Detecting Large Repetitive Structures with Salient Boundaries

Changchang Wu[1], Jan-Michael Frahm[1], and Marc Pollefeys[2]

[1] Department of Computer Science
UNC Chapel Hill, NC, USA
{ccwu,jmf}@cs.unc.edu
[2] Department of Computer Science
ETH Zürich, Switzerland
marc.pollefeys@inf.ethz.ch

Abstract. This paper presents a novel robust and efficient framework to analyze large repetitive structures in urban scenes. A particular contribution of the proposed approach is that it finds the salient boundaries of the repeating elements even when the repetition exists along only one direction. A perspective image is rectified based on vanishing points computed jointly from edges and repeated features detected in the original image by maximizing its overall symmetry. Then a feature-based method is used to extract hypotheses of repetition and symmetry from the rectified image, and initial repetition regions are obtained from the supporting features of each repetition interval. To maximize the local symmetry of each element, their boundaries along the repetition direction are determined from the repetition of local symmetry axes. For any image patch, we define its repetition quality for each repetition interval conditionally with a suppression of integer multiples of repetition intervals. We determine the boundary along the non-repeating direction by finding strong decreases of the repetition quality. Experiments demonstrate the robustness and repeatability of our repetition detection.

1 Introduction

Repetition and symmetry are frequently used in the design of urban architecture. In fact, buildings often consist of a hierarchy of repetitions and symmetries (e.g. Fig. 1). Particularly, most of the basic repeating elements on facades (such as doors and windows) are symmetric by themselves, repetition and symmetry coexist and interplay at different scales. This paper introduces a new method to detect repeating elements with salient boundaries in facade images.

The symmetry and repetition patterns together with the appearance of the repeating/symmetric elements provide a strong characterization of the scene. Given that, particularly for urban scenes, the symmetries and repetitions of a scene describe its high-level structure, they can be used for wide baseline matching. One area where this representation would be useful is in the reconstruction from urban photo collections as in [1]. The reliable boundaries of the detected repeating elements and the symmetric structure can be used as compact image

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part II, LNCS 6312, pp. 142–155, 2010.

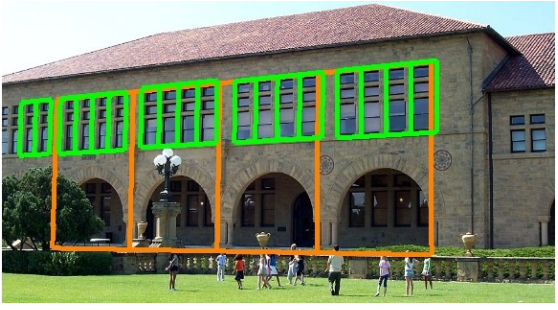

Fig. 1. Example of our detected repetitive structures. Note that the vertical boundaries are selected automatically to distinguish between the interesting elements and high frequency repetition of the roof.

features for effective recognition. Since such structures encode significantly more scene semantics than, for example, SIFT features [2], the matching is significantly less ambiguous. The known scene symmetries and repetitions allow us to automatically extract the facade grammars [3,4,5] as well as the semantic parsing of the images. Additionally, the known structure of the facades allows to regenerate facades based its grammar or compensate for occlusions by replacing occluded parts through their symmetric or repetitive equivalent.

Reliably detecting repetitions and extracting their boundaries is a significantly challenging problem. Even though images of planar facades can be rectified to a frontal view by using the vanishing points of the facade, the appearance of repeating elements may still significantly change, due to reflections and occlusions. In addition, the perspective change for non-planar structures on the facade plane severely affects the local symmetries.

A particularly challenging scenario that draws our attention is where the large repetitive structures repeat only along the horizontal direction (e.g. Fig. 1). Homogeneous regions, edges along vanishing directions, and high-frequency repetitions cause additional ambiguities in choosing meaningful boundaries for the repeating elements. To reliably detect the boundaries of such structures, we need to distinguish between regions that belong to different repetition groups (with different repetition intervals).

The remainder of the paper is organized as follows. Section 2 briefly discusses the related work. Section 3 discusses the few of our observations on repetition in urban scenes. Section 4 gives our vanishing point detection and sparse repetition analysis. Sections 5 introduces our repetition quality functions. Section 6 proposes our dense repetition detection algorithm with salient boundary detection. Experiments are discussed in Section 7 and conclusions are given in Section 8.

2 Related Work

Repetitions are usually hypothesized from the matching of local image features, and repetition are often detected as a set of sparse repeated features by growing

or tracking from the small sets of initial features towards their immediate spatial neighbors [6,7,8,9,10,11]. Dense detection of repetition requires the determination of the boundaries of repeating elements. Liu et. al [12] determine the boundary of repeating elements by maximizing the local symmetries. A limitation of their method is the requirement of a 2D repetition grid, which are not always available in urban environment. This paper shares their idea of maximizing local symmetries, but beyond that we separate different repetition groups by evaluating the local repetition quality conditionally for different repetition intervals.

Additional assumptions about the shape of the repeating elements are sometimes used to define the boundaries of repeating elements. Korah et. al [13] assume the repeating elements to be rectangular and extract them based on the edge segments in the rectified images. Their assumptions is often not completely valid in urban scenes because curved structures are very common. Our method uses a less restrictive assumptions only requiring the repeating elements to be approximately symmetric.

The general symmetry includes translational symmetry (we refer to as repetition), reflective symmetry and rotational symmetry. Many researches have proposed frameworks that can solve both translational symmetry and reflective symmetry(e.g. [14], [9]). Our method also handles both but in a joint fashion. We use the coexistence of repetition and symmetry to define the boundaries for our detected repetition regions along the repeating direction.

Perhaps most closely related to this paper is the work of Müller et. al [15]. They also aim to recover the architectural grammar describing the structure of the facade. The results are impressive, but require significantly stronger assumptions than for our approach. Besides rectification as in our approach, this approach requires a tight rectangular boundary delineating the facade (which seems to be a manual step as no automated solution is provided). It is further assumed that within this region vertical repetitions occur over the whole width and horizontal repetitions over the whole height. This is more restrictive than the bottom up approach we propose in this paper which only requires local support. As [15] only demonstrates their approach in the presence of both horizontal and vertical repetitions this seems to be required. Our approach works in the presence of horizontal repetition (or symmetry) alone. Finally, in [15] boundaries between elements are chosen based on edge support and distance heuristics and can yield undesired results. An important contribution of our work is to propose a principled approach to determine those boundaries based on the symmetry assumption and on direct image support. Beyond the scope of our paper, [15] refines the subdivision of facade elements and enables manual depth adjustments to yield detailed 3D facade reconstructions which is ideally suited for rendering.

3 Observations and Assumptions

Urban scenes are often designed with many repetitive structures, this section lists some observations that guided the design of our detection algorithm.

1. Dominant repetition(s) are mostly along the vanishing point direction(s) with equal 3D spacing. This gives us the opportunity to refine the vanishing point(s) based on repetition;
2. While many existing approaches require 2D repetition, many buildings lack vertical repetitions and symmetries. Our approach is specifically developed to handle this case.
3. Repeating architectural elements typically also exhibit reflective symmetry around vertical axes. Symmetry axes occur at twice the frequency of the repetition, in the middle and in between repeated elements. We use this to localize the vertical boundary between repeating elements (up to a two-fold ambiguity). Only in very few cases have we observed buildings where this principle is not satisfied. Note that the rectangle structure assumptions used for example by [13,15] is a special case of this assumption.

4 Sparse Repetition and Symmetry Detection

In this paper, we denote the extraction of repetition and symmetry from keypoints as sparse detection. This section first introduces our improved vanishing point detection, and then discusses our sparse detection in the rectified images.

4.1 Vanishing Point Refinement by Maximizing Overall Symmetry

Accurate vanishing point (VP) detection is important in our framework because we assume the repetitions along vanishing directions. Inaccuracy in VP locations will disturb the finding of optimal repetition interval and symmetry axes since the pairwise distances between the matched features change gradually. In our approach we use the cascaded hough transform [16] to compute the vertical and one or more horizontal vanishing points from edge pixels as initialization.

We propose a VP refinement by maximizing the overall symmetry in the entire image using both edges and features. Given a pair of horizontal and vertical vanishing points, VP_H and VP_V, a homography $T = T(VP_H, VP_V)$ can be determined to rectify the image. We define the transformation to keep the original resolution as much as possible to avoid too much shrinking and expanding. By matching SIFT [2] features extracted in the original image along both vanishing directions and keeping the closest matches (closest in the image), three sets of feature pairs can be extracted. We use R_H for horizontal repetition, R_S for horizontal symmetry, R_V for vertical repetition.

Consider a set of point pairs $R \in \{R_H, R_V, R_S\}$ in the original image and a transformation T, we use $X^T(R)$ to denote the distribution of their horizontal distances after rectification, $Y^T(R)$ for the distribution of their rectified vertical distances and $C^T(R)$ for the distribution of the horizontal coordinates of their rectified midpoints. Typically in urban scenes images, there exist only a few strong symmetry axes and repetitions intervals. Correspondingly, we expect to see only a few strong peaks in the histogram of $X^T(R_H)$, $Y^T(R_V)$ and $C^T(R_S)$. These strong peaks correspond to the minimum information that are required to represent most of the data distribution. Therefore, we expect low entropies

from those histograms. This paper optimizes the rectification by minimizing the summed entropy, so that the vanishing directions are better aligned with repetition directions and symmetry axes.

We use H to denote the entropy function. It can be proven that $H(X^T(R_H))$ and $H(Y^T(R_V))$ are invariant to any affine transformations, and $H(C^T(R_S))$ is invariant to transformation in the form of $\begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$. However, such an affine ambiguity can be resolved by using the point distances in the direction that is perpendicular to repetition or symmetry, $Y^T(R_H)$, $X^T(R_V)$ and $Y^T(R_S)$, because they are only invariant to transformations in the form of $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ given a finite resolution of histogram.

Consider a distance distribution $D(x) \in \{Y^T(R_H), X^T(R_V), Y^T(R_S)\}$ that are expected to be close to zero, we use the entropy of $D(x) + D(-x)$ in our minimization to reduce both drift from zero mean and large variance. We denote such entropy function by $\widehat{H}$. In our case, this can apply to $Y^T(R_H)$, $Y^T(R_S)$, and $X^T(R_V)$. Optionally, the edge information can be incorporated straightforwardly. Given the set of edge segments G_H and G_V that are corresponding to the two vanishing point, $Y^T(G_H)$ and $X^T(G_V)$ can be used the same as repetition.

By assuming the different distributions independent of each other and ignoring their joint distributions, we define an energy function for the repetition and symmetry of an image as

$$\begin{aligned} Q(VP_H, VP_V) = {} & H(X^T(R_H)) + H(Y^T(R_V)) + H(C^T(R_S)) + \widehat{H}(Y^T(R_H)) \\ & + \widehat{H}(Y^T(R_S)) + \widehat{H}(X^T(R_V)) + \widehat{H}(Y^T(G_H)) + \widehat{H}(X^T(G_V)) \end{aligned}$$

and the vanishing points VP_H, VP_V are then recovered at the minimum as

$$(VP_H, VP_V) = \underset{VP_H, VP_V}{\operatorname{argmin}} Q(VP_H, VP_V) \tag{1}$$

It can be seen that our method still optimizes both vanishing points when vertical repetition R_V is missing because the horizontal symmetry constraints the vertical vanishing points. Liu [17] has pointed out the potential of using symmetry in rectification, which were used by [7] to rectified facade images of 2D repetition grids. Our paper goes beyond to work with more general cases.

In this paper, individual entropies are weighted by the number of points to avoid bias from small point sets, and gradient descent is used to solve Eq 1. Our experiments show the VP refinement significantly reduces the drift of the estimated repetition interval when the initial detection is not accurate enough.

4.2 Repetition Intervals and Symmetry Axes

With the detected VPs, the original images are rectified to be fronto-parallel, and afterwards upright SIFT features are extracted (similar to the concept of U-SURF [18]). The single fixed orientation for all features is a natural choice given that the rotation is compensated through the rectification. Hence, our

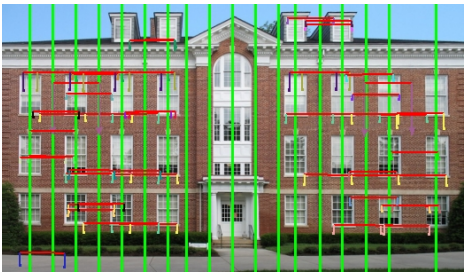

Fig. 2. An example of detected repeating features and symmetry axes. Only the feature pairs for the strongest repetition interval are displayed. It can be seen that the symmetry axes are repeating at half the interval of the window repetition.

feature matching does not suffer under descriptor changes from the erroneous orientation detections. The upright SIFT features are then matched along the horizontal and vertical direction. Note that the feature matching for reflective symmetry detection uses the mirrored matching [8].

By matching features along the horizontal direction and vertical direction, histograms of possible horizontal repetition intervals, vertical repetition intervals, and symmetries can be obtained from the features pairs. Local maxima are extracted from histograms to get a set of repetition intervals $\{I\}$ and symmetry axes $\{\Lambda\}$. In this paper, we do not try to recover vertical symmetries since they typically do not show up in urban scenes. We also skip any repetition intervals that are smaller than 30 pixels to focus on only large repetitive structures.

For each repetition interval and symmetry axis, the bounding box of their matches features gives rough regions for the repetition and symmetry. Unfortunately these regions are often inaccurate due to noise in their appearance and the ambiguity caused by small repetitive structures. To find the correct region, a dense measurement should be used.

Consistent with our assumption #3, the local symmetries and the symmetries between neighbouring repeating elements repeat with an interval of half of the structure size. See Fig. 2 for an example. Selecting the horizontal boundaries at the position of those symmetry axes maximizes the local symmetry of the repeating elements.

5 Evaluation of Repetition Quality

In order to define salient boundaries for repeating elements, we need to densely evaluate how well each location fits the repetition interval under consideration. While it is important to have some invariance to lighting changes and other small variations, non-repeating elements have to be identified. In addition, it is also important to suppress spurious support that could come from homogeneous regions and repetitions at higher frequencies (for example, the roof eaves in Fig. 2 has a repetition interval of $\frac{1}{5}$ of the window distances). We first use

image patches to evaluate the similarity between any two locations. In order to be invariant to scale changes and different rectification, the patch size W_I is selected proportionally to the repetition interval I. Through our experiments we have determined that $W_I = \frac{I}{4}$ consistently provides good results.

To provide robustness to small variations and lighting, patch similarity is evaluated by comparing SIFT descriptors, which is effeciently computed on GPU [19]. Given a repetition interval I and a location x, we use $D_R(x, I)$ to denote the distance between the normalized SIFT descriptor at x and $x + I$. Similarly, the matching distance wrt. a symmetry axis Λ is denoted as $D_S(x, \Lambda)$.

It can be verified that if an element is repeated many times, then if I is a valid repetition interval $2I, 3I, \ldots$ will also be valid. Therefore, we are interested in the smallest valid repetition interval and want to suppress its multiples. It is therefore important to verify that for a repetition interval I, the repetition intervals $\{\frac{I}{2}, \frac{I}{3}, \ldots\}$ are not valid repetition intervals. In fact, this would only have to be verified for $\frac{I}{p}$ with p prime numbers. In practice, verifying for the first few prime numbers is sufficient (we go up to 7). Notice that this automatically also covers the issue of homogeneous regions as those would verify repetition for any interval. Inspired by the widely used ratio test in SIFT matching, we choose a set of translations $T_I = \{0, \pm\frac{I}{2}, \pm\frac{I}{3}, \ldots\}$, compute the set of matching distances for them $V = \{D(x, I + t) | t \in T_I\}$, and define the following quality function

$$f(x, I) = \min(\alpha \frac{V_{(2)} + \sigma}{D(x, I) + \sigma}, 1) \tag{2}$$

where $V_{(2)}$is the second smallest distance in V. α is a parameter used for truncating the quality so that the quality function evaluates to 1 when $D(x, I)$ is significantly smaller than $V_{(2)}$ (we use $\alpha = 0.7$ as typical for the SIFT ratio test [2]). Adding a small number σ reduces noise when all distances are very small, which can be seen as a variance in the SIFT distance distribution (we use $\sigma = 0.1$). It can be seen that $f(x, I) > \alpha$ only when I is local minimum. Note that the definition works for both single patch or a patch set.

In feature matching, a small ratio between the smallest distance and the second often corresponds to a high probability of being a correct match [2], and such a ratio test filters out both ambiguous matches and poor matches. Similarly, a high $f(x, I)$ indicates high probability of x being salient repetition for interval I. The quality measure will be low if appearances change too much or if a patch matches better under other intervals or everywhere. This strategy gives penalty to both noise and ambiguous high frequency regions (e.g. Fig. 3).

As evaluation of single patches is very noisy, we define similarity and quality measures to evaluate repetition for image regions. The distance between two patch sets is defined as its the median distance: $D_R(X, I) = \text{median}\{D_R(x, I) | x \in X\}$. For quality function, we use a pre-learned threshold[1] T to select a inlier patch set $X_I\{x | x \in X, D_R(x, I) < T\}$ of an image region X, and use the inlier set to evaluate the repetition quality as $F(X, I) = f(X_I, I)$. Our experiments show that this quality function is very robust to outliers even for low inlier ratios.

[1] $T = 0.64$ learned from the distributions in labeled images is used in this paper.

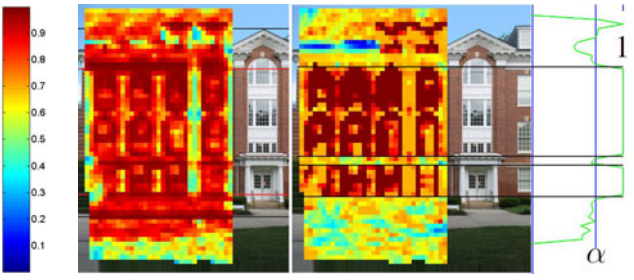

Fig. 3. Our similarity and quality measurement. The colored-patches in the left image gives the distance map (The visualization uses $1-\frac{1}{2}d^2$ to map distance $[0, \sqrt{2}]$ to $[0, 1]$). The colored patches in the right image gives the quality map and the curve gives the quality for each row. The distance map shows good matching for the grass, roof eaves and the horizontal edges, but our quality function is able to penalize them. The black lines in the right image gives the places where the vertical boundary are detected.

In order to avoids unreliable evaluation from noise, we set the quality measure to 0 when inlier ratio is less than 20%.

To correctly handle the first and last element of a repetition sequence, we define the bidirectional distance and quality

$$D^+(X, I) = \min(D(X, I), D(X, -I))$$
$$f^+(X, I) = \max(f(X, I), f(X, -I))$$

The distance map and quality map refer to D^+ and f^+ unless specified otherwise. Similar with $F(X, I)$, only inliers $X_I^+ = \{x | x \in X, D^+(x, I) < T\}$ are considered while evaluating f^+ for a patch set instead of a single patch.

6 Dense Detection

Our dense detection uses the detected sparse repetition and symmetry to obtain their initial regions, and refines them by dense matching and propagation.

6.1 Region Initialization and Propagation

It is a natural choice to select the horizontal boundaries of the repeating elements according the detected repeating symmetry axes (e.g. Fig. 2) since such boundaries generate elements with maximal local symmetry. As illustrated in Fig. 4, the initial horizontal extent of repetition region is defined by a group of symmetry axes that have horizontal distances of $\frac{I}{2}$ or I with each other. The initial vertical range is chosen to cover the matched feature pairs whose line segments intersect with the symmetry axes.

Detecting repeating elements is more complicated than detecting symmetry because the larger repetition count requires propagation and verification in order to get the full correct regions. Due to perspective change and noise, not all

Algorithm 1. The Repetition Detection Algorithm

```
Detect vanishing points and rectify image.
Find sparse repetitions {I} and symmetry axes {Λ}  (Section 4)
for each un-processed repetition interval I do
    Find sets of repeating symmetry axes {Λ_I}
    while {Λ_I} is not empty do
        Find a consecutive set of axes with gap I/2 or I
        Initialize region from the symmetry axes (Section 6.1)
        Propagate the region by matching at interval I
        Find region boundaries and sub-regions. (Section 6.2)
        Search and analyze vertical repetition.
        Find further decompositions of regions.  (Section 6.3)
        Save detected repeating elements
        Remove covered symmetry axes from {Λ_I}
    end while
    Mark repetitions that can be modeled as processed
end for
```

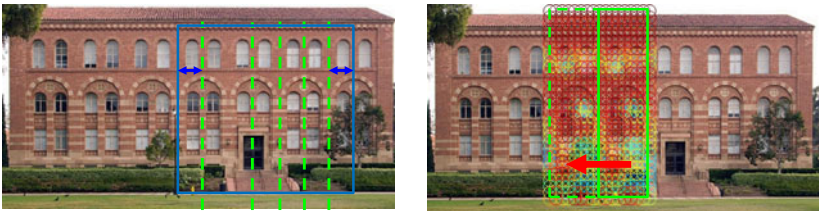

Fig. 4. Our region initialization from symmetry and propagation by dense matching

symmetry axes can be perfectly detected from initial feature matching. The initialization in previous step is likely to miss some parts of the actual repetition region. To extend the repetition region horizontally, we take steps of $\pm I$ or $\pm\frac{I}{2}$ to match a rectangular region of width I at the desired location. If the inlier ratio for both the left and right $\frac{I}{2}$ are high enough, the region is extended by the step size. Given the large window sizes, it is actually not necessary to match all the pixels. Typically, a sparse grid of locations can be used instead like Fig. 4.

6.2 Boundary Detection

Without using vertical repetition, we select the vertical boundaries based on the quality evaluation of row scanlines. Basically, we exclude regions that lack salient repetitions at interval I by simply setting boundaries where the quality of rows $F^+(X, I)$ drops from 1 to α (e.g. roof eaves and grass in Fig. 3).

With the determined vertical boundaries, multiple repeating elements can be defined after filtering out the rows without salient repetition between them.

After the horizontal repetition analysis, sparse vertical repetition analysis is applied in the detected region, and the boundaries for the vertical repetition are

Fig. 5. Example of decomposition. The color stripes in the left image shows the continuation score, and the black vertical lines give the detected element boundary edges. The right image shows the final decomposition as 4 different repetition groups.

then detected from vertical repetition quality map in a similar way. The initial region is then decomposed to sub-regions that have both horizontal and vertical repetition, and sub-regions that have only horizontal repetition.

6.3 Decomposition

As shown in Fig. 5, the possible mistakes of initializing from symmetry axes is the over-grouping of different repeating elements that have the same repetition intervals. In this case, the matching distances between neighboring elements will change over the entire horizontal range, it particularly gets large matching distances at the places where the repetition elements change. We define a continuation score function to evaluate how the repetition continues over a range of 4 times the repetition interval. Similar with the quality function, we define a continuation score from X to $X + I$ based on the ratio of distances

$$Cont(X, I) = \frac{\min(D(X_I^+ - I, I), D(X_I^+ + I, I)) + \sigma}{D(X_I^+, I) + \sigma}$$

The ratio threshold α used in repetition quality functions is basically a closeness threshold. For regions where we have good continuation of repetition, the continuation score should be in $(\alpha, 1/\alpha)$. At places where the repetition changes to something else, there will be much smaller continuation score. We particularly look for local minimums along horizontal direction that satisfy

$$Cont(X, I) < \min(Cont(X - I, I), Cont(X + I, I), \alpha)$$

Such local minimums give the possible locations that separate different repetition elements, and connecting such points vertically defines the edges between different repetition elements. To be robust to noises, we use regions of size $I \times I$ to evaluate the continuation score. Fig. 5 gives an example of the continuation score and the resulting decomposition.

7 Experiments

This section presents our qualitative results and quantitative results. We run our experiments with the same setting on all the results included in this paper.

7.1 Qualitative Results

Fig. 6 gives a few of our detection results. It can be seen that our detection algorithm robustly finds salient boundaries for both horizontal direction and vertical direction. The boundary detection is robust to occlusions, illumination changes, perspective changes, and existence of homogeneous regions and high-frequency repetition regions. As shown in of 2, 3, 4, 9, 13, 18 of Fig. 6 our algorithm detects vertical boundaries based on our quality function and correctly generates multiple repetition regions vertically.

Although our algorithm initializes the regions from symmetry axes, we do not enforce strong symmetry constraint on the detected elements. This allows the repetition detection under very large viewpoints, where the symmetry is

Fig. 6. Detection shown in the original images. Best viewed in color with 4× zoom.

Table 1. Our detection performance on the ZübuD dataset

Category	#	Percentage
No detection due to VP detection failure	25	4%
No detection due to other algorithmic limitations	34	5%
Partial Detection; Missing major repetitions	88	12%
Full detection of all major repetitions; Some boundaries errors	67	9%
Full detection of all major repetitions; Good boundaries	509	70%

very weak (e.g. 4, 9, 14, 16 in Fig. 6). In such cases, the repeating elements are detected with imperfect symmetries, and the horizontal boundaries may not be optimal. The experiments also show several of our limitations. In Fig. 6.3, the repetition from the left tower to the right tower is missing because the repetition interval is much larger than the tower width. Our current proportional patch size will not work, unless the ratio is allowed to vary. Fig. 6.8 does not detect the pure vertical repetition on the right side because our implementation currently only looks for vertical repetitions for horizontally repeating elements. We do have small errors in boundary detection like in Fig. 6.4 where too much is occluded for correct boundary detection. Fig. 6.17 has detected a wrong repetition due to inaccuracy of the second vanishing point pair.

7.2 Quantitative Evaluation

We use the ZübuD database [20] to evaluate out detection. ZübuD contains 1005 images of 201 buildings in Zürich taken from different viewpoints and illuminations conditions. We first manually filtered out 282 images that do not have clear repetitions that satisfy our assumptions (Due to occlusions, curved surface, etc). Fig. 6.17-19 are 4 examples from ZübuD. Table 7.2 is the statistics of our detection on the 723 remaining images. It can be seen that our algorithm has high successes rate for both VP detection and repetition detection.

Furthermore, we run an image retrieval experiment to evaluate the repeatability of our detection. We select the 140 buildings that have clear repetitions on at least 4 images. Our algorithm detects 10096 features in total (eacch element is counted as one; average 14 per images). Similar with SIFT descriptor, for each repeating element, we compute a 4x4 and a 8x8 gradient orientation histogram grid aligned with repeated elements to get a 128D resp. 512D feature descriptor. Paticularly, uniform weighting is used instead of Gaussian weighting to give equal importance to each cell. The distance of a feature to an image is defined as its smallest distance to all the features in that image. Given a single feature, images can be retrieved by selecting the closest ones. In this experiment, a feature-image retrieval is considered correct if the image is one of the other 4 images of the same building. For comparison, we select the 10/100 SIFT features that have the largest scales in each image to run the same experiment. Fig. 7 shows the retrieval precisions for the first 4 nearest neighbors, where our detection of repeating elements demonstrates relatively high repeatability. It is worth

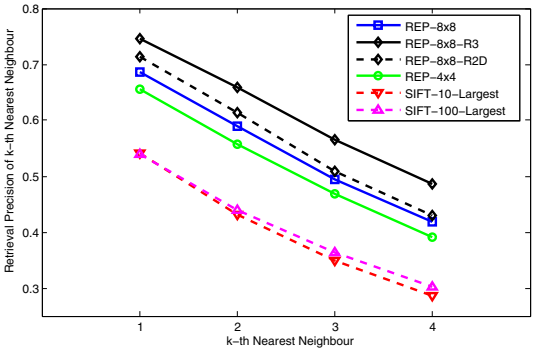

Fig. 7. Evaluation by single-feature image retrieval. REP refers to our repeating elements. 8x8 and 4x4 refers to the grid size for feature descriptor. R3 refers to the elements that repeat at least 3 times. R2D refers to the features that belong some 2D repetition grids. It can be seen that our repetition-based features achive better repeatability compared with standard image features. We believe that further improvements can be achieved with a new descriptor to capture more details. Additionally, features in R2D and R3 have better precision because they are easier to detect.

pointing out that many of the retrieval failures are due to the similar structures (especially windows) on different buildings.

8 Conclusion and Future Work

We propose a novel method to detect repeating elements on architectural facades. The main contributions are the new boundary selection for the dense repetition detection. We initialize our detection from symmetry axes to maximize the local symmetry. We also propose a quality function to conditionally evaluate how image patches fit a repetition interval, which leads to accurate vertical boundary detection. Our method is very efficient by evaluating repetition and symmetry with adaptiveness to the scale of repetitions. Typical images require only 2-4 seconds to complete the full analysis with the help of GPU. We evaluate our detection on large datasets and demonstrate the robustness and repeatability of our algorithm. Our method works particularly well for low-count and purely horizontal repetitions which has not been addressed by most previous work.

In future work, we hope to use the proposed repetition and symmetry detection scheme to automatically extract architectural grammars from images. We also hope to be able to recover missing 3D information by finding gradual changes of repetition and symmetry at different depths and generate true ortho-photos of facades from oblique views. Due to perspective changes, repeating elements at different depth that have a same 3D repetition interval will show different 2D repetition intervals in a rectified image (e.g. 9 and 14 in Fig. 6). Building further on the preliminary experiment presented in the evaluation, an interesting area of future work is to use the repetition/symmetry regions as invariant feature extractor and develop specific appearance and repetition descriptors.

References

1. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. ACM Transactions on Graphics (SIGGRAPH Proceedings) (2006)
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
3. Ripperda, N., Brenner, C.: Application of a formal grammar to facade reconstruction in semiautomatic and automatic environments. In: Proc. of the 12th AGILE Conference on GIScience (2009)
4. Wonka, P., Wimmer, M., Sillion, F., Ribarsky, W.: Instant architecture. In: SIGGRAPH 2003 (2003)
5. Mueller, P., Wonka, P., Haegler, S., Ulmer, A., Gool, L.V.: Procedural modeling of buildings. In: SIGGRAPH 2006 (2006)
6. Leung, T., Malik, J.: Detecting, localizing and grouping repeated scene elements from an image. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1064, pp. 546–555. Springer, Heidelberg (1996)
7. Schindler, G., Krishnamurthy, P., Lublinerman, R., Liu, Y., Dellaert, F.: Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In: CVPR. IEEE Computer Society Press, Los Alamitos (2008)
8. Loy, G., Eklundh, J.O.: Detecting symmetry and symmetric constellations of features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 213–225. Springer, Heidelberg (2006)
9. Wenzel, S., Drauschke, M., Förstner, W.: Detection of repeated structures in facade images. Pattern Recognition and Image Analysis 18, 406–411 (2008)
10. Schaffalitzky, F., Zisserman, A.: Geometric grouping of repeated elements within images. In: BMVC 1999 (1999)
11. Park, M., Collins, R., Liu, Y.: Deformed lattice discovery via efficient mean-shift belief propagation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 474–485. Springer, Heidelberg (2008)
12. Liu, Y., Collins, R.T., Tsin, Y.: A computational model for periodic pattern perception based on frieze and wallpaper groups. IEEE Transactions on PAMI 26, 354–371 (2004)
13. Korah, T., Rasmussen, C.: 2D lattice extraction from structured environments. In: ICIP 2007, vol. II, pp. 61–64 (2007)
14. Turina, A., Tuytelaars, T., Van Gool, L.: Efficient grouping under perspective skew. In: CVPR 2001, vol.1, pp. I-247–I-254 (2001)
15. Mueller, P., Zeng, G., Wonka, P., Gool, L.V.: Image-based procedural modeling of facades. In: Proceedings of ACM SIGGRAPH 2007/ACM Transactions on Graphics, vol. 26. ACM Press, New York (2007)
16. Tuytelaars, T., Van Gool, L., Proesmans, M., Moons, T.: The cascaded hough transform as an aid in aerial image interpretation. In: Proc. ICCV, pp. 67–72 (1998)
17. Liu, Y., Collins, R.: Skewed symmetry groups. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 872–879 (2001)
18. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
19. Wu, C.: SiftGPU: A GPU implementation of scale invariant feature transform (SIFT) (2007), `http://cs.unc.edu/~ccwu/siftgpu`
20. Shao, H., Svoboda, T., Gool, L.V.: ZuBuD — Zürich buildings database for image based recognition. Technical Report 260, Computer Vision Laboratory, Swiss Federal Institute of Technology (2003)