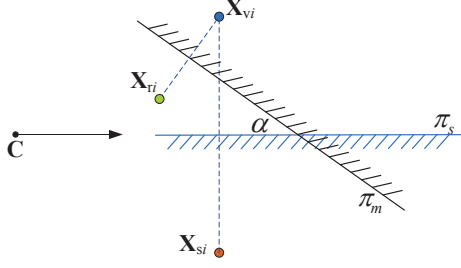
Fig. 2. A top-down view of the configuration of π_s and π_m. π_s is placed to pass the camera principal axis.

To handle this problem, we employ an affine invariant feature extraction algorithm ASIFT which has advantages over SIFT when dealing with large perspective changes [22]. Once a correspondence $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{s}i}$ is identified, the R-V pair $\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}$ is readily established based on the known mapping between $\mathbf{x}_{\mathbf{v}i}$ and $\mathbf{x}_{\mathbf{s}i}$. Algorithm 1 summarizes how quadruples are constructed.

Algorithm 1: ASIFT-based Quadruple Extraction

Input : Two images I and I'
Output: A set of quadruples $\{\mathbf{Q}_k\}$

1 flip I left-right (or up-down) to get I_f;
2 find point matches $\{\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{s}i}\}$ between I and I_f using ASIFT;
3 map $\mathbf{x}_{\mathbf{s}i}$ in I_f back to $\mathbf{x}_{\mathbf{v}i}$ in I to establish R-V correspondences $\{\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}\}$;
4 apply steps 1-3 to I' to obtain $\{\mathbf{x}'_{\mathbf{r}j} \leftrightarrow \mathbf{x}'_{\mathbf{v}j}\}$;
5 find cross-view matches $\{\mathbf{x}_{\mathbf{r}k} \leftrightarrow \mathbf{x}'_{\mathbf{r}k}\}$ from $\{\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}\}$ and $\{\mathbf{x}'_{\mathbf{r}j} \leftrightarrow \mathbf{x}'_{\mathbf{v}j}\}$ using ASIFT;
6 construct quadruples $\{\mathbf{Q}_k\}$ from $\{\mathbf{x}_{\mathbf{r}i} \leftrightarrow \mathbf{x}_{\mathbf{v}i}\}$ and $\{\mathbf{x}'_{\mathbf{r}j} \leftrightarrow \mathbf{x}'_{\mathbf{v}j}\}$ according to $\{\mathbf{x}_{\mathbf{r}k} \leftrightarrow \mathbf{x}'_{\mathbf{r}k}\}$;
7 **return** $\{\mathbf{Q}_k\}$;

B. Maximum Likelihood Estimation (MLE)

To apply RANSAC framework, we need to estimate $\mathbf{n}_m$, $\mathbf{n}'_m$ and d_m using the quadruples $\{\mathbf{Q}_i\}$ from the inlier set by minimizing a cost function. Assuming measurement errors are Gaussian, then the estimation is MLE if reprojection error is employed as the cost function. Let us derive this metric.

For $\mathbf{Q}_i$, let $\mathcal{X}_i = (\tilde{x}_{\mathbf{r}i}, \tilde{y}_{\mathbf{r}i}, \tilde{x}_{\mathbf{v}i}, \tilde{y}_{\mathbf{v}i}, \tilde{x}'_{\mathbf{r}i}, \tilde{y}'_{\mathbf{r}i}, \tilde{x}'_{\mathbf{v}i}, \tilde{y}'_{\mathbf{v}i})^\mathsf{T}$ be a 8-vector formed by concatenating the inhomogeneous coordinates of $\mathbf{x}_{\mathbf{r}i}, \mathbf{x}_{\mathbf{v}i}, \mathbf{x}'_{\mathbf{r}i}$ and $\mathbf{x}'_{\mathbf{v}i}$. Given points $\{\mathcal{X}_i\}$ in the measurement space $\mathbb{R}^8$, the task of estimating $\mathbf{n}_m$, $\mathbf{n}'_m$ and d_m becomes finding a variety that passes through the points $\{\mathcal{X}_i\}$ in $\mathbb{R}^8$. Because of noise, it is impossible to fit a variety exactly. In this case, let $\mathcal{V}$ be the variety corresponding to $\mathbf{n}_m$, $\mathbf{n}'_m$ and d_m, and let $\widehat{\mathcal{X}}_i$ be the closest point to $\mathcal{X}_i$ lying on $\mathcal{V}$.

Given $\mathbf{n}_m$, $\mathbf{n}'_m$ and d_m, define

$$\mathcal{C}_\mathcal{V}(\widehat{\mathcal{X}}_i) := \begin{bmatrix} (\widehat{\mathbf{x}}_{\mathbf{r}i} \times \widehat{\mathbf{x}}_{\mathbf{v}i})^\mathsf{T} \mathrm{K} \mathbf{n}_m \\ (\widehat{\mathbf{x}}'_{\mathbf{r}i} \times \widehat{\mathbf{x}}'_{\mathbf{v}i})^\mathsf{T} \mathrm{K} \mathbf{n}'_m \\ \widehat{\mathbf{m}}'_i \times \mathrm{H} \widehat{\mathbf{m}}_i \end{bmatrix},$$

where $\mathrm{H}, \widehat{\mathbf{m}}_i$ and $\widehat{\mathbf{m}}'_i$ are intermediate variables computed using (9) and (4), respectively, and $\widehat{\mathbf{x}}_{\mathbf{r}i} = (\widehat{x}_{\mathbf{r}i}, \widehat{y}_{\mathbf{r}i}, 1)^\mathsf{T}$, and similarly for $\widehat{\mathbf{x}}_{\mathbf{v}i}, \widehat{\mathbf{x}}'_{\mathbf{r}i}$ and $\widehat{\mathbf{x}}'_{\mathbf{v}i}$. Then the MLE method is to find $\mathbf{n}_m$, $\mathbf{n}'_m$, d_m and $\widehat{\mathcal{X}}_i$ that minimize the error function

$$\sum_i \|\mathcal{X}_i - \widehat{\mathcal{X}}_i\|^2_{\Sigma_i}, \tag{10}$$

subject to $\mathcal{C}_\mathcal{V}(\widehat{\mathcal{X}}_i) = \mathbf{0}, \forall i$, where Σ_i is the covariance of $\mathcal{X}_i$, and $\| \cdot \|_\Sigma$ represents the Mahalanobis distance.

Although minimizing the reprojection error is MLE, it involves solving a high-dimensional non-linear optimization problem, which is quite complex and time-consuming. To speed up the algorithm, we derive Sampson error approximation. Instead of finding the closest point $\widehat{\mathcal{X}}_i$ on the variety $\mathcal{V}$ to the measurement $\mathcal{X}_i$, the Sampson error function estimates a first-order approximation to $\widehat{\mathcal{X}}_i$. For given $\mathbf{n}_m$, $\mathbf{n}'_m$ and d_m, any point $\mathcal{X}_i$ lying on $\mathcal{V}$ will satisfy $\mathcal{C}_\mathcal{V}(\mathcal{X}_i) = \mathbf{0}$. Then the Sampson approximation to (10) is $\sum_i \epsilon_i^\mathsf{T} (\mathrm{J}_i \Sigma_i \mathrm{J}_i^\mathsf{T})^{-1} \epsilon_i$ where $\epsilon_i = \mathcal{C}_\mathcal{V}(\mathcal{X}_i)$ and $\mathrm{J}_i = \frac{\partial \mathcal{C}_\mathcal{V}}{\partial \mathcal{X}_i}$.

C. Applying RANSAC Framework

Algorithm 2: Robust Mirror Estimation using RANSAC

Input : Two images I and I'
Output: Mirror plane π_m or **no mirror**

1 obtain a set $\mathcal{S}$ of quadruples using Algorithm 1;
2 $N = \infty$;
3 **for** $k \leftarrow 1$ **to** N **do**
4   randomly sample 2 quadruples from S;
5   compute $\mathbf{n}_m^{(k)}$, $\mathbf{n}'^{(k)}_m$ and $d_m^{(k)}$ using (1) and (7);
6     $\mathscr{I}_k = \emptyset$;   // initialize inlier set
7   **for** $\mathbf{Q}_i \in S$ **do**
8         $D_i = \sqrt{\epsilon_i^\mathsf{T} (\mathrm{J}_i \Sigma_i \mathrm{J}_i^\mathsf{T})^{-1} \epsilon_i}$;
9   **if** $|D_i| < T_D$ **then**
10             $\mathscr{I}_k = \mathscr{I}_k \cup \mathbf{Q}_i$;
11   update N using (4.18) from [25] (Page 119);
12 $k^\star = \arg\max_k |\mathscr{I}_k|$;
13 $\mathscr{I}^\star = \mathscr{I}_{k^\star}$;
14 **if** $|\mathscr{I}^\star| < T_N$ **then**
15   **return** **no mirror**;
16 **else**
17   re-estimate $\mathbf{n}_m$, $\mathbf{n}'_m$ and d_m with $\mathscr{I}^\star$ by minimizing Sampson error using the Levenberg-Marquardt algorithm;
18   (guided matching): find correspondence inliers consistent with the optimal estimation;
19   **return** π_m;

We are now ready to apply the RANSAC to the set S of quadruples to estimate π_m. The whole algorithm is summarized in Algorithm 2. There are two thresholds used: inlier-outlier threshold T_D and mirror detection threshold T_N. Threshold T_D in step 9 is used to determine whether the quadruple belongs to the current inlier set. The threshold T_D is chosen based on 8 DOFs of the decision variables. With a preset probability threshold of 0.95, $T_D = \sqrt{15.51\sigma^2}$ according to [25] where σ is the standard deviation of the measurement error for feature points. The algorithm returns "no mirror" when the maximum inlier set is smaller than T_N (Step 15). T_N will be determined experimentally through in-lab tests in Section VI-A. In step 11, the maximum sample iteration N is chosen adaptively (Page 119 of [25]). Steps 17 and 18 of Algorithm 2 can be iterated until the number of correspondence inliers is stable.

VI. EXPERIMENTS

We have implemented the proposed algorithm using Mat-lab under a Windows 7 operating system. For the ASIFT algorithm, we use the open source implementation in [26]. Images are taken by a pre-calibrated Vivicam 7020 camera with a resolution of 640×480 pixels. We first test the algorithms in our lab to determine the algorithm accuracy under the controlled settings and to determine the proper threshold before extensive field tests.
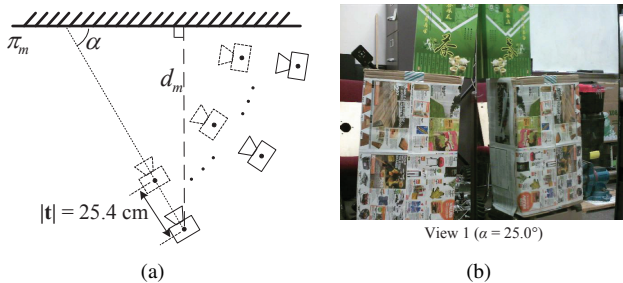
A. In-lab Tests

Fig. 3. In-lab experiment setup: (a) Experiment configurations. (b) A sample view.

Fig. 3(a) illustrates the setup of in-lab tests. Define α as the angle between the camera optical axis and the mirror plane π_m. This is usually the robot approaching angle towards the mirror plane. It is important to know how α affects the estimation accuracy of π_m for the collision avoidance purpose. Data are collected in 6 different α values ranging from $5°$ to $60°$. Scene structure is kept to be the same during the test (see Fig. 3(b)) with abundant features. The baseline distances between the first and second views are 25.4cm while maintaining the same optical axis. Ground truth data are obtained using physical measurements.

Fig. 4 illustrates that both angular errors of mirror plane normal and relative depth errors are reasonably small under different α values. Note that 100 trials have been carried out for each α setting. The results are desirable because errors are not sensitive to α values. Note that we have not
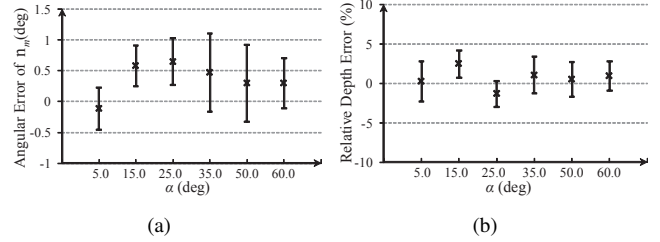

Fig. 4. The accuracy of estimated mirror plane with respect to α values: (a) Angular error of the mirror normal. (b) Relative depth error for the mirror plane. The vertical bar and the middle cross represent the one standard deviation range and sample mean, respectively.

performed experiments for cases with large angle values (i.e. $\alpha > 60°$). At large angles, the camera/robot almost faces the mirror directly. Since a regular camera has a horizontal field of view larger than $55°$, the robot can see itself in the mirror. For such cases, the problem becomes trivial because it is reduced to self-appearance-based mirror detection, which is less challenging.
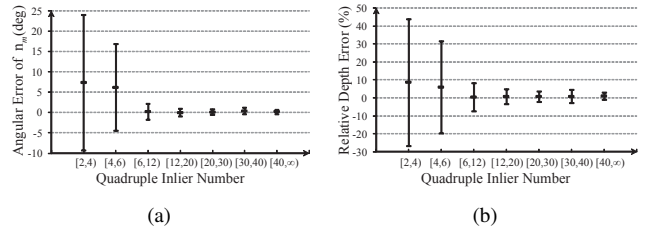

Fig. 5. The mean and standard deviation plot of mirror plane parameters vs. number of quadruple inliers: (a) Angular error of the mirror normal. (b) Relative depth error for the mirror plane. The vertical bar represents one standard deviation range.

The second experiment is to explore the relationship between the quadruple inlier number and the estimation accuracy and hence determine threshold T_N in Algorithm 2. We use the same data set from the first experiment. For every pair of images, the mirror parameters are computed each time as the number of quadruple inliers is changed through incrementally adjusting the ASIFT feature detection threshold. Then we group the estimation results according to their corresponding quadruple inlier numbers and compare the estimation error across groups. The results are shown in Fig. 5. As expected, the standard deviation of estimation generally decreases as the quadruple inlier number increases. When the quadruple inlier number drops below 6, the estimation accuracy becomes untrustable due to its large standard deviation. Hence we set $T_N = 6$ for our field tests.

B. Field Tests

We have tested our algorithm in the field. A data set of 100 pairs of images is collected from real world scenes with or without mirrored walls, such as gymnasiums, corridors, campus, and shopping malls (see Fig. 6). In the data set, 50% of the image pairs contain mirrored walls such as wall mirrors, window glasses, and water surfaces.

Fig. 6. Sample images from the data set.

		Predicted	
		Positive	Negative
Actual	Positive	45	5
	Negative	4	46

The detection result is presented in a confusion matrix in Table I, where "Positive" indicates the existence of mirrored walls. In the confusion matrix, true positive rate and true negative rate are both high, indicating desirable recognition ability. The false positive cases are typically caused by objects with strong symmetric appearances, e.g., sample image 12 in Fig. 6. The false negative cases are mainly due to lack of features in the scene. The overall detection accuracy is 91.0%.

VII. CONCLUSION AND FUTURE WORK

We addressed PMDP using two views from an on-board camera. First, we derived geometric constraints for corresponding real-virtual features across two views. Based on the geometric constraints, we employed RANSAC framework and ASIFT to develop a robust mirror detection algorithm. We implemented the algorithm and tested it under both in-lab and field settings. The algorithm has achieved an overall accuracy of 91.0%. In the future, we will study how to segment the mirror region out of the background image. This is important for object recognition (e.g. glassy doors or windows).

REFERENCES

[1] NSTC, "National science and technology council, Federal research and development agenda for net-zero energy, high-performancegreen buildings," http://www.bfrl.nist.gov/buildingtechnology/documents/FederalRDAgendaforNetZeroEnergyHighPerformanceGreenBuildings.pdf, Oct. 2008.

[2] W. G. Walter, "An imitation of life," *Scientific American*, vol. 182, no. 2, pp. 42–45, 1950.

[3] G. Gallup, "Chimpanzees: Self-recognition," *Science*, vol. 167, no. 3914, pp. 86–87, January 1970.

[4] F. Patterson and W. Gordon, "The case for the personhood of gorillas," *The great ape project: Equality beyond humanity. London: Fourth Estate*, pp. 58–77, 1993.

[5] D. Reiss and L. Marino, "Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence," *PNAS*, vol. 98, no. 10, pp. 5937–5942, May 2001.

[6] H. Prior, A. Schwarz, and O. Gntrkn, "Mirror-induced behavior in the magpie (pica pica): Evidence of self-recognition," *PLoS Biol*, vol. 6, no. 8, p. e202, 08 2008.

[7] M. Hauser, C. Miller, K. Liu, and R. Gupta, "Cotton-top tamarins (saguinus oedipus) fail to show mirror-guided self-exploration," *American journal of primatology*, vol. 53, no. 3, pp. 131–137, 2001.

[8] P. Michel, K. Gold, and B. Scassellati, "Motion-based robotic self-recognition," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3. Sendai, Japan: IEEE, September 2004, pp. 2763–2768.

[9] P. Haikonen, "Reflections of consciousness: The mirror test," in *Proceedings of the 2007 AAAI Fall Symposium on Consciousness*, Arlington, Virginia, USA, November 2007, pp. 67–71.

[10] J. Gluckman and S. Nayar, "Catadioptric stereo using planar mirrors," *International Journal of Computer Vision*, vol. 44, no. 1, pp. 65–79, 2001.

[11] G. Mariottini, S. Scheggi, F. Morbidi, and D. Prattichizzo, "Planar catadioptric stereo: single and multi-view geometry for calibration and localization," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. Kobe, Japan: IEEE, May 2009, pp. 1510–1515.

[12] K. N. Kutulakos and E. Steger, "A theory of refractive and specular 3d shape by light-path triangulation," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 13–29, January 2008.

[13] S. Rozenfeld, I. Shimshoni, and M. Lindenbaum, "Dense mirroring surface recovery from 1d homographies and sparse correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 325–337, February 2011.

[14] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 73–82, January 2004.

[15] S. Rahmann and N. Canterakis, "Reconstruction of specular surfaces using polarization imaging," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, Kauai, Hawaii, USA, December 2001.

[16] M. Oren and S. K. Nayar, "A theory of specular surface geometry," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 105–124, September 1997.

[17] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 4, pp. 91–110, Nov. 2004.

[18] R. Ma, J. Chen, and Z. Su, "Mi-sift: mirror and inversion invariant generalization for sift descriptor," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. Xi'an, China: ACM, July 2010, pp. 228–235.

[19] X. Guo and X. Cao, "Find: A neat flip invariant descriptor," in *2010 International Conference on Pattern Recognition*. Istanbul, Turkey: IEEE, August 2010, pp. 515–518.

[20] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, September 2004.

[21] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.

[22] J. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, April 2009.

[23] A. Agha-mohammadi and D. Song, "Robust recognition of planar mirrored walls using a single view," in *Robotics and Automation, Proceedings 2011 IEEE International Conference on*. Shanghai, China: IEEE, May 2011, pp. 1186–1191.

[24] Y. Lu, D. Song, Y. Xu, A. Perera, and S. Oh, "Automatic building exterior mapping using multilayer feature graphs," in *IEEE International Conference on Automation Science and Engineering (CASE), Madison, Wisconsin*, Aug. 2013.

[25] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Pr, 2003.

[26] G. Yu and J.-M. Morel, "ASIFT: An Algorithm for Fully Affine Invariant Comparison," http://www.ipol.im/pub/algo/my_affine_sift/, 2011.