

Structure and Pose from Single Images of Symmetric Objects with Applications to Robot Navigation

Allen Y. Yang Wei Hong Yi Ma
Electrical & Computer Engineering Department
University of Illinois at Urbana Champaign
Email: {yangyang, weihong, yima}@uiuc.edu

Abstract

In this paper, we provide a principled and unified explanation how knowledge in global 3-D structural invariants, typically captured by a group action on a symmetric structure, can significantly facilitate the task of reconstructing a 3-D scene from one or more images. More importantly, the “absolute” pose between the camera frame and the canonical frame that a symmetric structure admits can be recovered, which explains why symmetric objects (e.g., buildings, hallways) provide us overwhelming clues to their orientation and position. We give the necessary and sufficient conditions under which this pose can be uniquely determined, and, when such conditions are not satisfied, exactly to what extent this pose can be recovered. We show how algorithms from conventional multiple-view geometry, after properly modified and extended, can be effectively applied to perform such recovery. Since now structure, pose and even camera calibration can be recovered from a single image, the techniques naturally apply to vision-based robot navigation where global position and orientation is important.

1 Introduction

In robot vision (and computer vision) literature, reconstruction of scene structure and camera pose from images is one of the fundamental problems that has been studied extensively for the past two decades or so and it is related to a variety of important applications in robotics, such as autonomous navigation. In most extant methods in robot vision, scene structure and camera pose (or motion) are computed based on structure from motion techniques (e.g., [4]) by exploiting geometric relationships such as the epipolar constraint among images. However, a well-known dilemma associated with those methods is that: with short baselines between images (which is often the case in robot vision), tracking features is relatively simple, but multiple-view geometric algorithms are often ill-conditioned; with large baselines, these algorithms become robust to noise

but establishing correspondence across multiple images is often a time-consuming task (even if it can now be done automatically to certain extent).

To resolve this problem, we take a somewhat extreme stand and pose the question: Can we recover all the 3-D information, including scene structure, camera pose and even camera calibration from a single image? If so, we can very much avoid the difficult tracking or matching problem, and on the other hand, we hope (as we will soon see) the resulting algorithms will be robust. How can this be possible? We must rely on something which will make our problem well-posed. Let us first take a look at some examples in Figure 1 below. It is not hard to convince ourselves that even from a single image, we are able to perceive the 3-D structure and relative orientation of the object involved, although the object can be distorted by the perspective projection.

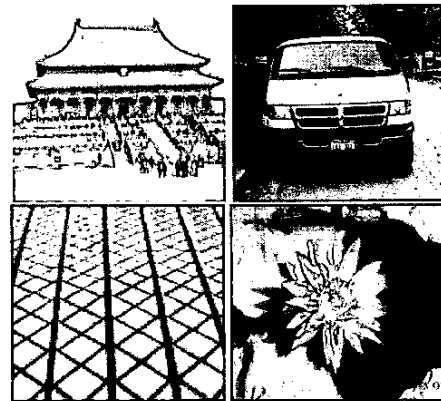


Figure 1: Examples of symmetric structures.

We note that all the objects in these examples exhibit something in common. They are all *symmetric* (or at least so can we assume). Symmetry, as a global structural invariant, turns out to be a key factor which may resolve almost all the problems we mentioned above in structure from motion. It in fact gives us very accurate and robust 3-D infor-

mation even when parallax (or stereo) information is not available. Figure 2 below demonstrates from another aspect that *symmetry is a fundamental modeling assumption that the human vision system often imposes on the environment, even if that means we occasionally get fooled by it.* This is probably the reason why, symmetry, as a useful ge-



Figure 2: Ames room, which is located in San Francisco Exploratorium (www.exploratorium.edu).

ometric cue to 3-D information, has been a traditional topic in cognitive science [11, 12].

In computer vision, symmetry has been exploited as an effective cue to compute surface orientation or camera position. But it is traditionally studied under the context of *texture* such as the study of isotropic (or weak isotropic) and homogeneous textures [3, 19, 1, 10, 13, 8]. It is the work of [1, 2] that provided people a wide range of efficient algorithms to recover the orientation based on the assumption of (weak) isotropy. [13] gave out a detailed comparison of isotropy and homogeneity in texture analysis.

On the geometric front, both the bilateral symmetry [16, 21, 14, 20] and the rotational symmetry [20] have been shown useful for the reconstruction of 3-D information from 2-D images. Other symmetry related partial scene knowledge, such as "orthogonality", "parallelism" and "vanishing points", has also been extensively exploited for the same purpose (e.g., see [7]). One step further, people sometimes use explicit 3-D models to compute the orientation of objects (e.g., human face, vehicle, building, etc.) [6, 15].

In this paper, we show, for the first time, that a unified theory for the study of perspective images of all types of symmetry (reflective, rotational, translational) can be obtained through the combination of the conventional multiple-view geometry with symmetry group actions. As a consequence, conventional multiple-view geometric algorithms and techniques, with slight generalization, can be directly applied to perform 3-D reconstruction from even a single image of symmetric objects. Furthermore, the theory also explains why the resulting algorithms are much better conditioned. This provides both a theoretical and an algorithmic foun-

ation for robot recovering 3-D pose and structure from an image of its surroundings.

2 Multiple-view geometry for symmetric structures

In this section we present our main theoretical results on multiple-view geometry for symmetry structure. For details of the proofs, the reader may refer to the technical report [5]. First, we give a mathematical definition of symmetry which takes into account all types of symmetry uniformly.

Definition 1 A set of 3-D points $S \doteq \{p\} \subset \mathbb{R}^3$ is called a symmetric structure if there exists a non-trivial subgroup G of the Euclidean group $E(3)$ that acts on it. That is, for any element $g \in G$, it defines an isomorphism (i.e. a one-to-one invertible map) from S to itself, $g \in G : S \rightarrow S$. If so, we say the structure S has a symmetry group G .

In particular, we have $g(S) = g^{-1}(S) = S$ for any $g \in G$. Mathematically, symmetric structures and groups are equivalent ways to capture symmetry: any symmetric structure is invariant under the action of its isomorphic group; any group (here as a subgroup of $E(3)$) defines a class of (3-D) structures that are invariant under this group action. Here we emphasize that G is a subgroup of the Euclidean group $E(3)$ but in general not of the special one $SE(3)$. The reason is, many symmetric structures that we are going to consider are invariant under reflection which is an element in $O(3)$ but not in $SO(3)$. For simplicity, here we only consider G to be a discontinuous (or discrete) group. It has been shown that any isometric symmetry in \mathbb{R}^3 falls into one of the following three types (or a combination of them) [18]: *reflection, rotation, and translation*.

In the following, we will first introduce multiple-view rank condition for a symmetric structure and the necessary and sufficient condition under which a unique reconstruction is possible, and we will present algorithms which compute the structure and pose by exploiting reflective, rotational, and translational symmetry, respectively.

2.1 Symmetric multiple-view rank condition

Using the homogeneous representation of $E(3)$, any element $g = (R, T)$ in the group G can be represented as a 4×4 matrix of the form

$$g = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4 \times 4}. \quad (1)$$

where $R \in \mathbb{R}^{3 \times 3}$ is an orthogonal matrix and $T \in \mathbb{R}^3$ is a vector. Note that in order to represent the group G in

this way, a world coordinate frame must have been chosen. For instance, if the symmetry is only rotational, it is natural to choose the origin of the world frame to be the center of symmetry and one of the coordinate axes to be the rotation axis. Often such a choice results in a simple canonical representation of the symmetry.

Now suppose an image of a symmetric structure S is taken from a vantage point $g_0 = (R_0, T_0) \in SE(3)$. Although the choice of the world coordinate frame in principle can be arbitrary, the relative pose g_0 from the viewer to the object, as we will soon see, can be uniquely determined from a single image as long as there is “enough” symmetry in the structure and it is “naturally” represented in the canonical world coordinate frame. The image of S is a set of image points $\mathcal{I}_0 \doteq \{\mathbf{x}\} \subset \mathbb{R}^3$. In homogeneous coordinates each image point \mathbf{x} satisfies

$$\lambda \mathbf{x} = \Pi_0 g_0 \mathbf{X} = [R_0, T_0] \mathbf{X}. \quad (2)$$

where $\Pi_0 \doteq [I, 0] \in \mathbb{R}^{3 \times 4}$, $\mathbf{X} \in \mathbb{R}^4$ is homogeneous coordinates of a point $p \in S$, and $\lambda \in \mathbb{R}$ is the depth.

Since $g(S) = S$ for all $g \in G$, we have $g(\mathcal{I}_0) = \mathcal{I}_0$. Although for an image \mathbf{x} of a point p , at the different vantage point $g_0 g$, it may become a different image, say \mathbf{x}' , but \mathbf{x}' must coincide with one of the points in the image \mathcal{I}_0 taken at the original vantage point. That is $\mathbf{x}' \in \mathcal{I}_0$. If we denote $\mathbf{x}' = g(\mathbf{x})$, then the group G also acts on the image set \mathcal{I}_0 , as it does on S . For simplicity, we always assume $g(\mathbf{x}) \in \mathcal{I}_0$ for all $\mathbf{x} \in \mathcal{I}_0$ and all $g \in G$. Hence the equation (2) yields a system of equations

$$\lambda' g(\mathbf{x}) = \Pi_0 g_0 g \mathbf{X}, \quad \forall g \in G. \quad (3)$$

Therefore, if we know the type of symmetry G in advance and know how its element acts on points in \mathcal{I}_0 , every image point in the set $\{g(\mathbf{x}) : g \in G\}$ then can be interpreted as a different image of the same point \mathbf{X} seen from different vantage points. That is, we effectively have as many as $|G|$ images of the same 3-D structure S “hidden” in the given one.¹

Now let $\{g_i = (R_i, T_i)\}_{i=1}^m$ be m different elements in the group G . Then, one image $\mathbf{x} \sim \Pi_0(g_0 \mathbf{X})$ (here “ \sim ” means equal up to a scale) of a symmetric structure with the symmetry G is equivalent to at least m different images that satisfy the following equations

$$\begin{aligned} g_1(\mathbf{x}) &\sim \Pi_0 g_0 g_1 g_0^{-1}(g_0 \mathbf{X}), \\ g_2(\mathbf{x}) &\sim \Pi_0 g_0 g_2 g_0^{-1}(g_0 \mathbf{X}), \\ &\vdots \\ g_m(\mathbf{x}) &\sim \Pi_0 g_0 g_m g_0^{-1}(g_0 \mathbf{X}). \end{aligned}$$

¹Here we use $|G|$ to denote the cardinality of the group G . In particular, when G is finite, $|G|$ is the number of elements in G .

It is well known from multiple-view geometry that these “hidden” images must be related by the so-called multiple-view rank condition [9]. That is, the so-called *multiple-view matrix*

$$M_g(\mathbf{x}) \doteq \begin{bmatrix} \widehat{g_1(\mathbf{x})} R(g_1) \mathbf{x} & \widehat{g_1(\mathbf{x})} T(g_1) \\ \widehat{g_2(\mathbf{x})} R(g_2) \mathbf{x} & \widehat{g_2(\mathbf{x})} T(g_2) \\ \vdots & \vdots \\ \widehat{g_m(\mathbf{x})} R(g_m) \mathbf{x} & \widehat{g_m(\mathbf{x})} T(g_m) \end{bmatrix} \quad (4)$$

where

$$\begin{aligned} R(g_i) &\doteq R_0 R_i R_0^T \in O(3), \\ T(g_i) &\doteq (I - R_0 R_i R_0^T) T_0 + R_0 T_i \in \mathbb{R}^3, \end{aligned} \quad (5)$$

for $i = 1, 2, \dots, m$, satisfies the multiple-view rank condition

$$\boxed{\text{rank}(M_g(\mathbf{x})) \leq 1, \quad \forall \mathbf{x} \in \mathcal{I}_0.} \quad (6)$$

Note that this rank condition is independent of any particular order of the group elements g_1, g_2, \dots, g_m and it captures the only fundamental invariant in a perspective image of a symmetric structure. Thus, we call it the *symmetric multiple-view rank condition*. Note that if $G \subseteq O(3)$ (i.e. $T_i = 0$ for all i), the expression for $T(g_i)$ is simplified to

$$T(g_i) = (I - R_0 R_i R_0^T) T_0, \quad i = 1, 2, \dots, m. \quad (7)$$

Therefore, one image of a symmetric structure S with its symmetry group G is equivalent to $m = |G|$ images of $n = |S|$ feature points.² The reconstruction of $(R(g_i), T(g_i))$ can be easily solved by a factorization algorithm based on the above rank condition [9]. Nevertheless, in order to solve for the “initial pose” $g_0 = (R_0, T_0)$, we need to further solve them from the equation (5) once $R(g_i)$ and $T(g_i)$ are recovered. To solve for R_0 , it requires us to solve a system of *Lyapunov* type equations:

$$R(g_i) R_0 - R_0 R_i = 0, \quad i = 1, 2, \dots, m \quad (8)$$

with $R(g_i)$ and R_i known. We denote the solution set to each Lyapunov equation as $\text{Ker}(L_i)$, $i = 1, 2, \dots, m$.

2.2 Necessary and sufficient condition for a unique pose recovery from symmetry group

The uniqueness of the solution R_0 depends on the relationship between the sets $\{R(g_i) = R_0 R_i R_0^T\}$ and $\{R_i\}$. In general we have:

Lemma 2 (Rotational and reflective symmetry group).

Given a (discrete) subgroup G of $O(3)$, a rotation R_0 is uniquely determined from the pair of sets $(R_0 G R_0^T, G)$ if and only if the only fixed point of G acting on \mathbb{R}^3 is the origin.

²It is possible that both $|G|$ and $|S|$ are infinite. In practice, one can conveniently choose only a finite subset of them in algorithms.

We refer the reader to the report [5] for a proof. Once R_0 is determined, it is then not difficult to show that, with respect to the same group G , T_0 can be uniquely determined from the second set of equations in (5). Thus, as a consequence to the above lemma, we have the following statement:

Theorem 3 (Unique initial pose from symmetry group).

If a 3-D symmetric structure S allows a rotational or reflective symmetry subgroup G which fixes only the origin of \mathbb{R}^3 , the initial pose g_0 can always be uniquely determined from one image of S .

Note that the group G does not have to be the only symmetry that S allows – as long as a subgroup of the total symmetry group of S satisfies above conditions, g_0 can be uniquely recovered.

From multiple-view geometry, we know that if the structure S happens to be a planar structure, the above results will need modification. This is because, planar structure is critical configuration for generic multiple-view algorithms (e.g., the standard 8-point algorithm will break down). Thus, we can no longer directly apply the above theorem to this case since we do not necessarily have all the elements of $g_0 G g_0^{-1}$. In particular, if $g \in G$ represents the reflection of the planar structure about its supporting plane, it fixes all points on the planar and hence, from the image point of view, g is not distinguishable from the identity map I in G . But having a planar structure is not necessary a bad thing: We can restrict the origin of the world frame to be inside this plane and the plane normal to be one of the axes (often the z -axis). Done so, the relative pose g_0 between the viewer to the plane can be specified often with less ambiguity.

For a reflection R with respect to a plane, we call the normal vector to its plane of reflection as the *axis of the reflection*.³ Using this notion, we have the following statement:

Corollary 4 (Initial pose from symmetry group of a planar scene). *If a planar symmetric structure S allows a rotational or reflective symmetry subgroup G (not including the reflection with respect to the supporting plane of S itself) with two independent rotation or reflection axes, the initial pose g_0 can always be uniquely determined from one image of S (with the world frame origin o restricted in the plane and its z -axis as the plane normal).*

As a consequence, to have a unique solution for g_0 , a planar symmetric structure S must allow at least two reflections

³The role of the axis of a reflection is very similar to that of the axis of a rotation once we notice that, for any reflection R , $-R$ is a rotation of angle $\theta = \pi$ about the same axis.

with independent axes, or one reflection and one rotation (automatically with independent axes for a planar structure).

Under the guideline of the general theory above, let us study each individual type of symmetry, i.e. the reflective, rotational, and translational symmetry, and show explicitly, to what extent the pose g_0 can be recovered if the necessary and sufficient condition is violated.

2.3 Reflective symmetry

In this subsection, we develop in detail for the reflective symmetry results regarding reconstruction from one image of a symmetric object. For rotational and translational symmetry, we will only present the results and refer the derivation to the technical report [5].

Figure 3 shows a reflective symmetry of a checker board and the correspondence between features before and after the symmetry is applied. Without loss of generality, we

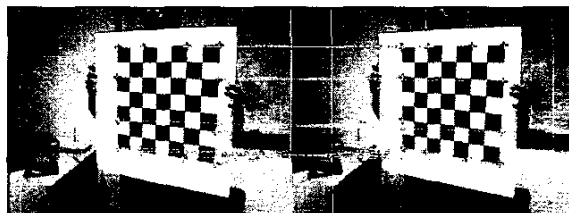


Figure 3: Left and right show the same image of a checker board with a vertical reflection (of feature points).

suppose the plane of reflection is the yz -plane of a pre-chosen world coordinate frame, or equivalently the reflection axis is the x -axis. Then the reflection can be described by the motion $g = (R, 0)$ where

$$R = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3} \quad (9)$$

is an element in $O(3)$ and it has $\det(R) = -1$. If one image of such a symmetric object is taken at $g_1 = (R_0, T_0)$, then we have the following two equations for each image point on this structure:

$$\lambda \mathbf{x} = \Pi_0 g_0 \mathbf{X}, \quad \lambda' g(\mathbf{x}) = \Pi_0 g_0 g \mathbf{X}. \quad (10)$$

We define $R' \doteq R_0 R R_0^T$ and $T' \doteq (I - R_0 R R_0^T) T_0$. Then the symmetric multiple-view rank condition, in the two view case, is reduced to the following epipolar constraint:

$$g(\mathbf{x})^T \widehat{T'} R' \mathbf{x} = 0. \quad (11)$$

Note that here $\det(R') = \det(R) = -1$ is not a rotation matrix. Therefore the matrix $E = \widehat{T'} R'$ is not a conven-

tional essential matrix and the 8-point (or 7-point) algorithm needs to be modified accordingly, if it is to be used to recover (R', T') .

Once $R' = R_0 R R_0^T$ is obtained, we can use it, together with R , to recover R_0 . The associated Lyapunov equation is

$$R' R_0 - R_0 R = 0. \quad (12)$$

In [5], we have proved that $\dim(\text{Ker}(L)) = 5$ in this case, but the intersection of $\text{Ker}(L)$ with $SO(3)$ is a 1-parameter family which corresponds to an arbitrary rotation of the y, z -axes inside the plane of reflection. The translation T_0 is recovered up to the form

$$T_0 = (I - R_0 R R_0^T)^\dagger T' + \beta v_2 + \gamma v_3, \quad \beta, \gamma \in \mathbb{R} \quad (13)$$

where $v_2, v_3 \in \mathbb{R}^3$ are two basis vectors for the null space of the rank-1 matrix $I - R_0 R R_0^T$.

In the case of Figure 3, the object is actually flat and all features (points, lines) lie in a 2-D plane, say P . This plane can be described by equation

$$N^T X = d \quad (14)$$

with $N \in \mathbb{R}^3$ as its normal vector and $d \in \mathbb{R}_+$ as the distance from the center of the camera to the plane P . In this plane, the image point x and its symmetric point $g(x)$ satisfy the planar homography [17]

$$\widehat{g(x)} H x = 0 \quad (15)$$

where the matrix H is the homography matrix of the form $H = R' + \frac{1}{d} T' N^T \in \mathbb{R}^{3 \times 3}$. With more than 4 points $\{x\}$, the matrix H can be recovered from the above equation (15). Decomposing the homography matrix H yields⁴

$$H \mapsto \left\{ R', \frac{1}{d} T', N \right\}. \quad (16)$$

If we choose the z -axis of the world frame to be the normal to the plane, the only solution for R_0 is

$$R_0 = \left[\pm v_1, \pm \widehat{N} v_1, N \right]. \quad (17)$$

where v_1 is the (only) real eigenvector of R' . If we further restrict the origin of the world frame to be in the plane itself, the ambiguity in T_0 is reduced from a 2-parameter family to a 1-parameter family which corresponds to an arbitrary translation along the y -axis (the intersection of the plane P and the plane of reflection).

⁴In general, one gets a total of four solutions from the decomposition, but only two of them are physically possible [17].

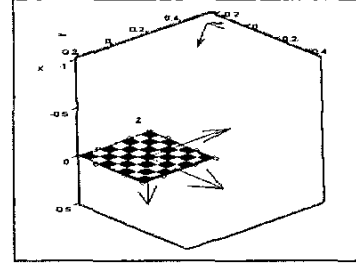


Figure 4: The reconstruction result from the reflective symmetry for Figure 3.

2.4 Rotational and translational symmetry

The checker board in Figure 3 also admits a set of rotational symmetry. Without loss of generality, we assume the rotation R of the form $R = e^{j\omega\theta}$ with $\|\omega\| = 1$ and $0 < \theta < \pi$, and hence it has three distinct eigenvalues $\{1, e^{+j\theta}, e^{-j\theta}\}$. In [5] we have proved that for the associated Lyapunov equation $\dim(\text{Ker}(L)) = 3$ but $\text{Ker}(L) \cap SO(3)$ is also a 1-parameter family which corresponds to an arbitrary rotation about the axis ω of R . The translation is recovered up to the form

$$T_0 = (I - R_0 R R_0^T)^\dagger T' + \beta v_1, \quad \beta \in \mathbb{R} \quad (18)$$

where $v_1 \in \mathbb{R}^3$ is in the null space of the rank-2 matrix $I - R_0 R R_0^T$.

For a planar structure however, with almost the same argument, we can recover R_0 as

$$R_0 = \left[u, \widehat{N} u, N \right], \quad \forall u \perp N. \quad (19)$$

After restricting the origin of the world frame to be inside the plane, T_0 is uniquely determined.

In translational symmetry case, we have $R = I$ and $T \neq 0$, and equation (5) is reduced to the following

$$R' = R_0 I R_0^T = I, \quad T' = R_0 T. \quad (20)$$

Thus, R_0 can be determined up to a 2-parameter family of rotations and T_0 is arbitrary.

If S is planar, the origin o of the world frame can be chosen in the supporting plane, the plane normal as the z -axis, and T as the x -axis. Thus $R_0 = [T', \widehat{N} T', N]$ and T_0 is determined up to an arbitrary translation in the plane.

We summarize in Table 1 ambiguity in determining the pose g_0 from each of the three types of symmetry, for both general and planar structures.

Due to the limit of space, we will not put here details about how symmetry may simplify the task of camera self-calibration. We only give the statement:

	$Ker(L)$	general	planar
Reflective	5-D	(1+2)*	(0+1)
Rotational	3-D	(1+1)	(1+0)
Translational	9-D	(2+3)	(0+2)

Table 1: Ambiguity in determining initial pose g_0 from three types of symmetry. (*: “ $(a + b)$ ” means there are an a -parameter family of ambiguity in R_0 of g_0 and b -parameter family of ambiguity in T_0 of g_0 .)

Remark 5 (Self-calibration from symmetry). *Not only is it possible to retrieve information about the camera intrinsic parameters from a single image of symmetric objects, but also the solution is always linear with each of the three types of symmetry applied.*

We refer the reader to the report [5] for details.

3 Experiments and applications

There are two distinctive features that allow symmetry-based techniques to be used in a wide range of applications. First, by imposing global structure invariants like symmetry, all multiple-view geometric algorithms become extremely robust. One way to understand this is that applying symmetry to a set of image feature typically results in a large displacement of features across the image (often in the range of hundreds of pixels). Thus, the signal to noise ratio is high. Another way to interpret this is that between the “hidden” images generated by a symmetry, the “baseline” of the corresponding 3-D motion is often large. Secondly, since reconstruction can now be done from a single image, tracking or matching features across multiple images is no longer needed. This dramatically reduces computational complexity, which is important in time-critical applications such as robot vision.

3.1 Global orientation for vision-based navigation

When we tour around a man-made environment, we typically have little problem orient and locate ourselves since objects of regular shape provide us overwhelming information about their relative position and orientation. Now such information can also be exploited by a robot (equipped with a vision system) using the techniques above. For instance, we now can easily compute the normal vectors of planes or surfaces surrounding the robot. These normal vectors provide a “global” reference of orientation and they can serve as a compass to the robot when navigating through the environment.

Figure 5 shows the result of reflective symmetry applied to outdoor surfaces. In the left image, we select nine sets of reflective structures, eight of them are associated to tiling patterns on the ground, and the last one is structural symmetry on the west wing of the Beckman Institute. We no-



Figure 5: Orientation recovery using reflective symmetry.

tice that results of the first and sixth sets are not fully correct. The explanation for it is that the reflective axes of them happen to be passing closely through the optical center of the camera, so the baseline between the “hidden” images is short. Thus, when we apply the 4-point algorithm to compute the homography matrix H , due to the noise, parallax information is not correctly extracted. Nevertheless, in this situation, the axis of reflection, say the (horizontal) x -axis, is always correct, and only the y, z -axes can be arbitrarily rotated in the plane of reflection, an ambiguity to be expected. To justify this explanation, we can change the reflective axis to be along the x -axis, and the result is shown in the image on the right. Now all the surface normals are correctly computed except for an immaterial ambiguity in the sign of the normal which is not chosen uniformly due to the way the algorithm is currently implemented.

We may also apply rotational symmetry assumption to the image (since the tiles on the floor are square), and the result is shown in Figure 6. For rotational symmetry, the only degenerate setting is when the optical axis of the camera lies up with the rotation axis, in which case the homography matrix H is a rotation matrix and the translation T , if computed anyhow, will be complete random. The disadvantage of rotational symmetry to reflective one is that in rotational case, only the normal to the surface (the z axis) can be determined as the rotation axis, but the x, y -axes can rotate arbitrarily.

Figure 7 shows an experiment of the translational symmetry. The reconstruction from translational symmetry is very robust, and it typically gives very accurate orientation of the surface. Another interesting application of translational symmetry is to test whether certain structures with the same reflective or rotational symmetry are within the same plane. For example, a staircase is shown in Figure 8. We first compute the frontal side for each stair using its reflective symmetry, and the normal vectors are all pointing towards us.

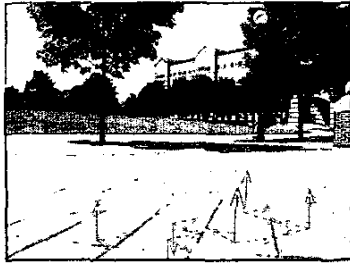


Figure 6: Surface normal recovered using rotational symmetry.

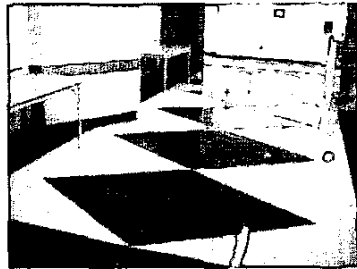


Figure 7: Floor normal recovered using translational symmetry.

But after we apply translational symmetry to two stairs together (as they were in the same plane), the recovered normal vector is pointing toward a different direction, in this case pointing a little upwards. This indicates (from the image alone) that these two surfaces are not in the same plane although they have the same surface normal and reflective symmetry with respect to the same plane. However, how

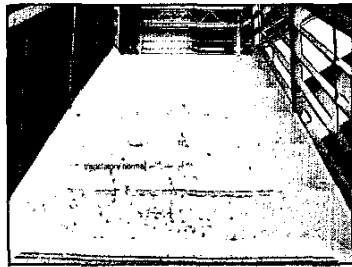


Figure 8: Pose of each stair is correctly computed as coordinate frames plotted on the right. If two stairs were treated as the same planar structure, the frame recovered (on the left) would be incorrect to either stair but more of the “normal” to the entire staircase.

all types of symmetry interplay with each other and give rise to a consistent 3-D representation for a scene (which may consist of many symmetric objects) remains as an open problem at this point. We hope this paper has provided people a good starting point from which many issues of this kind can be systematically addressed.

3.2 3-D object pose for visual surveillance

We can also apply our algorithm to detect vehicle poses in traffic surveillance. In model-based vehicle surveillance [15], the most computation-consuming part is to compute *pan angle* θ of the vehicle which indicates its heading. But as we notice, since vehicles are bilaterally symmetric, we can compute its heading using the techniques provided. In this experiment, features points (such as the signal lights) are chosen manually, and the results are shown in the following figure.



Figure 9: Vehicle pose detection using reflective symmetry.

3.3 Vision-based remote guidance

Currently, we are applying these techniques to vision guided mobile robot navigation. The algorithm can not only give us the robot motion (R, T) and the ground normal N , but at the same time the relative pose of the robot from the camera $g_0 = (R_0, T_0)$, which gives us directly the relative position of the robot and the camera. Thus, the camera can send guidance command to the robot in its own coordinate frame. We use this experiment to mimic a scenario where an aerial vehicle guides a ground robot. The camera is hung on the ceiling and pointing downwards, and we have a remote-controlled mobile robot with color markers. The reconstruction result is shown in Figure 10. As one can

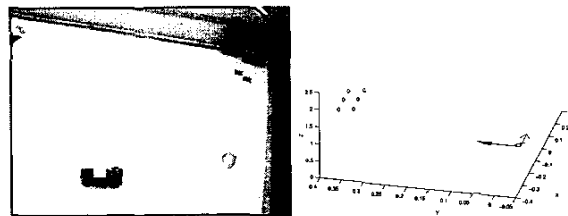


Figure 10: Using symmetry, the location and orientation of the robot can be fully recovered from each image.

see, although the size of the markers on the robot is relatively small, the reconstruction is considerably accurate.

4 Conclusions and discussions

Points, lines, and planes, as special symmetric objects, have been extensively studied before as geometric primitives for reconstructing a 3-D scene from 2-D images. This paper examines primitives at a higher level and with a larger scale: primitives that admit any type of symmetry. Symmetry is a more principled concept to unify various assumptions about 3-D structure that people have exploited in multiple-view geometry before, such as orthogonality and parallelism. Our results demonstrate that, with the extra symmetry assumptions, multiple-view geometric algorithms start to become extremely well-conditioned, due to almost always large baseline between “hidden images”. Nonlinear optimization (such as bundle adjustment) is no longer needed to improve the accuracy of the final reconstruction. In addition, since reconstruction can now be done reliably from even a single image, feature correspondence across different images is no longer a crucial step for 3-D reconstruction.

In experiments above, the knowledge about the presence of symmetry in the image is imposed manually. We are currently developing a system which can *automatically detect and match* 2-D and 3-D symmetry from image sequences, based on a combination of image segmentation and the presented multiple-view geometric techniques. Experimental results on real-world images have shown great promise and we will present these advances in future work. These new detection, matching, and reconstruction techniques for symmetric objects have significantly improved the quality and efficiency of image based 3-D reconstruction for man-made environments.

In this paper, the developed techniques mainly target at automatic map building or guidance for robot navigation. More examples and experimental results will be available at “<http://decision.csl.uiuc.edu/~yangyang/symmetry/>”. At the mean time, we also find these techniques extremely useful in applications such as camera self-calibration, image-based modeling, and understanding visual illusion.

References

- [1] J. Garding. Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, 2(4):327–350, 1992.
- [2] J. Garding. Shape from texture and contour by weak isotropy. *J. of Artificial Intelligence*, 64(2):243–297, 1993.
- [3] J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950.
- [4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2000.
- [5] W. Hong, A. Y. Yang, and Y. Ma. On group symmetry in multiple view geometry: Structure, pose and calibration from single images. *Technical Report, UILU-02-2208 DL-206, September 12, 2002*.
- [6] S. Jeng, H. Liao, Y. Liu, and M. Chern. An efficient approach for facial feature detection using geometrical face model. pages 426–430, 1996.
- [7] J. Košecká and W. Zhang. Efficient computation of vanishing points. In *IEEE Int. Conference on Robotics & Automation*, pages 223–228, 2002.
- [8] T. Leung and J. Malik. On perpendicular texture or: Why do we see more flowers in the distance? In *Int. Conference on Computer Vision & Pattern Recognition*, pages 807–813, 1997.
- [9] Y. Ma, J. Kosecka, and K. Huang. Rank deficiency condition of the multiple view matrix for mixed point and line features. In *Asian Conference on Computer Vision*, 2002.
- [10] J. Malik and R. Rosenholtz. Computing local surface orientation and shape from texture for curved surfaces. *Int. Journal on Computer Vision*, 23:149–168, 1997.
- [11] D. Marr. *Vision*. W. H. Freeman and Company, 1982.
- [12] S. E. Plamer. *Vision Science: Photons to Phenomenology*. The MIT Press, 1999.
- [13] R. Rosenholtz and J. Malik. Surface orientation from texture: Isotropy or homogeneity(or both)? *Vision Research*, 37(16):2283–2293, 1997.
- [14] C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J. L. Mundy. Extracting projective structure from single perspective views of 3d point sets. In *IEEE International Conference on Computer Vision*, pages 573–582, 1993.
- [15] T. Tan, G. D. Sullivan, and K. D. Baker. Efficient image gradient-based object localization and recognition. In *Int. Conference on Computer Vision & Pattern Recognition*, pages 397–402, 1996.
- [16] T. Vetter and T. Poggio. Symmetric 3d objects are an easy case for 2d object recognition. *Spatial Vision*, 8:443–453, 1994.
- [17] J. Weng, T. S. Huang, and N. Ahuja. *Motion and Structure from Image Sequences*. Springer Verlag, 1993.
- [18] H. Weyl. *Symmetry*. Princeton Univ. Press, 1952.
- [19] A. P. Witkin. Recovering surface shape and orientation from texture. *J. of Artificial Intelligence*, 17:17–45, 1988.
- [20] H. Zabrodsky, S. Peleg, and D. Avnir. Symmetry as a continuous feature. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 17(12):1154–1166, 1995.
- [21] H. Zabrodsky and D. Weinshall. Using bilateral symmetry to improve 3d reconstruction from image sequences. *Comp. Vision and Image Understanding*, 67:48–57, 1997.