

# Discovering and Exploiting 3D Symmetries in Structure from Motion

Andrea Cohen<sup>1</sup>   Christopher Zach<sup>2</sup>   Sudipta N. Sinha<sup>3</sup>   Marc Pollefeys<sup>1</sup>  
ETH Zurich<sup>1</sup>   Microsoft Research<sup>2</sup>   Microsoft Research<sup>3</sup>  
Switzerland   Cambridge, UK   Redmond, USA

## Abstract

Many architectural scenes contain symmetric or repeated structures, which can generate erroneous image correspondences during structure from motion (Sfm) computation. Prior work has shown that the detection and removal of these incorrect matches is crucial for accurate and robust recovery of scene structure. In this paper, we point out that these incorrect matches, in fact, provide strong cues to the existence of symmetries and structural regularities in the unknown 3D structure. We make two key contributions. First, we propose a method to recover various symmetry relations in the structure using geometric and appearance cues. A set of structural constraints derived from the symmetries are imposed within a new constrained bundle adjustment formulation, where symmetry priors are also incorporated. Second, we show that the recovered symmetries enable us to choose a natural coordinate system for the 3D structure where gauge freedom in rotation is held fixed. Furthermore, based on the symmetries, 3D structure completion is also performed. Our approach significantly reduces drift through "structural" loop closures and improves the accuracy of reconstructions in urban scenes.

## 1. Introduction

We address the problem of automatic recovery of scene structure and camera motion from multiple images, which is referred to as the structure from motion (SfM) problem. Significant progress has been achieved towards point-based SfM techniques, primarily in the area of efficient algorithms for scalable image matching [17], and large-scale bundle adjustment [1, 8, 16, 11, 28], which is a core component of all SfM approaches. As a result, it is nowadays possible to easily reconstruct large scenes from image sequences. The topic of urban 3D reconstruction and architectural modeling from images, in particular, has received considerable attention in the last decade [7, 10, 14, 22, 23, 29].

Man-made or architectural structures typically contain rich intrinsic symmetries and structural regularities by de-

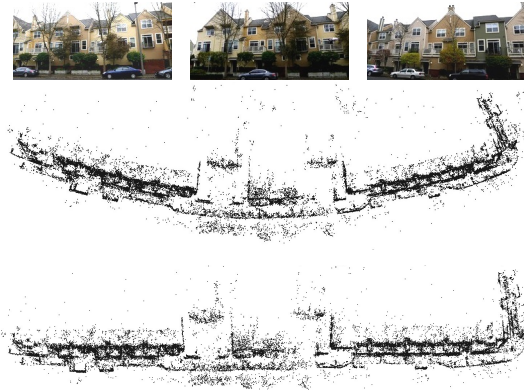


Figure 1. [TOP]: Symmetric building facades on a street. [MIDDLE]: Traditional SfM reconstruction from an open sequence. [BOTTOM]: Our method discovers 3D symmetries and enforces them during SfM to produce an accurate reconstruction.

sign. For example, repetitions of identical 3D structural elements are common in many building facades. Sometimes, multiple identical or mirrored instances of large substructures may exist in the scene, as in the case of two identical wings of a large building complex. Furthermore, the arrangement of these parts is rarely random, but often exhibits an underlying geometric regularity. In the past, various model-based techniques have attempted to exploit such rich geometric constraints in image-based architectural modeling [26, 7]. However the need for prior knowledge of the 3D geometry makes those methods difficult to use in practice. This limitation was recently addressed in [19, 20, 22], where symmetries and structural regularities in 3D geometry were automatically inferred from 3D models without using any prior knowledge.

In this paper, we address the goal of discovering symmetries and repetitions in the scene structure from multiple images and imposing these symmetry constraints to improve the accuracy of structure from motion algorithms. Starting from a set of images, our approach automatically recovers structural regularities and symmetry relations in the *unknown* 3D structure from visual feature correspondences in the images. These symmetry relations are then exploited

in various ways within a new constrained bundle adjustment formulation. We say that a symmetry relation exists, when two particular subsets of 3D points in the structure are related by a similarity transformation.

To our knowledge, there is very little prior work in structure from motion where symmetry relations are automatically recovered from multiple images and also used to impose structural constraints in bundle adjustment without prior knowledge of the scene. Unlike [19, 20, 22], neither a dense 3D model, nor a dense 3D point set is available in our case. Since we start from images and the 3D scene structure recovered by traditional SfM is uncertain, the recovery of symmetries in our case is more challenging. Also the SfM point clouds are sparser and more irregularly sampled in comparison to the 3D point sets used in prior work [22].

Our goal is also different from [19, 22]. We aim to recover symmetry relations in situations where the initial SfM reconstruction suffers from drift and can be inaccurate. Discovering symmetries using only geometric cues could be difficult in such cases. By discovering symmetry constraints within uncertain 3D structure, and imposing them during SfM, we are able to recover a much more accurate reconstruction where the inferred structural constraints are respected. Our approach for detecting 3D symmetries is based on image cues. We use pairwise image matches that contradict geometric relations between corresponding cameras, induced by a global SfM reconstruction, to detect subsets of 3D points that are symmetric to each other. Related image-based approaches for finding repetitions in multi-view sequences have proposed removing these contradicting matches during SfM [34, 35, 24]. In contrast, we show that these matches are extremely useful for discovering symmetry relations in the scene, which can be further exploited to achieve a much more accurate reconstruction.

To this end, we propose a new bundle adjustment formulation where structural constraints are enforced between various subsets of 3D point sets related by similarity transforms. Various symmetry priors are incorporated into this new formulation. The geometric constraints between distant symmetric point sets act like "structural" loop closures, which addresses the problem of drift in the reconstruction.

Symmetry knowledge in the 3D structure has another advantage. It allows a natural coordinate system for the structure to be chosen during bundle adjustment. Depending on the specific family of symmetries recovered, various degrees of gauge freedom can be held fixed in the bundle adjustment step. The similarity transformations induced by the symmetries have low model complexity when expressed in the natural coordinate system. Finally, we also show that the underlying symmetries can be used for 3D model completion. The 3D structure is completed with hallucinated 3D points which respect the symmetry transformations and this generates more densely sampled reconstructions.

## 2. Related Work

In earlier work on architectural modeling from images, model-based approaches were quite common. Symmetries and geometric regularities were enforced using parameterized primitives such as polyhedrons [7, 26], planes [2], and parallelepipeds [31]. Constrained 3D modeling was performed with these primitives in conjunction with Euclidean constraints [4], coplanarity constraints [25] and piecewise planar representations [2]. In contrast, almost all modern SfM approaches [10, 1, 6, 11, 8] represent the scene structure as an unstructured 3D point cloud. In ideal circumstances where enough overlapping images and accurate correspondences are available, the point-based SfM methods often work well. However, SfM algorithms are known to be prone to errors due to drift and instabilities caused by the lack of sufficient images or image observations<sup>1</sup>. Our work aims to incorporate some of the advantages of model-based constrained modeling techniques into modern bundle adjustment methods [1, 11, 8] but does not require prior knowledge of the scene structure. Other ways to improve SfM algorithms are known – such as using GPS [15]. Our work complements such techniques.

Various approaches have been proposed to detect symmetries in images. These include detecting regular lattices [21], repetitive structures [32], bilateral and mirror symmetries [18] in single images. Most recently, approaches for lattice detection from multiple views for non-planar scenes have also been proposed [13]. Mirrored descriptors were used in [18] to find mirror symmetries. Such symmetries detected in single images can be exploited for image-based modeling in specific cases [12, 33, 14].

There also exists prior work on symmetry detection in 3D geometry from range-scans or 3D models. These methods address the discovery of structural regularity in architecture [22], using lines-based features [3] and hierarchical reasoning [27] for symmetry detection. Approaches for enforcing symmetries on 3D models have also been investigated [20]. However as mentioned earlier, in order to exploit symmetries in structure from motion, our symmetry detection approach must deal with uncertain 3D structures and rely on image cues and multi-view constraints between different images to discover the existing symmetries. This is a key distinction between our method and existing approaches for image-based symmetry detection and geometric approaches for detecting symmetries in 3D data.

The symmetries we aim to recover can involve self-similarities or mirror symmetries of arbitrary 3D structures in the scene. The structural elements and their arrangements are therefore more general than lattices [21, 13] and are not limited to planar repetitions [32], which have been addressed in prior work. As addressed in [24, 34, 35], detect-

---

<sup>1</sup>Errors due to drift are often significant in open loop sequences.

ing these sort of symmetries and repetitions within a scene from uncalibrated images can be extremely challenging as a pair of images where the same structure instance is observed twice can be confused with an image pair where the different instances are observed.

### 3. Overview

The first stage of our approach involves performing robust SfM on the input image sequence, making sure our initial reconstruction does not get affected by the presence of repetitions or symmetries using an approach similar to [34, 35, 24]. A set of calibrated cameras, a set of 3D points and a set of 2D image observations for each 3D point are obtained. In Section 4 we describe how appearance cues (using SIFT feature matching) are used to find out the subsets of 3D points that contain a symmetry. This step produces a set of hypotheses for potential repetitive structures in the form of 3D similarity transformations. Once all the 3D transformations have been extracted, the symmetry planes and Euclidean transformations that best explain these matches are estimated. We also explain how the discovered symmetries allow us to select a natural coordinate system for the 3D structure. The symmetries and transformations are then used as geometric structural constraints in the bundle adjustment algorithm, which is described in Section 6. The natural coordinate system that was estimated is used for parameterizing the structure in our constrained bundle adjustment algorithm. Additionally, the symmetry constraints are used for adding more 3D points to the sparse 3D point cloud. This is described in Section 5.

### 4. Repetition and Symmetry Detection

The first step in our method is to detect repetitions and symmetries between different parts of the 3D model. This procedure can be divided into three main steps. First, we search for pairwise image matches that are erroneous but survived the RANSAC step. We refer to these matches as false positive pairwise matches. Then an additional step of image matching is performed to detect mirror symmetries on the 3D structure. Next, all the additional 2D-2D correspondences obtained from these two steps are used to estimate 3D similarity transformations corresponding to repetitions and symmetries. We also describe in this section how a natural coordinate frame can be estimated by using the discovered symmetry transformations.

**Search for false positive image matches:** With the initial sparse 3D point cloud and associated camera poses, it is possible to detect false positive matches between two images. More specifically, we are interested in the matches that allow us to compute a relative orientation that survives the geometric verification, but is not in agreement with the

ones induced by the globally consistent camera poses. Such pairwise image matches provide strong evidence of visual resemblance between two different parts of the model. False positive matched pairs are calculated in the following way: if  $\hat{E}_{i,i'}$  is the essential matrix<sup>2</sup> of a potential false positive, and  $E_{i,i'}$  is the one computed from the global poses (i.e.  $E_{i,i'} = [t_{i'} - R_{i,i'} t_i]_{\times} R_{i,i'}$  with  $R_{i,i'} = R_{i'} R_i^T$ ), then we consider the image pair  $(i, i')$  to be a false positive candidate if the essential matrices differ significantly (in our experiments we used:  $\|\hat{E}_{i,i'} - E_{i,i'}\|_F \geq 1.5$ .) Note that the use of a normalized baseline allows the essential matrices to be compared without ambiguity.

**Additional matching for symmetry detection:** Since symmetric structures are not only symmetric in terms of the underlying 3D geometry, but usually also symmetric in their appearance, we match original feature descriptors with their mirrored counterparts which are just extracted from the mirrored source images. Feature matching between original and mirrored descriptors including geometric verification is applied to obtain a set of putative reflective symmetries. Fig. 2 shows 2D feature correspondences for a mirror symmetric transformation obtained by this procedure.

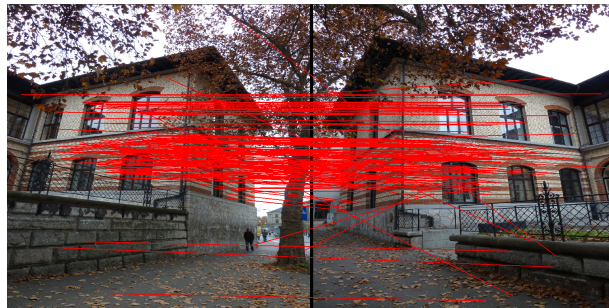


Figure 2. Geometrically consistent, mirror-symmetric feature matches between two wings of a building seen in different images, found using mirrored descriptors.

**Estimation of repetitions and symmetries:** For all false positive feature correspondences and mirrored descriptor matches that have (different) 3D points associated in the initial model, a putative 3D-3D correspondence, induced by a repetition or a planar symmetry, can be established. The set of all putative 3D-3D correspondences found is used, at first, to robustly estimate pure 3D translations and symmetries by exhaustive search. Sequential RANSAC is then used on the remaining 3D-3D correspondences to estimate general Euclidean transformations. In order to make the repetition and symmetry detection robust to different scales, inlier verification is done by using the projection of each 3D point on the image observations associated with the opposite 3D point. Fig. 3(a) depicts 3D correspondences verified

<sup>2</sup>We assume to have approximately calibrated cameras using EXIF tags.

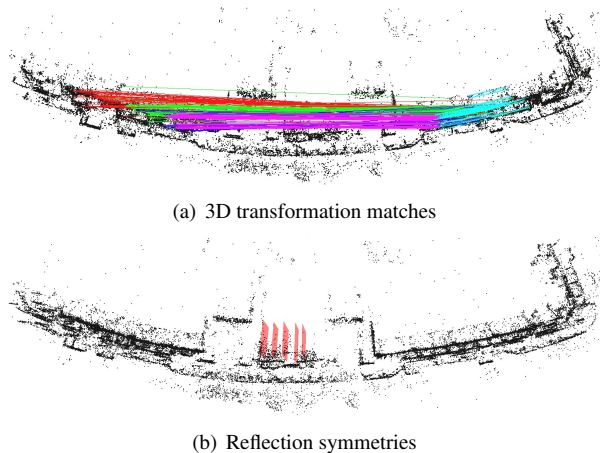


Figure 3. Estimated similarity transformations and symmetry planes for dataset 1.

via similarity transformations, and Fig. 3(b) illustrates extracted symmetry planes.

**Natural coordinate frame estimation:** Symmetries and repetitions found in architectural scenes provide a strong cue for the directions of the principal coordinate axes, which form a natural coordinate frame for the 3D structure. In particular, knowing the vertical direction allows us to place the 3D model, which usually resides in a gauge-free space, onto a natural ground plane. Of less importance, but still useful, can be the knowledge of the other principal directions, and of the center of the dominant symmetry, which can be thought of as a natural model origin. We consider the dominant symmetry to be the one with the largest number of 3D-3D correspondences. The vertical direction is estimated by exhaustively testing rotation axes and pairwise intersections of symmetry planes for the largest support. Translation vectors perpendicular to, and rotation axes and symmetry planes collinear with the hypothesized vertical direction support the current sample. Knowledge of the vertical direction fixes the ground plane normal. The principal direction in the ground plane is estimated in a similar manner<sup>3</sup>. With the estimated principal directions and the center of dominant symmetry, the selected coordinate frame will be a natural, object-centric one.

## 5. Model Completion

Knowledge of the symmetries intrinsic to the object of interest allows to hypothesize 3D patterns via those transformations, even if they are not directly seen in the images. Using a purely geometric plausibility approach, it would be possible to transfer 3D structures to completely unobserved parts (e.g. to “predict” the back side of a building that was not captured at all). In this work we focus on partial 3D

<sup>3</sup>One could also use the entropy method proposed in [23].

model completion such that any additional 3D point is seen in at least one of the registered images. This is naturally achieved in our system, as for a 3D point  $X_{j1}$  to be part of a detected symmetry  $k$  with associated transformation matrix  $M_k$ , it needs to have had at least a matching feature on the opposite side of the transformation. If enough observations were present on both sides of the transformation,  $X_{j1}$  would approximately map to some  $X_{j2}$  under transformation  $k$ . However, sometimes not enough observations were present to reconstruct  $X_{j2}$  (we require a minimum of 3 observations to triangulate points in our models). We can then construct  $\hat{X}_{j2}^k = M_k \circ X_{j1}$  and verify that it projects sufficiently close to the matching feature locations. If this is the case, the new 3D point  $\hat{X}_{j2}^k$  is added to the model. The matching features are attached to it so that it can be treated seamlessly like any other 3D point in the bundle adjustment described in the following section. Notice that we could also triangulate additional 3D points seen in only one or two images from each side of the transformation, but we refrain from this as those would typically not be very accurate.

## 6. Extended Bundle Adjustment

Standard bundle adjustment is a non-linear least-squares optimization step refining all estimated unknowns (extrinsic and optionally intrinsic camera calibration, 3D points) to match the observed image measurements according to an assumed noise model [28, 9, 8]. The objective is to minimize the distance between the reprojected 3D points and the corresponding image observations. Bundle adjustment can be also seen as an inference procedure in a graphical model represented by a factor graph (e.g. [6]). This view allows us to easily augment the standard objective function in bundle adjustment with additional terms representing priors we hypothesize on the final 3D model. Formally, let  $C_i$  denote the parameters of camera  $i$  and  $X_j$  be the  $j$ -th 3D point. The standard image reprojection error term is defined as

$$E_{\text{repro}} = \sum_{(i,j) \in \mathcal{M}} \rho_{\text{repro}}(\pi(C_i, X_j) - p_{ij})^2, \quad (1)$$

where  $\pi$  is the image projection function,  $\rho_{\text{repro}}$  an optional robust cost function, and  $p_{ij}$  is the observed measurement of  $X_j$  in image  $i$ . The sum runs over  $i, j$  with  $(i, j) \in \mathcal{M}$ , which is the set of observed image projections. We utilize the Huber cost function for  $\rho_{\text{repro}}$  by re-weighting the squared residuals. We extend the standard objective  $E_{\text{repro}}$  by adding several terms to incorporate the detected structural symmetries.

### 3D consistency of repeating and symmetric elements:

The detection of repetitions and symmetries (see Section 4) provides a list of inlier 3D point correspondences  $X_{j1} \leftrightarrow X_{j2}$  that are related either via a similarity transformation or

a reflective symmetry, e.g. if  $X_{j_1}$  and  $X_{j_2}$  are related by a transformation  $k$ , then

$$\sigma_{\text{transf}}^{k^{-1}} X_{j_2} = R_{\text{transf}}^k X_{j_1} + t_{\text{transf}}^k + \eta, \quad (2)$$

where  $R_{\text{transf}}$ ,  $t_{\text{transf}}$  and  $\sigma_{\text{transf}}$  are the parameters of the similarity transformation  $k$  (rotation, translation and scale resp.) and  $\eta$  is additive noise. We favor rigid/symmetric alignment in the final solution by adding appropriate terms to the cost.

One could directly add a term based on the 3D distance between points  $X_{j_2}$  and the image of  $X_{j_1}$  under the transformation, but this has the following disadvantage. The image reprojection term  $E_{\text{repro}}$  is typically measured in pixels whereas there is no natural unit for measuring 3D distances. Hence, it is difficult to choose an appropriate weight between  $E_{\text{repro}}$  and a penalty term measured in terms of 3D distances. Repetitions and symmetries may also be detected at very different scales within the same 3D model, e.g. symmetries could exist between two distant parts of a building but also between nearby facade elements at a smaller scale. Therefore the noise level in the 3D distance estimates can vary a lot. This would be difficult to capture solely with a 3D distance based cost.

Thus, it is better to use the 2D image observations corresponding to  $X_{j_1}$  to guide the alignment of  $X_{j_2}$ . This also naturally handles the position uncertainty of the 3D points  $X_{j_1}$  and  $X_{j_2}$  which is caused by the triangulation of 2D measurements in nearby cameras where the corresponding viewing rays are almost parallel in 3D. The additional term in the objective is as follows.

$$E_{\text{transf}} = \mu_{\text{transf}} \sum_{k \in \mathcal{T}} \rho_{\text{transf}}(\pi(\hat{X}_{j_1}^k, C_i) - p_{i,j_2})^2 + \mu_{\text{transf}} \sum_{k \in \mathcal{T}} \rho_{\text{transf}}(\pi(\hat{X}_{j_2}^k, C_i) - p_{i,j_1})^2, \quad (3)$$

where  $\hat{X}_{j_1}^k = R_{\text{transf}}^k X_{j_1} + t_{\text{transf}}^k$  and  $\hat{X}_{j_2}^k = R_{\text{transf}}^{k^{-1}}(X_{j_2} - t_{\text{transf}}^k)$  according to the  $k$ -th transformation parameters.  $\mathcal{T}$  is the set of all the detected euclidean transformations. The terms exist only if the corresponding image observation is available.  $\mu_{\text{transf}}$  is a weight parameter to balance between the reprojection term  $E_{\text{repro}}$  and this term. Since both terms are based on the same units, choosing  $\mu_{\text{transf}}$  is relatively noncritical. Since one is less confident in matches across transformations,  $\mu_{\text{transf}}$  will typically be at most 1. If  $\mu_{\text{transf}} = 1$ , the same image observations are associated with  $X_{j_1}$  and  $X_{j_2}$ , thus effectively identifying both 3D points as the same one. We choose  $\rho_{\text{transf}}$  again to be the Huber cost.

A similar term is added for the detected symmetries and the respective 3D point correspondences. It reads as

$$E_{\text{symm}} = \mu_{\text{symm}} \sum_{k \in \mathcal{S}} \rho_{\text{symm}}(\pi(\tilde{X}_{j_1}^k, C_i) - p_{i,j_2})^2 + \mu_{\text{symm}} \sum_{k \in \mathcal{T}} \rho_{\text{symm}}(\pi(\tilde{X}_{j_2}^k, C_i) - p_{i,j_1})^2, \quad (4)$$

where  $\mathcal{S}$  is the set of detected symmetries and  $\tilde{X}_j^k$  is the obtained 3D point after applying the  $k$ -th symmetry relation.

**Priors for man-made environments:** Man-made environments are dominated by collinear and orthogonal 3D elements, hence it is self-evident to add appropriate collinearity and orthogonality priors. Examples of such similarity transformations are pure translations or rotations in the ground plane with angles that are multiples of  $\pi/2$  ( $90^\circ$ ). Such transformations are favored over general 3D similarities. For a rotation angle  $\theta$ , let  $[\theta]$  denote the closest angle that is a multiple of  $\pi/2$ . For each detected similarity transformation we add the following prior,

$$E_\theta = \mu_\theta \sum_{k \in \mathcal{T}} \rho_\theta(\cos(\theta_k) - \cos([\theta_k]))^2, \quad (5)$$

where  $\theta_k$  is the rotation angle of the  $k$ -th estimated Euclidean transformation. We use the difference of the cosines in order to avoid numerical instabilities at  $\theta_k = 0$ . Since we only want to enforce the angular prior if  $\theta_k$  is close to  $[\theta_k]$ , we employ the non-convex but robust Cauchy M-estimator.

**Model complexity reduction:** Another feature of man-made environments is, that not only the 3D structure, but also the relations between symmetric and repeating parts are highly compressible. Thus, the overall model complexity to describe the transformation and symmetry parameters is expected to be small. One way to achieve complexity reduction in terms of the detected transformations is to introduce an additional cost term,

$$E_{\text{comp}} = \mu_{\text{comp}} \sum_{k,k' \in \mathcal{T}, k \neq k'} \rho_{\text{comp}}(d_{\mathcal{T}}(T_k, T_{k'})) + \mu_{\text{comp}} \sum_{k,k' \in \mathcal{S}, k \neq k'} \rho_{\text{comp}}(d_{\mathcal{S}}(S_k, S_{k'})). \quad (6)$$

$d_{\mathcal{T}}$  and  $d_{\mathcal{S}}$  are distance functions to compare transformations and symmetries, respectively.  $d_{\mathcal{T}}$  is proportional to the angle between the translation vectors for transformations with very similar rotation matrices.  $d_{\mathcal{S}}$  is proportional to the angle between the symmetry planes for symmetries with non-parallel planes, or to their plane distance in the opposite case. Since very different transformation parameters should not influence the solution, we use the Cauchy M-estimator for  $\rho_{\text{comp}}$  (i.e. very dissimilar transformation pairs are excluded). This generic term for complexity reduction can be replaced by the one described in the following paragraph, which aligns transformation parameters with a ‘‘natural’’ coordinate frame implied by the detected transformations and symmetries.

**Natural coordinate frame:** Another important assumption is that most symmetry planes and rotation axes should

be vertical, and many translations components should be horizontal. Given the initial orthonormal basis obtained in Section 4 the model can be transformed to the canonical coordinate frame e.g. with the vertical axis aligned with the  $z$ -axis. In this canonical frame, every transformation parameter (translation directions, rotation axes, symmetry plane normal) which is close to one of the principal directions, can be hypothesized to align exactly with the respective principal axis. In addition to the rotation angle prior we add a robust term favoring the transformation parameters to be collinear with the closest principal direction, e.g. we have for symmetry plane normals  $n_k$

$$E_{\text{symm-layout}} = \mu_{\text{layout}} \sum_{k \in S} \rho_{\text{layout}} \left( \cos \left( \angle(n_k, [n_k]) \right) \right), \quad (7)$$

where  $[n_k]$  is the closest principal direction ( $x$ ,  $y$ , or  $z$ -axis or its negated version). Since this prior should only be effective if  $n_k$  is close to  $[n_k]$ , we utilize the robust Cauchy cost for  $\rho_{\text{layout}}$ . These terms link transformations via the global coordinate frame, hence it indirectly leads to model complexity reduction without the need for the explicit pairwise terms in  $E_{\text{comp}}$  (but only for transformations that are almost aligned with the natural axis, e.g. diagonal repetitions are thereby not made collinear).

**Numerical Procedure:** We use a standard Levenberg-Marquardt method in combination with minimum-degree column reordering and sparse Cholesky decomposition to optimize the combined non-linear least squares objective. Mostly due to  $E_{\text{transf}}$  the approximated Hessian is not as sparse as the one derived solely from  $E_{\text{repro}}$ , but this exactly corresponds to loop closures in standard SfM. In the reduced camera matrix there is an additional non-zero block at  $(i, i')$  if there exists a transformation/symmetry linking 3D points  $X_j$  and  $X_{j'}$  with image observations in cameras  $i$  and  $i'$ , respectively. By detecting repeating and symmetric patterns we establish “structural” loop closures implied by similar 3D patterns, not identical 3D points. The transformation parameters induce a dense non-zero pattern in the respective Hessian, but this additional cost is negligible.

## 7. Results

In this section we present quantitative and qualitative results illustrating the improvement in our 3D reconstructions in comparison to traditional bundle adjustment where 3D symmetries are not enforced. We evaluated our method on four urban datasets where the number of images ranged from 99 to 191 images. We use a robust SfM approach suited for unstructured image collections to obtain the initial sparse 3D reconstruction and camera poses, and utilize the focal length estimates in the image EXIF tags to operate in a semi-calibrated setup. The approximate intrinsic parameter

estimates can produce distortions in the 3D reconstructions, especially when loop closures are not detected. Datasets 1–3 are shown in Figures 1, 5 and 6 whereas the fourth dataset is shown in the supplementary material. The 3D models estimated using standard bundle adjustment have significant distortion, e.g. significantly curved roads and facades (Figure 1 and 6) and lack of parallelism in the walls and other vertical sub-structures (see Figure 5).

We estimated symmetries in the underlying 3D geometry as described in Section 4 to obtain a set of similarity transformations (see Figures 3(a), 5(b) and 6(b)) and mirror symmetries (see Figures 3(b), 5(c) and 6(c)). In the four datasets, the number of extracted similarity transformations and reflection symmetries were between 15 to 50 and the number of reflection symmetries ranged from 5 to 30 respectively. Depending on the dataset, 3D model completion introduces only 2.5% new points or up to almost 20% additional 3D points (see Table 1, column 6 for the exact numbers). With the knowledge of putative 3D symmetries, our extended bundle adjustment is applied to rectify the model. We use  $\mu_{\text{transform}} = \mu_{\text{symm}} = 1$ . Unfortunately, the choice of  $\mu_{\theta}$ ,  $\mu_{\text{layout}}$  is rather critical and has a substantial impact on the result. We had to set  $\mu_{\theta} = \mu_{\text{layout}} = 100$  for data set 2, and to 1000 for the other ones.

Table 1 also displays the evolution of the bundle adjustment objectives. The total residual and the contribution of the image reprojection term,  $E_{\text{repro}}$ , are shown separately. As expected, the overall residual decreases, but only partially at the expense of the image reprojection term. For instance, in dataset 1,  $E_{\text{repro}}$  is decreased on one order of magnitude. We conjecture that sometimes the additional terms in the objective enable the initial solution to escape from a local minimum.

The degree of alignment of the resulting model with the major coordinate axis is illustrated in Figure 4, which displays histograms of angles between respective transformation parameters and the closest principal direction. Fig. 4(a) shows the histogram before our proposed bundle adjustment (with all datasets merged into a single histogram), and Fig. 4(b) shows the histogram after our adjustment. Note that all angles larger or equal to 10 degrees are collapsed into a single bin. These histograms show that the well aligned symmetries are moving closer to being perfectly aligned without affecting the rest of the structure. The improvement in the 3D reconstructions is evident from the visualizations in Figures 1, 5(e) and 6(e), which demonstrate significant straightening of the model layout and reduction in drift. Figure 7 shows the extracted natural coordinate and principal directions overlaid on our reconstructions.

## 8. Conclusions

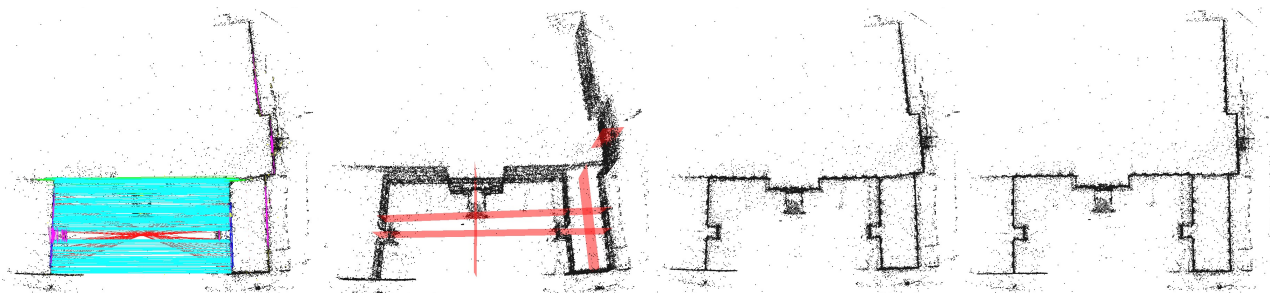
In this paper we have proposed a new approach for structure from motion, where symmetry relations in the

Dataset	#images	#3D points	init. image error	init. total error	#added points	final image error	final total error
1	175	43553	2.1751e+07	7.8620e+07	1497	2.1418e+06	1.0644e+07
2	186	47756	6.1872e+05	8.7939e+07	5605	4.8943e+06	2.7847e+07
3	99	31876	1.7700e+06	1.2126e+08	5747	6.7547e+06	3.3324e+07
4	191	60997	3.3342e+06	3.9434e+07	1556	2.4533e+06	5.7418e+06

Table 1. Dataset specification (first three columns),  $E_{\text{repro}}$  and the total objective before running our bundle adjustment, and  $E_{\text{repro}}$  and the total objective after convergence.



(a) Input images (out of 186)



(b) 3D transformation matches

(c) Reflection symmetries

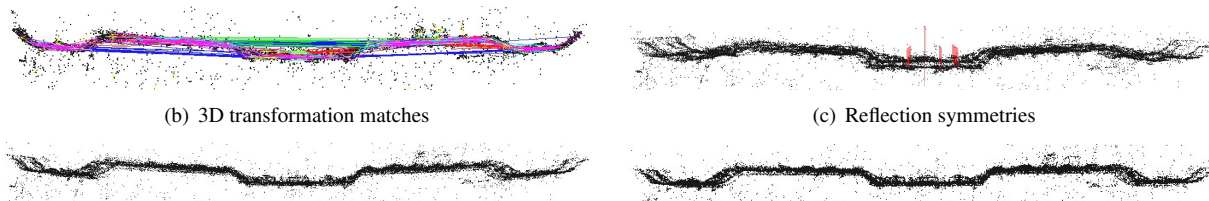
(d) Standard BA result

(e) Proposed BA result

Figure 5. Dataset 2 and respective top views.



(a) Input images (out of 99)



(b) 3D transformation matches

(c) Reflection symmetries

(d) Standard BA result

(e) Proposed BA result

Figure 6. Dataset 3 and respective top views.



(a) Dataset 2

(b) Dataset 3

Figure 7. Estimated natural coordinate frames and the overlaid sparse reconstruction.

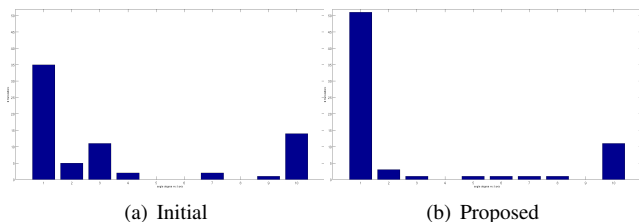


Figure 4. Histograms of angles indicating the alignment of transformations with respect to the closest principal direction. Histogram bins are in degrees, bin 10 includes angles  $\geq 10$  degrees. (a) Before the proposed BA, and (b) result of our adjustment.

3D structure are automatically recovered from multiple images. Structural constraints derived from those symmetries are imposed within a new constrained bundle adjustment formulation that incorporates robust priors on the expected model shape. We have demonstrated that our approach leads to improved appearance in various types of urban and architectural scenes. We also showed that for scenes where symmetries exist, a natural coordinate system can be used to parameterize the structure which has several advantages. Furthermore, discovering symmetries also facilitates 3D model completion.

## 9. Acknowledgements

We gratefully acknowledge the support of the 4DVideo ERC Starting Grant Nr. 210806.

## References

- [1] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski. Bundle adjustment in the large. In *ECCV (2)*, pages 29–42, 2010.
- [2] A. Bartoli and P. Sturm. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *IJCV*, pages 45–64, 2003.
- [3] M. Bokeloh, A. Berner, M. Wand, H.-P. Seidel, and A. Schilling. Symmetry detection using feature lines. *Comput. Graph. Forum*, 28(2):697–706, 2009.
- [4] D. Bondyfalat and S. Bougnoux. Imposing euclidean constraints during self-calibration processes. In *IN PROC. SMILE WORKSHOP*, pages 224–235, 1998.
- [5] M. Cummins and P. Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647, 2008.
- [6] F. Dellaert and M. Kaess. Square root SAM: Simultaneous location and mapping via square root information smoothing. *International Journal of Robotics Research*, 2006.
- [7] A. R. Dick, P. H. S. Torr, and R. Cipolla. Modelling and interpretation of architecture from several images. *Int. J. Comput. Vision*, 60:111–134, November 2004.
- [8] C. Engels, H. Stewenius, and D. Nister. Bundle adjustment rules. In *In Photogrammetric Computer Vision*, 2006.
- [9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.
- [10] A. Irschara, C. Zach, and H. Bischof. Towards wiki-based dense city modeling. In *ICCV*, pages 1–8, 2007.
- [11] Y. Jeong, D. Nister, D. Steedly, R. Szeliski, and I.-S. Kweon. Pushing the envelope of modern methods for bundle adjustment. In *CVPR*, pages 1474–1481, 2010.
- [12] N. Jiang, P. Tan, and L.-F. Cheong. Symmetric architecture modeling with a single image. *ACM Trans. Graph.*, 28:113:1–113:8, December 2009.
- [13] N. Jiang, P. Tan, and L.-F. Cheong. Multi-view repetitive structure detection. In *ICCV*, 2011.
- [14] K. Köser, C. Zach, and M. Pollefeys. Dense 3d reconstruction of symmetric scenes from a single image. In *DAGM-Symposium*, pages 266–275, 2011.
- [15] M. Lhuillier. Fusion of gps and structure-from-motion using constrained bundle adjustments. In *CVPR*, pages 3025–3032, 2011.
- [16] M. A. Lourakis and A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Software*, 36(1):1–30, 2009.
- [17] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, 2004.
- [18] G. Loy and J. olof Eklundh. Detecting symmetry and symmetric constellations of features. In *In ECCV*, pages 508–521, 2006.
- [19] N. J. Mitra, L. J. Guibas, and M. Pauly. Partial and approximate symmetry detection for 3d geometry. *ACM Trans. Graph.*, 25:560–568, July 2006.
- [20] N. J. Mitra, L. J. Guibas, and M. Pauly. Symmetrization. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH ’07, 2007.
- [21] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu. Deformed lattice detection in real-world images using mean-shift belief propagation. *PAMI*, 31(10):1804–1816, 2009.
- [22] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, and L. Guibas. Discovering structural regularity in 3D geometry. *ACM Transactions on Graphics*, 27(3):#43, 1–11, 2008.
- [23] M. Pollefeys, D. Nister, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, and H. Towles. Detailed real-time urban 3d reconstruction from video. *Int. J. Comput. Vision*, 78:143–167, 2008.
- [24] R. Roberts, S. N. Sinha, R. Szeliski, and D. Steedly. Structure from motion for scenes with large duplicate structures. In *CVPR*, pages 3137–3144, 2011.
- [25] R. Szeliski and P. H. S. Torr. Geometrically constrained structure from motion: Points on planes. In *SMILE’98*, pages 171–186, 1998.
- [26] C. Taylor, P. E. Debevec, and J. Malik. Reconstructing polyhedral models of architectural scenes from photographs. In *ECCV*, pages 659–668, 1996.
- [27] S. Thrun and B. Wegbreit. Shape from symmetry. In *ICCV*, pages 1824–1831, 2005.
- [28] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment a modern synthesis. In *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.
- [29] T. Werner and A. Zisserman. Model selection for automated architectural reconstruction from multiple views. In *BMVC*, pages 53–62, 2002.
- [30] T. Werner and A. Zisserman. New techniques for automated architecture reconstruction from photographs. In *ECCV*, pages 541–555, 2002.
- [31] M. Wilczkowiak, G. Trombtoni, C. Jermann, P. Sturm, and E. Boyer. Scene modeling based on constraint system decomposition techniques. In *ICCV*, volume II, pages 1004–1010, October 2003.
- [32] C. Wu, J.-M. Frahm, and M. Pollefeys. Detecting large repetitive structures with salient boundaries. In *ECCV: Part II*, pages 142–155, 2010.
- [33] C. Wu, J.-M. Frahm, and M. Pollefeys. Repetition-based dense single-view reconstruction. In *CVPR*, 2011.
- [34] C. Zach, A. Irschara, and H. Bischof. What can missing correspondences tell us about 3d structure and motion? In *CVPR*, 2008.
- [35] C. Zach, M. Klopschitz, and M. Pollefeys. Disambiguating visual relations using loop constraints. In *CVPR*, pages 1426–1433, 2010.