Repetition-based Dense Single-View Reconstruction

Changchang Wu
University of Washington
ccwu@cs.washington.edu

Jan-Michael Frahm
UNC-Chapel Hill
jmf@cs.unc.edu

Marc Pollefeys
ETH-Zürich
marc.pollefeys@inf.ethz.ch

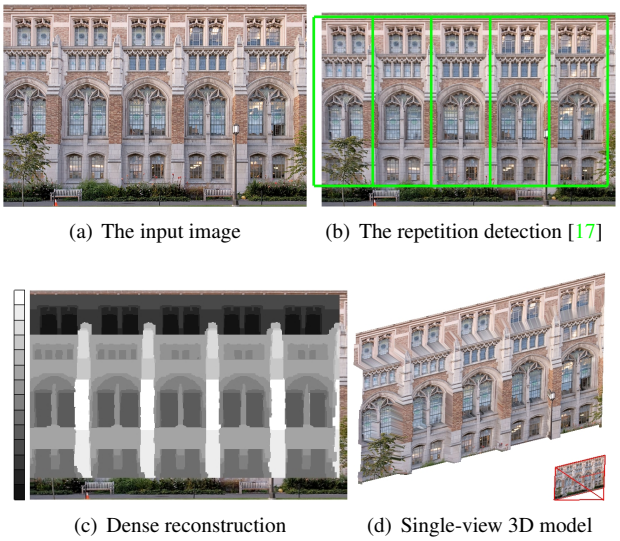
Abstract

This paper presents a novel approach for dense reconstruction from a single-view of a repetitive scene structure. Given an image and its detected repetition regions, we model the shape recovery as the dense pixel correspondences within a single image. The correspondences are represented by an interval map that tells the distance of each pixel to its matched pixels within the single image. In order to obtain dense repetitive structures, we develop a new repetition constraint that penalizes the inconsistency between the repetition intervals of the dynamically corresponding pixel pairs. We deploy a graph-cut to balance between the high-level constraint of geometric repetition and the low-level constraints of photometric consistency and spatial smoothness. We demonstrate the accurate reconstruction of dense 3D repetitive structures through a variety of experiments, which prove the robustness of our approach to outliers such as structure variations, illumination changes, and occlusions.

1. Introduction

The existence of repetitive and symmetric structures is a pervasive phenomenon in urban scenes. In typical images, the perspective distorted repetition and symmetry encode the relative 3D geometry between the repeating elements. If the repetition is mostly on a plane, the perspective distortion can be modeled by a planar homography. Detecting such planar repetition and symmetry allows us to recover *vanishing points* and camera calibrations [13, 7]. While non-planar repeating structures can not be accurately modeled by one homography, this paper exploits the visual differences between the repeating elements to recover the 3D details. We propose to obtain the 3D repetition information by modeling it as energy minimization yielding a dense 3D reconstruction of the repetitive structures.

The main contribution of this paper is a novel model to use high-level geometric information, such as repetition and reflective symmetry, in an optimization framework, which allows to enforce geometric consistency between repetitive pixels that are not immediate image neighbors. By enforcing consistency between the 3D reconstruction of the repeating elements, more accurate reconstructions even filling in occluded structures can be achieved (e.g. the tree in Fig. 1). One application of our method is stereo reconstruction of urban scenes with repetitive 3D structures. Additionally, we demonstrate the extension of the proposed concept to multi-view reconstruction.



(a) The input image (b) The repetition detection [17]

(c) Dense reconstruction (d) Single-view 3D model

Figure 1. An example of our reconstruction. The repeating elements in (b) show differences in due to their depth. Particularly, the distance between two columns is larger than the distance between the upper windows given that the columns are closer. Our approach recovers consistently repeating structures despite the varying reflections and occlusions.

The remainder of the paper is organized as follows. Section 2 briefly discusses the related work. In Section 3 we discuss the 3D geometric constraints deployed for reconstruction followed by the repetition detection in Section 4 and the optimization framework for dense reconstruction in Section 5. Experiments and applications are shown in Section 6.

2. Related Work

Dense reconstruction of repetitive structures strongly connects the work on repetition analysis and on stereo reconstruction. Due to the importance of repetition and symmetry in man-made scenes, many methods have been proposed to detect repetitive and symmetric structures especially from facade images (e.g. [10, 17]).

Reconstruction from Repetition and Symmetry Sparse reconstruction based on repetition and symmetry has been well studied. Hong et al. [7] explores general symmetry (including translational, reflective, and rotational symmetry) to recover camera pose and the orientation of scene objects, as well as some sparse geometry, from a single view. Francois et al. [5] use a virtually mirrored viewpoint to frame the single view reconstruction as a two view reconstruction.

Based on the sparse reconstruction, several work has furthur achieved dense reconstruction. Gool et al. [6] propose an optimization framework that uses the sprase feature matches as control points to recover dense depth maps. Shimshoni et al. [14] recover symmetric human face models by propagating the correspondences of manually-given pixels pairs based on photometric stereo. Our work proposes a global energy minimization framework to model both repetition and reflective symmetry in the image domain with a simple interval map model.

Markov Random Field Stereo MRF-based stereo optimization typically uses a data term to enforce photometric consistency between matched pixels, and a smoothness term to penalize the inconsistency of disparities between pixel neighbors [3, 11, 16]. Most stereo algorithms enforce the consistency only in the traditional pixel neighborhood, while consistency between non-neighboring pixels is often considered intractable. We propose a novel repetition and symmetry-based energy function that enforces high-level consistency between the disparities of non-neighboring pixels. Furthermore, we show that high-level 3D information can be modeled in the image domain by graphcut.

Symmetric Stereo Symmetric stereo methods treat all the images equally. Particularly, to recover multiple depth maps that are consistent with each other, the interactions between the depth maps need to be modeled. However, the interactions between pixels in different images are depth dependent posing a challenge to an image based model. Several methods have been proposed to enforce the consistency indirectly through visibility and occlusions. For example, [9] define an interaction set among multiple images and enforce the hard visibility constraint, and [15] uses an occlusion term to penalize the occlusion, which indirectly makes depth maps consistent. In this paper, the proposed energy function will directly enforce the consistency between multiple depth maps (and different parts within the depth map).

3. The 3D Geometry of Rectified Images

We consider the dominant horizontal repetition and symmetry often found in man-made environments, which can be robustly detected. Without loss of generality, we assume the following properties for one set of 3D repeating structures that have equal spacing:

- The camera center is at $(0,\ 0,\ 0)$;
- The 3D repetition step is 1 along direction $(1,\ 0,\ 0)^T$.

The projection matrix of the original image can be written as $P = KR[I\,|\,0]$, where K is the intrinsic calibration and R is the camera orientation.

Consider a rectified image generated according to two orthogonal vanishing points in the original image. The homography H for rectification must satisfy $HKR\,(1,\ 0,\ 0)^T \sim (1,\ 0,\ 0)^T$ and $HKR\,(0,\ 1,\ 0)^T \sim (0,\ 1,\ 0)^T$. The matrix H can be written as

$$\begin{bmatrix} a & 0 & b \\ 0 & c & d \\ 0 & 0 & 1 \end{bmatrix} (KR)^{-1},$$

where a, b, c, d are decided by the choice of rectification and the calibration K. Given a 3D point $(X,\ Y,\ Z)^T$, the corresponding pixel $(x, y)^T$ in the rectified image is

$$(x, y)^T = (\frac{aX}{Z} + b,\ \frac{cY}{Z} + d)^T. \tag{1}$$

Given a set of repeating 3D points $(X + k,\ Y,\ Z)^T$ with $k = 0, \ldots, N$ and N the number of repetitions, their projections lie on the same scanline $y = cY/Z + d$ in the rectified image, and the distance between any two neighbouring projections is a/Z. Therefore, the repetition interval is actually a function of depth Z, which we denote as

$$I_Z = \frac{a}{Z}. \tag{2}$$

This implies for the reconstruction: 1) the pixel correspondences need to be considered only within scanlines, 2) the relative depth of the 3D point can be recovered from the image repetition interval if the camera calibration is known. When the camera calibration K is available, the camera pose can be solved. Hence, a, b, c, and d can be obtained from H, and the 3D location of each pixel (p_x, p_y) in the rectified image can be then recovered as

$$(X_p,\ Y_p,\ Z_p)^T = (\frac{p_x - b}{I_Z},\ \frac{a(p_y - d)}{cI_Z},\ \frac{a}{I_Z})^T. \tag{3}$$

For uncalibrated cameras, we assume the principle point of the original image at the image center, and recover the focal length and the camera pose by enforcing the orthogonality of vanishing directions, which is similar to [13]. In case of degeneracy where the vanishing points are almost at infinity, we choose the focal lengths based on the EXIF header of the JPEG if its available or as a typical viewing angle.

In addition to pure repetition, it is possible to model local reflective symmetries for the repetitive structures. Let the range along $(1,\ 0,\ 0)^T$ of the first element be $[X_0 - 0.5,\ X_0 + 0.5]$, the symmetry planes in 3D are $X_i = X_0 + \frac{i}{2}, i \in \mathbb{N}$, and their corresponding positions in the rectified image are

$$x = \frac{a(X_0 + \frac{i}{2})}{Z} + b = (X_0 + \frac{i}{2})I_Z + b. \quad (4)$$

Hence, the symmetry axes at different depths are transformed to different locations in the rectified image. Besides the depth, symmetry axes is constrained by the global unknown X_0, which we can automatically recover in Section 5. Note that $x = b$ is the vanishing line of any planes that is perpendicular to $(1, 0, 0)$.

4. Repetition Detection

As the input to our algorithm, we deploy the repetition detection of [17] to obtain rectified images and their repetition regions. In addition, the detection provides the vanishing points of the original image (from which we recover K and R) and the homography H used by the rectification (from which we recover a, b, c and d). The detection [17] is particularly robust to the appearance differences between the repeating elements, and we exploit these differences for the dense reconstruction.

Each detected repetition region P fits the geometric model discussed in Section 3, which allows the repetition-based dense reconstruction. For each region P, we choose a repetition interval range according to the feature matches along the scanlines in the region, where the feature matches are byproduct of the repetition detection. Let D_X be the horizontal distances of the matched feature pairs. We empirically choose the repetition interval range as $L = \{l \mid l \in \mathbb{N},\ |l - \text{mean}(D_X)| < 2\sqrt{var(D_X)}\}$. Experiments show that such a range is typically larger than the actual interval range, but we can search for a more accurate range from a first-pass reconstruction with L.

5. Optimization Framework

Instead of direct recovery of depth, we densely estimate the repetition intervals for pixels of a repetition region in the rectified image, which are inversely proportional to the depths. Similar to the disparity map in two-view stereo, we call this **interval map**. An interval map is basically a labeling f over a repetition region P, such that each pixel p matches either $p - f(p)$ or $p + f(p)$[1] [2]. As a key property of repetitive structures, the interval map should satisfy:

$$f(p - f(p)) \text{ and } f(p + f(p)) \text{ are similar to } f(p).$$

[1]To simplify the notations, we use $p \pm I$ to denote $(p_x \pm I, p_y)$.

[2]Throughout the paper, we do not use any pixels outside the region P.

Now we propose a novel energy function to jointly model photometric appearance similarity, neighborhood smoothness and repetition consistency:

$$E(f) = E_{data}(f) + E_{smooth}(f) + E_{repetition}(f), \quad (5)$$

where the data term E_{data} is the matching cost of repeating pixels, the smoothness term E_{smooth} penalizes the differences between the repetition intervals of neighboring pixels, and the repetition term $E_{repetition}$ penalizes the differences between the repetition intervals of corresponding pixels.

Data term since repetition is bidirectional, the data cost should combine the cost from the left matching and right matching. Let $D(p, q)$ be the matching cost of two pixels p and q, we define the data cost for a pixel p as

$$D_f(p) = \begin{cases} \text{mean } \{\ D(p,q) \,|\, q = p \pm f(p),\ q \in P\ \} \\ \qquad \text{if } p + L_{\max} \in P \text{ or } p - L_{\max} \in P \\ 0 \qquad \text{otherwise.} \end{cases}$$

The two cases of $D_f(p)$ ensure that a pixel either has valid costs for all labels or 0 for all labels, which handles the margin when the repetition count is two. The data cost for the entire image is defined as

$$E_{data} = \sum_{p \in P} D_f(p). \quad (6)$$

Given two pixels p and q, we design their matching cost based on the maximum absolute difference from the three color channels, which we denote as $D_O(p, q)$. We employ two standard techniques to improve the matching cost: We first apply the Birchfield-Tomasi (BT) sampling [1] to reduce errors caused by the resampling during the image rectification, which is denoted by $D_{BT}(p, q)$. Second, we truncate with T_D the matching cost to improve robustness to occlusion leading to the pairwise cost $D(p, q) = \min(D_{BT}(p, q), T_D)$.

A significant difference between the repetition-based reconstruction and multi-view stereo is that we match the appearances of multiple surfaces within one image, while multi-view stereo matches the appearance of the same surface across multiple images. Hence we aim for more robustness as real scenes often do not have perfect repetition. The cost truncation supports the robustness of large appearance differences of the repeating elements.

Smoothness term we define the smoothness cost based on the truncated L_1 distance of the repetition intervals of the neighboring pixels. Let N_P be the set of neighboring pixels in the repetition region P for the 4-neighborhood system, our smoothness term is defined as

$$E_{smooth} = \omega_{smooth} \sum_{(p,q) \in N_P} V(p, q), \quad (7)$$

where ω_{smooth} is a positive penalty for violating the smoothness constraint. To boost robustness we again chose a truncated cost $V = \min(T_V, |f(p) - f(q)|)$ with T_V being the truncation threshold. The cost V is small for small interval differences, and avoids overly punishing large interval changes at object boundaries.

Repetition term We design a novel repetition term to penalize the deviation between different instances $f(p \pm f(p))$ of the same repetition $f(p)$. Treating the different parts of the interval map as the disparity map of the repeating elements, the repetition term essentially provides a new formulation for symmetric stereo by explicitly enforcing that matching pixels have similar disparities.

For convenience of notation, we define a function

$$\rho(\textbf{condition}) = \textbf{condition} \text{ is true } ? \ 1 \ : \ 0.$$

The repetition term between two pixels only needs to be enforced when their distance is equal to one of their repetition intervals. The key difference between the repetition term and the smoothness term is that the pixel pairs are relying on the labels of the pixels, which defines a dynamic neighborhood for each pixel. We then define a function R_f to check if two pixels p and q should be compared.

$$R_f(p, q) = \rho(|p_x - q_x| \ \in \ \{f(p), \ f(q)\}). \qquad (8)$$

We also consider other requirements for enforcing the repetition. As discussed in the data term, the matching of repeating pixels often involves large appearance changes from occlusions, reflections, noise etc. It is not optimal to enforce the repetition consistency without considering the photometric similarity, as the global optimization may try to avoid those matching cost and propagate incorrect intervals. Hence it is natural to loosen the repetition constraint when the two locations significantly deviate in appearance. This is similar in spirit to the idea reducing the smoothness constraint across image edges. We define a guiding function $G(p, q)$ to evaluate if the repetition-based consistency should be considered for two pixels.

$$G_{local}(p, q) = \rho(D_{BT}(p, \ q) \ < \ T_G), \qquad (9)$$

where T_G is a threshold for testing the pixel similarity. Consequently, the repetition cost is applied only between similar pixels, and larger T_G gives stronger constraint.

In Section 6, we compare G_{local} to two other choices of the guiding function: 1) No repetition constraint with $G_{none} = 0$; 2) The global repetition term G_{global} that enforces repetition based on the region decomposition of [17] without considering the photometric similarity. In the following G_{local} from Equation (9) is always used unless otherwise specified.

We define the repetition cost of the entire image as

$$E_{repetition} = \omega_{rep} \sum_{q_y = p_y, \, R_f(q,q)=1} G(p, q) \ \rho(f(p) \neq f(q)). \qquad (10)$$

where ω_{rep} is a positive penalty for violating the repetition consistency. Equation (10) uses a dynamic neighborhood for the pixel nodes in the graph, which can not directly be handled by traditional optimization methods. However, the above equation can be rewritten as

$$E_{repetition} = \omega_{rep} \sum_{|q_x - p_x| \, \in \, L, \, q_y = p_y} R_f(p, q) \, G(p, q) \, \rho(f(p) \neq f(q)). \qquad (11)$$

Now the neighborhood is fixed and the standard energy optimization methods are applicable. Since the number of edges for the repetition term is in $O(|P||L|)$, one limitation of this work is the possibly large memory consumption.

In this work, the efficient α-expansion graph-cut [3, 2] is used to minimize the proposed energy. Kolmogorov and Zabih [8] proved that α-expansion can minimize the class of energy functions that satisfy the regularity constraint $e(\alpha, \alpha) + e(\beta, \gamma) \ \leq \ e(\beta, \alpha) + e(\alpha, \gamma)$. We now prove that our repetition term fulfills the regularity constraint.

Given an edge between two pixels p and q, let $\omega = \omega_{rep} \, G(p, q)$ and $\delta = |p_x - q_x|$, the repetition cost $e(f(p), f(q))$ of the edge is a function of their labels

$$\begin{aligned} e(f(p), \ f(q)) &= \omega \ R_f(p, \ q) \ \rho(f(p) \neq f(q)) \\ &= \omega \rho(f(p) = \delta \text{ or } f(q) = \delta) \ \rho(f(p) \neq f(q)), \end{aligned}$$

which is obviously non-negative and symmetric.

Now, we consider three labels α, β and γ to prove the regularity property. If $\alpha = \beta$, $\alpha = \gamma$ or $\beta = \gamma$ the inequality can be simply proved by substitution, which this paper will skip. If the three labels are all **different**, at most one of them can be equal to δ, and there are four different cases:

$$\begin{array}{llll} \text{IF } \alpha = \delta: & e(\beta, \gamma) = 0, & e(\beta, \alpha) = \omega, & e(\alpha, \gamma) = \omega; \\ \text{IF } \beta = \delta: & e(\beta, \gamma) = \omega, & e(\beta, \alpha) = \omega, & e(\alpha, \gamma) = 0; \\ \text{IF } \gamma = \delta: & e(\beta, \gamma) = \omega, & e(\beta, \alpha) = 0, & e(\alpha, \gamma) = \omega; \\ \text{Otherwise}: & e(\beta, \gamma) = 0, & e(\beta, \alpha) = 0, & e(\alpha, \gamma) = 0. \end{array}$$

Since $e(\alpha, \alpha) \ \equiv \ 0$, the inequality below holds true:

$$e(\alpha, \alpha) + e(\beta, \gamma) = e(\beta, \gamma) \ \leq \ e(\beta, \alpha) + e(\alpha, \gamma).$$

It is also worth noting that our choice of repetition cost is non-trivial, we find that other choices such as having $e(f(p), f(q))$ proportional to (truncated) $|f(p) - f(q)|$ or penalizing only the non-occluded pixels will violate the regularity constraint of α-expansion.

Reflective symmetry term Similar to the repetition term, any available knowledge about the reflective symmetries

can be incorporated. To model them it is required to know the symmetry axis X_0 in Equation (4). The symmetry axis is initially unknown but can be recovered from the sparse feature correspondences or alternatively from the dense correspondences. For improved robustness we opt for the latter and propose a two-step approach:

1. Recover the interval map f_0 using only $E(f)$ from Equation (5), and locate X_0,
2. Refine the interval map by a graphcut that includes the reflective symmetry term:

$$E^+(f) = E(f) + E_{sym}(f). \tag{12}$$

To extract X_0 we render an orthogonal view f_0' of the interval map f_0 by generating the 3D model from f_0 by using Equation 3 and reprojecting the 3D model along $(0,\ 0,\ 1)^T$. For robustness 3D points that have $f_0(p + f_0(p)) \neq f_0(p)$ or $f_0(p - f_0(p)) \neq f_0(p)$ are excluded. Next X_0 is chosen from all possible locations in the repeating elements to maximize the number of consistent pixels pairs in f_0'.

Next we introduce our reflective symmetry term E_{sym}. Given two pixels (p, q) that have different labels, we would like to give a penalty if they are at symmetric positions w.r.t. the depth of either $f(p)$ or $f(q)$. First, Equation 4 gives the set of symmetry axes for an interval I, thus we can check if the two pixels are symmetric w.r.t to one of these symmetry axes. Second, to enforce reflective symmetry only within each element, we require $|p_x - q_x| < L_{\min}$. We denote the function that tests the two aforementioned conditions as $C(I, p, q)$. Given an interval map f, we define the indicator function $S(f, p, q)$ for a pixel pair (p, q) as

$$S(f, p, q) = \max(C(f(p), p, q), C(f(q), p, q)). \tag{13}$$

Enforcing reflective symmetry is harder than enforcing repetition due to the occlusions we often have from oblique viewpoints. For example, in Figure 2.3, the right halves of many repeating elements are severely occluded, and it would create new problems if enforcing reflective symmetry naively everywhere. In particular, enforcing reflective on pixels whose symmetric structures are occluded would require to perturb the occluding pixels. On the contrary, only a few pixels are occluded in terms of pure repetition. It is less reliable to recover depth from reflective symmetry than from repetition, meaning the reflective symmetry should be a waker constraint than the repetition.

In this paper, we realize the reflective symmetry term as a refinement such that we do not contaminate the pixels $\Lambda(f_0)$ that already satisfy the reflective symmetry in f_0:

$$\Lambda(f_0) = \{p \,|\, \exists_q (q_y = p_y, f_0(p) = f_0(q), S(f_0, p, q) = 1)\}.$$

The graph edges $\Phi(f_0)$ for enforcing reflective symmetry are chosen to include all possible symmetric pairs but to not have any pixels in $\Lambda(f_0)$:

$$\Phi(f_0) = \{(p, q)\,|p_y = q_y,\ p, q \notin \Lambda, \exists_{l \in L}(C(l, p, q) = 1)\},$$

The symmetry term for the entire region is then given by

$$E_{sym}(f) = \omega_{sym} \sum_{(p,\ q) \in\ \Phi(f_0)} S(f, p,\ q)\ \rho(f(p) \neq f(q)),$$

where ω_{sym} is the penalty for violating reflective symmetry.

The proposed symmetry term also satisfies the regularity constraint. Consider the cost $e(f(p), f(q))$ of an edge (p, q), if $e(\beta, \gamma) = \omega_{sym}$, one of $C(\beta, p, q)$ and $C(\gamma, p, q)$ must be 1. As a result, either $e(\beta, \alpha)$ or $e(\alpha, \gamma)$ will be equal to ω_{sym}, and $e(\beta, \gamma) \leq e(\beta, \alpha) + e(\alpha, \gamma)$ is satisfied. Although it seems intuitive to incorporate occlusion information, we unfortunately find such reflective symmetry terms violating the regularity constraint.

6. Experiments

This section demonstrates the experimental results of the proposed repetition-based single view reconstruction to show the advantages of the novel optimization framework.

6.1. Repetition-based Single-view Reconstruction

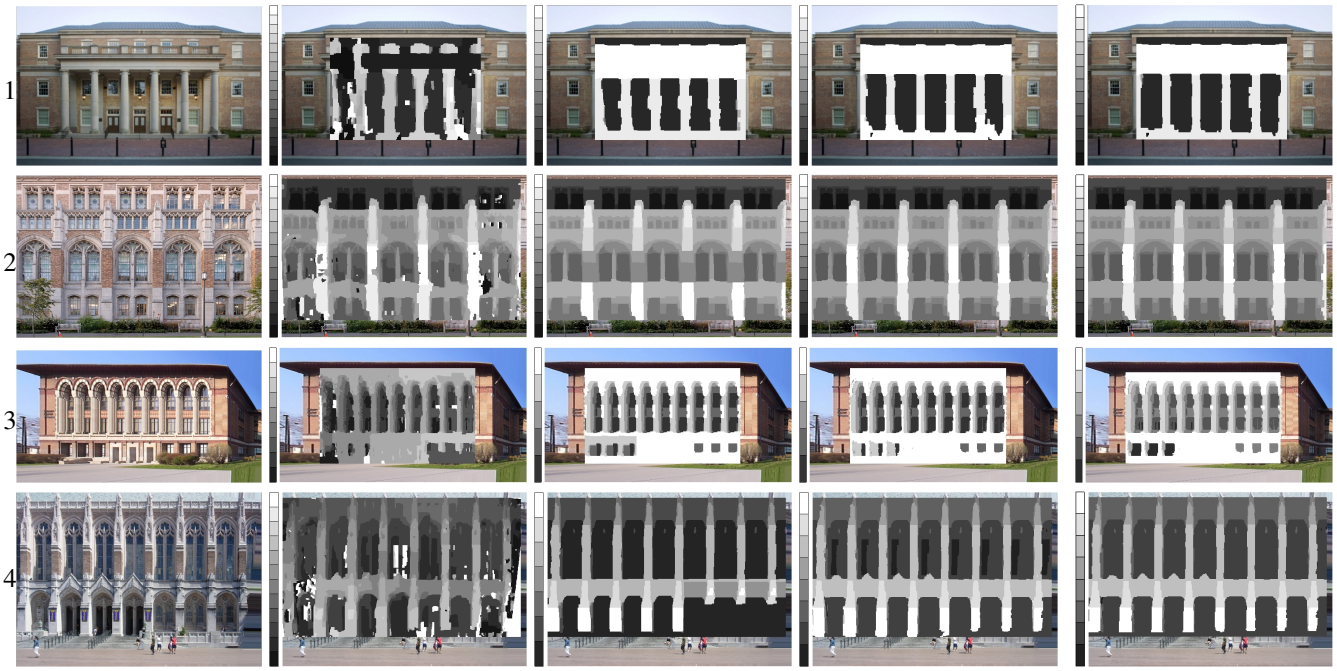
We first present our single-view reconstruction on a variety of challenging urban scenes. While the pervasive existence of the present repetition allows us to recover the dense 3D geometry from single images. The pipeline we deploy for urban scene reconstruction is:

1. α-expansion graphcut to minimize $E(f)$
2. Find the refined interval range L' from recovered f
3. α-expansion graphcut to minimize $E(f)$ on L'
4. Extract 3D symmetry axis parameter X_0 using f
5. α-expansion graphcut to minimize $E^+(f)$ on L'

The refined interval range is extracted by excluding the labels that are assigned to very few pixels. Let $r(l) = |\{p|f(p) = l\}|/|P|$, the new range is chosen to be $L' = [\min\{l|r(l) \geq r_{\min}\},\ \max\{l|r(l) \geq r_{\min}\}]$, where $r_{\min} = 1\%$ is used. This filtering won't affect most of the pixels, but it does improve the robustness by neglecting rarely used labels. We will show the effect of the label filtering in the supplemental material. In all our experiments we use $T_D = T_G = 25$, $T_V = 2$, $\omega_{smooth} = \omega_{rep} = 10$, and $\omega_{sym} = 2$ unless explicitly stated.

Figure 2 shows the different results on four challenging images when using different repetition constraints. We first compare the interval maps recovered with three different repetition terms (after step 1-3).

- **No repetition constraint** $G(p, q) = 0$. The standard graphcut is able to recover some correct intervals for the repetition regions, but the results show its sensitivity to noise and outliers in real scenes.



(a) Input rectified image (b) Reconstruction with no repetition term (c) Reconstruction with global repetition term (d) Reconstruction with local repetition term (e) Refined reconstruction w. reflective symmetry

Figure 2. Comparison of different repetition terms. The brighter colors in the interval maps correspond to larger intervals and closer surface, and the gray scale bars on the left give the number of filtered interval labels. The comparison shows the local repetition term reconstructs correct interval maps despite the variations between the repeating elements, while the global repetition term tends to over-smooth and propagate errors. The last column show2 the improvement after enforcing reflective symmetry.

- **Global repetition term** $G_{global}(p,q)$ enforces repetition everywhere within each repetition region group. Despite the most continuous repetition results, this constraint is error-prone by propagating local errors (e.g. matching with occlusion boundaries). The quality is limited by the accuracy of the initial region segmentation, for example, the triangle structures in the 4th row in Figure 2 are smoothed away.
- **Local repetition term** $G_{local}(p,q)$ enforces repetition only for similar pixels. With the local repetition term, it is possible to reconstruct repeating structures even under large occlusions. It specifically handles the repetition of different structures (e.g. Figure. 2.3 and 2.4).

Table 1 lists the computation time for the experiment shown in Figure 2. It can be seen that enforcing the local repetition term takes **3** times the time of the standard optimization. For the local repetition term, no edges will be constructed in the graph for the pairs that satisfy $G(p,q) = 0$, and the total number of edges for repetition is $|\{(p,q)|G(p,q) \neq 0, |q_x - p_x| \in L, q_y = p_y\}|$. Consequently, the optimization with the local repetition term runs faster than that with the global constraint term.

The refinement of interval maps by reflective symmetry is demonstrated in Figure 2.(e). The repeating structures in this experiment are all reflective symmetric, but the reconstruction with neither local nor global repetition term can produce symmetric interval maps. By including the reflective symmetry term into the optimization, the inconsistencies between reflectively symmetric regions are corrected.

	CPU: 3Ghz P4	Time (seconds)/ $\|L'\|$			refine
	$Dim(P)\times\|L\|$	none	global	local	time
1	443x298x17	11/17	38/13	32/13	+14
2	781x461x17	22/14	79/13	67/14	+44
3	865x504x10	17/9	47/5	35/5	+14
4	996x582x13	29/12	91/8	79/8	+39

Table 1. Reconstruction timing (step 1-3) and label filtering under different repetition terms. The last column gives the extra time on enforcing reflective symmetry (step 4-5).

We reconstruct 3D structures by incorporating the recovered interval map and calibration according to Equation 3 and present the high quality results in Figure 1(b), Figure 3 and Figure 5.(a) . The camera positions and calibrations shown in Figure 3 are recovered based on vanishing points, while for Figure 1(b) and Figure 5(a) they are selected according to the EXIF data of the images because their recovered vanishing point locations are close to infinity. Hence, after scaling the 3D structure along the Z direction as well as the focal length, the resulting image does not

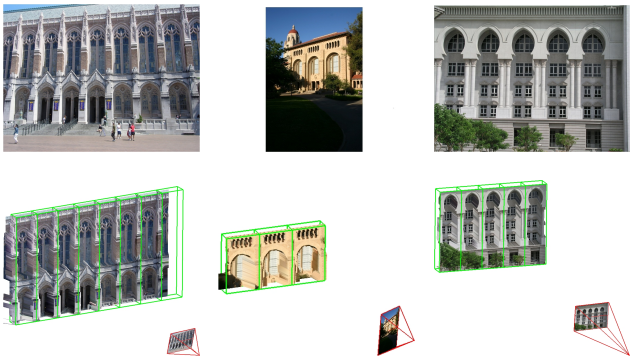


Figure 3. Dense reconstruction results. The first row gives the original images, and the second row shows the recovered 3D models. Each 3D bounding box also indicates a symmetric repeating item in 3D. More details can be found the supplemental video.

change. Such an ambiguity can not be solved without additional information (e.g. the third vanishing point).

Experiments show that our repetition term is robust to small errors in repetition detection and small radial distortions. It is mainly because we do not explicitly enforce the consistency between non-neighboring repeating elements thus preventing disturbance from larger distortions. In addition, the Birchfield-Tomasi sampling compensates partially for the errors. The data term and smoothness term. The typical impact would be a small shift for the location of the occlusion boundaries. For example, Figure 2.1 is an image that contains obvious radial distortion.

6.2. Repetition-based Symmetric Stereo

The proposed repetition-based reconstruction can also be applied to the standard two-view stereo and multi-view stereo of cameras that are equally spaced. For two-view stereo, the proposed method can be viewed as a symmetric stereo that explicitly enforces the consistency between multiple disparity maps.

Given a stero pair (left image $\mathcal{L}$ and right image $\mathcal{R}$), the input to our algorithm is their combined image $P = [\mathcal{R}, \mathcal{L}]$. We use such combined images to evaluate the proposed algorithm on the Middlebury datasets [11, 12], where the interval range is known from the ground truth disparity range. We slightly modified our pipeline to disable the steps that are specific to urban scenes by skipping the label filtering and reflective symmetry refinement.

Compared with the urban scene reconstruction experiments, we selected looser parameters to account for the higher image quality. In addition, we also make the smoothness cost adaptive to image edges, which are more reliable than those in the highly-textured and noisy urban images. The adaptive smoothness is chosen as:

$$V'(p,q) = \begin{cases} 0.5\,\rho(f(p) \neq f(q)) & D_O(p,q) > T_E \\ \min(T_V, |f(p) - f(q)|) & \text{otherwise} \end{cases}$$

T_E, ω_{smooth} T_G, ω_{rep}	Avg. Rank	Tsukuba all	Venus all	Teddy all	Cones all	Bad Pixels
∞,2, 10, 0	66.9	3.74	1.65	24.6	23.7	13.4
∞,2, 10, 1	61.0	3.67	1.65	19.7	14.5	10.4
∞,2, 10, 2	59.9	3.60	1.47	19.2	14.1	10.1
∞,2, 10, 4	59.9	3.46	1.20	17.9	12.9	9.79
∞,2, 20, 4	59.9	3.47	1.27	18.6	12.5	9.96
5, 4, 10, 0	68.2	3.53	2.10	24.6	31.7	15.1
5 ,4, 10, 1	58.2	3.06	1.50	19.0	18.1	10.3
5, 4, 10, 2	53.2	2.78	1.28	17.6	15.0	9.22
5 ,4, 10, 4	44.7	1.93	0.83	16.0	13.9	8.33
5, 4, 20, 4	43.3	2.02	0.80	15.8	12.8	7.99

Table 2. Evaluation of the repetition-based symmetric stereo. We choose two sets of smoothness setting: $\{T_E = \infty, \omega_{smooth} = 2\}$ and $\{T_E = 5, \omega_{smooth} = 4\}$, and we gradually increase the strength of the repetition term by increasing ω_{rep} and T_G, which demonstrates the increasing improvements of reconstruction quality. Extended version of this table can be found in the supplemental material.

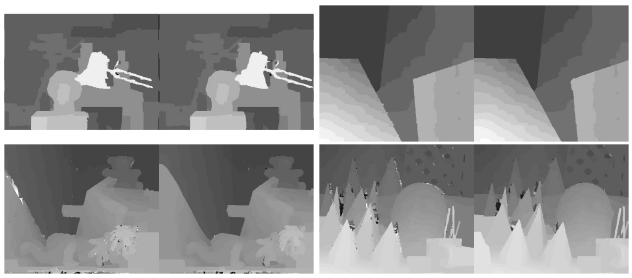


Figure 4. Interval maps recovered from the Middlebury dataset [11, 12]. Two smooth disparity maps are generated simultaneously, and the results are particular accurate for the non-occluded areas.

where T_E is the edge detection threshold. Consequently, smaller penalties are given to image edges. Note that the edge adaptiveness can be disabled by setting $T_E = \infty$.

Table 2 shows the evaluation of the recovered interval map under different settings (We fix the remaining parameters: $T_D = 5, T_V = 2$, and $\omega_{sym} = 0$). The evaluation proves that the additional repetition term in fact improves the reconstruction by explicitly enforcing the consistency between the two disparity maps. One example set of the reconstruction is given in Figure 4. The repetition-based symmetric stereo can be furthur improved by combining with other existing techniques, such as segmentation-based smoothness constraint.

6.3. Ortho-rectified Images

One of the interesting application of our single view reconstruction is to generate ortho-rectified views, an invariant view of 3D structures. Based on the texture synthesis proposed by [4], we generate the ortho-rectified images by

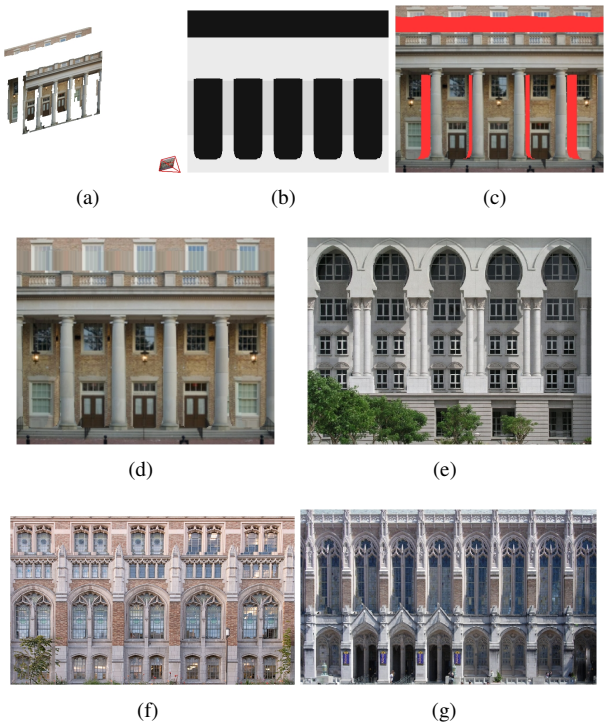


Figure 5. Examples of generating ortho-rectified views. (a) Dense surface model. (b) Fused orthogonal interval map. (c) Frontal view with missing pixel marked red. (d-g) Results of synthesized orthogonal views that accurately align the symmetry axes at different depths. Indeed, there are a few artifacts in (g), for example, the wrong door is copied to the last element, the error around the depth discontinuity above the center door. We suggest zooming to see the details.

filling the missing pixels from the best copy out of their repeating counterparts when they are visible in other instances of the repeating element. To reduce the noise in the reconstruction, we first apply a simple fusion of the orthogonal interval map to refine the reconstruction by enforcing smoothness, repetition and reflective symmetry in the 2.5D space of the interval map. Figure 5 shows examples of the fused orthogonal interval map and the final synthesized image. In Figure 5, the occluded parts of the white windows in the first copy and the last copy are filled in smoothly.

Due to the difference between the original viewpoints and the orthogonal projections, there are pixels in the ortho-rectified image for which all the copies are invisible in the original image (compare Figure 5 and Figure 2.1). Generating truly realistic ortho-rectified images would require more complicated model, which is beyond the scope of this paper.

7. Conclusion and Future Work

We proposed a novel framework for dense single view reconstruction enforcing the high-level constraints provided by repetition and symmetry along with the photometric consistency and neighbor smoothness in a single unified model. We demonstrate the power of our energy minimization framework for single view dense reconstruction by accurately extracting repetitive structures.

In the future, we would like to apply the proposed optimization framework to multi-view reconstruction and improve the quality of the ortho-rectified facade images. Given that our intervals are corresponding to planes at different depths, it is straightforward to extend our work to plane sweeping multi-view stereo.

References

[1] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *TPAMI*, 1998. 3

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMPI*, 26:359–374, 2004. 4

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *TPAMI*, 2001. 2, 4

[4] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *ICCV99*, pages 1033–1038, 1999. 7

[5] A. Francois, G. Medioni, and R. Waupotitsch. Mirror symmetry $\Rightarrow$ 2-view stereo geometry. *IVC*, 21:137–143, 2003. 2

[6] L. J. V. Gool, G. Zeng, F. V. den Borre, and P. Müller. Towards mass-produced building models. In *Photogrammetric Image Analysis (PIA07)*, pages 209–220, 2007. 2

[7] W. Hong, A. Yang, K. Huang, and Y. Ma. On symmetry and multiple-view geometry: Structure, pose, and calibration from a single image. *IJCV*, 60(3):241–265, 2004. 1, 2

[8] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *TPAMI*, 26:65–81, 2004. 4

[9] V. Kolmogorov, R. Zabih, and S. Gortler. Generalized multi-camera scene reconstruction using graph cuts. In *Proc.of the International Workshop on Energy Minimization Methods in CVPR*, pages 501–516, 2003. 2

[10] M. Park, K. Brocklehurst, R. Collins, and Y. Liu. Deformed lattice detection in real-world images using mean-shift belief propagation. *TPAMI*, 31(10):1804–1816, 2009. 2

[11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2001. 2, 7

[12] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR03*, 2003. 7

[13] G. Schindler, P. Krishnamurthy, R. Lublinerman, Y. Liu, and F. Dellaert. Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In *CVPR08*, 2008. 1, 2

[14] I. Shimshoni, Y. Moses, and M. Lindenbaum. Shape reconstruction of 3d bilaterally symmetric surfaces. *IJCV*, 2000. 2

[15] J. Sun, Y. Li, S. Bing, and K. H. yeung Shum. Symmetric stereo matching for occlusion handling. In *CVPR05*, pages 399–406, 2005. 2

[16] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *ICCV03*, pages 900–907, 2003. 2

[17] C. Wu, J.-M. Frahm, and M. Pollefeys. Detecting large repetitive structures with salient boundaries. In *ECCV10*, volume 2, pages 142–155, 2010. 1, 2, 3, 4