Frontiers of Computer Science CN 10-1014/TP
DOI 10.1007/s11704-017-6457-2 ISSN 2095-2228
RESEARCH ARTICLE

Automatic Facade Recovery from Single Nighttime Image

Yi ZHOU, Qichuan GENG, Zhong ZHOU, Wei WU

State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China

Front. Comput. Sci., **Just Accepted Manuscript** • 10.1007/s11704-017-6457-2
http://journal.hep.com.cn on January 13, 2017

Front.Comput.Sci.
DOI

RESEARCH ARTICLE

Frontiers of Computer Science
Automatic Façade Recovery from Single Nighttime Image

Yi ZHOU, Qichuan GENG (✉), Zhong ZHOU (✉), Wei WU

State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China

Abstract Nighttime images are difficult to be processed due to insufficient brightness, lots of noise and lack of details, so they are always removed from time-lapsed image analysis. It's interesting that the nighttime images have a unique and wonderful feature for buildings that has robust and salient lighting cues of human activities. The lighting variation depicts both the statistics and individual habitation, and it has inherent man-made repetitive structure from the architectural theory. Inspired by this, we propose an automatic nighttime façade recovery method that exploits the lattice structures of the window lighting. Firstly, a simple but efficient classification method is employed to get the salient bright regions which are possibly lighten windows. Then we group windows into multiple lattice proposals with respect to façades by patch matching, and after that, we greedily remove the overlapped lattices. With the horizon constraint, we solve the ambiguous proposals problem and obtain the correct orientation. At last, we complete the generated façades by filling the missing windows. This method is well suitable for urban environments, and its results can be used as a good single-view compensation for daytime images or act as a semantic input to other learning-based 3D reconstruction. The experiment demonstrates that our method works well in the nighttime image datasets, and obtain a high lattice detection rate of 82.1% in 82 challenging images, while a low mean orientation error of 12.1 ± 4.5 degrees.

Keywords façade recovery, nighttime images, lattice detection.

Received Sep 19th, 2016; accepted Jan 13rd, 2017

E-mail: *zhaokefirst@buaa.edu.cn*
zz@buaa.edu.cn

1 Introduction

Analyzing the image sequences captured from surveillance cameras has become a hot research field in modern computer vision. The analysis of daytime images has been well studied in the past few years. However, for the nighttime images, due to insufficient brightness, researchers cannot use them to extract enough useful features and further information, including 3D reconstruction [24]. In fact, nighttime images have salient visual features which can be used to achieve many tasks: repetitive window lights can be used to predict planar surface; the light which changes over time reflects the laws of peoples' work and rest; the spatial distribution of the light can be used in the recognition and classification of the visual places. These salient light features make nighttime images distinctly different from daytime images, and we should fully excavate these rich data and make a wide use of them. Furthermore, the images captured from one static camera is nearly pixel-level aligned. The analyzed results from nighttime images can be used to verify these from daytime images and the comprehensive analysis makes the result more reliable. In conclusion, the analysis of nighttime images has an important significance in computer vision. It might expand the scope of visual information acquisition and open up a new door of thought for further big data analysis.

In this paper, we focus on the geometric analysis of single nighttime image, and explore repetitive light structures on façades, which form regular grids or lattice structures. We apply this kind of light cues to parse night scenes into façade planar surfaces in man-made scenes, and estimate surface depth based on the consistency assumption of storey heights.

The presented method in this paper is an automatic algorithm to generate discrete planar surfaces. It includes three contributions to 3D parsing of a single-view night image in man-made environments, which mainly include:

- The first to take advantage of the salient window lights in a nighttime image to detect planes;
- Distinguish multiple façades of a single image effectively, and calculate their floor structures;
- An orientation estimation with low error is obtained without using any explicit geometric information, which is essential in other extraction methods of planar structures.

2 Related work

Nighttime image is divided into lighting region and dark region. Previous works related to nighttime images mainly focus on enhancement techniques [1–6], which brighten dark region, but keep light region. Most of traditional methods use contrast-based techniques such as histogram equalization [1] or tone mapping [2] to adjust the local contrast in different regions of the image and the night is improved. Raskar et al. [4] and Cai et al. [5] combine images taken at different times by using image fusion techniques. Dong et al. [6] finds a simple but efficient law for night image, and use the dark channel prior to enhance the night video in a real-time speed.

Lighting areas are both considered as visual salience. Both gradients and intensities are important and useful information. In their methods, they thought that lighting region is useful information to keep, but do not consider the latent information from lighting regions. We focus on the geometric structure from artificial lights. They all retain the high intensity or high gradient pixels. Not only these pixel is clear enough and do not need to be enhanced, lighting areas are also considered as visual salience.

Our work is closely related to façade extraction from single image, which includes three main categories: geometric properties, redundancy symmetry properties and sparse distribution.

The method based on geometric properties commonly detect vanishing points responding to families of lines which define the planesąŕ 3D orientations. The works in [7,8] detect obvious rectangular structures (such as windows), which contain two main orthogonal parallel lines in the same plane. In contrast, David et al. [9] decide the orientation map by grouping lines, which only fit Manhattan scenario, our method has a wider application in non-Manhattan ones. More importantly, their methods are built on the top of exact line detection and merging. And they are directly not suitable for featureless images, such as nighttime images.

Others discover the repetitive or similar textures and use them to infer the planar orientations [10, 11]. Even the near-regularity structures can be understood by an iterative lattice-finding method [12–15]. Park [13] conducts a MRF to represent the lattice, and use MSBP to improve the solve speed. However, the number of the planar surfaces in their experiment is often assumed as only one, and don't consider the overlapping problem between multiple lattices. Based on his single lattice finding method, Park then [14] extends his method to multiple lattices in urban scenes by perceptual grouping, which first sorts the lattice proposal, iteratively greedy adding the next-best lattice proposal with highest A-score, and remove the overlapped proposal in the rest of proposal pool. This method aims at a local optimal result and do not make sure obtain the global optimal result. Our method belongs to this category, and detects a global optimal non-overlapped (or mild overlapped) multiple façades from single nighttime image.

Additionally, the sparsity-based methods [16, 17] named TILT has been proposed to rectify near-regular texture to its frontal-view under the global affine /perspective transformation. These methods explored the intrinsic property of texture, and find the optimal planar orientation of whose rank is minimal. However, this method requires good initialization of a façade region and cannot automatically locate façade regions, which limits its application in façade detection. Another sparse method [18] detects the façade regions from aerial images via maximizing local regularity by Gini-Index, and greedily adding regions with consistent dominant orientations. Their method only describes the façade as a bounding box, and cannot obtain the detailed structure deployment.

3 Overview

Given an outdoor urban image at night, our goal is to group its window-like regions into planar façades, and get their lattice structures and orientation estimations. To achieve this goal, we first analyze the characteristics of input images and their possible reasonable assumptions.

As mentioned in section 1, the images used in our work are the nighttime image which are barely used in the field of scene geometry analysis. Compared to daytime images, nighttime images have two characteristics: low scene com-

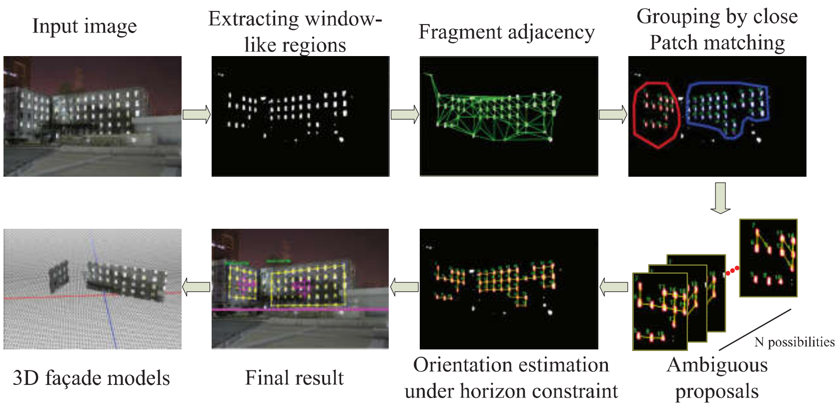
Input image
Extracting window-like regions
Fragment adjacency
Grouping by close Patch matching
N possibilities
Ambiguous proposals
Orientation estimation under horizon constraint
Final result
3D façade models

Fig. 1 Main steps of the automatic façade recovery method

plexity and salient visual feature of the light. The former means clutters in the scene become faintly visible in the dark condition and are blended into the background. In this situation, the scenes are much simpler and are easier to analyze. The latter means the illumination intensities of light regions are much stronger than the shade regions of the images. This high contrast makes light regions salient visual features which are easy to recognize.

Features, such as the color, texture or line segment [7, 8] can be effectively extracted from daytime images. However, they are not apparent any more at night caused by the poor lighting condition. Inspired by recent works on façade parsing, we use the architectural feature as the visual pattern: a salient lattice composed by a group of repetitive windows. The feature used here is different from the feature used before, and the 'line' used in our method is the high-level adjacency relation of man-made structures.

We follow two assumptions in our work. The first one is Local Manhattan World (LMW). The assumption suggests that the whole scene not only contains a Cartesian coordinate system (CCS), but also several CCSs. The parallel lines or VPs which belongs to one CCS are orthogonal to each other. All the CCSs share the same vertical axis, i.e. Except for the vertical vanishing point (or zenith), other horizontal vanishing points lie close to a single line in the image plane known as the horizon.

We also assume that the smallest recognizable information of nighttime images is the lighting window. The spaces between some windows are near the same, and by connecting these windows, we can get lattice structures. Multiple different lattices and some other clutters, such as street lamps or specular highlight, compose a man-made scene.

Based on LWM and lattice structure assumptions, we solve the façade recovery problem in three procedures. The pipeline is showed in Fig. 1. The first step is to find the window-like candidates. Then, we group candidates into multiple non-overlapped façade proposals by patch matching and greedy removal. After that, a global energy model is proposed to decide façade's orientation. And this model combines low-level representation (façade structure error) with high-level constraints (the zenith and the horizon), which optimizes both the center locations of windows and the position of zenith and horizon.

In the following sections, we detail the procedures of our automatic façade recovery method in the section 4 to 6, experimental results in section 7, and two applications of our planar façades in section 8.

4 Extracting window-like regions

We start our method by extracting window-like regions. We find that the lighten windows at night are usually salient separate regions with higher intensity than its neighboring pixels. And then they can be formulated as a set of distinguished closed connected regions.

We first detect Maximally Stable Extremal Regions (MSER) [19] $\mathbf{R}' = \{R'_i\}$ in the input image I, and use the brighter regions as the lighten candidates. However, the detection on a single image resolution often miss regions due to blur and incomplete boundary. In order to capture all the window in the image, we adopt the concept of multi-scale MESRs proposed by [20]. Each MSER region only has one level of intensity, and cannot represent a complete lighting window. So we merge adjacent regions into connected regions $\mathbf{R} = \{R_j | R_j = R'_{i_1} \cup \cdots \cup R'_{i_k}\}$. We also utilize the morphological opening and closing operations for further merging, which enforce that very close window pieces are merged

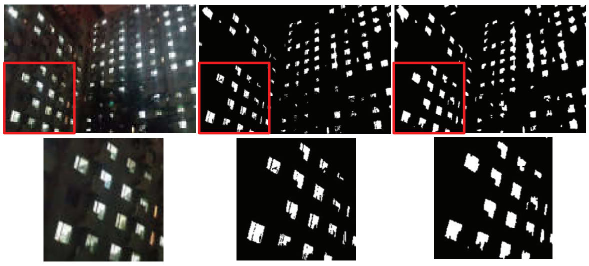

Fig. 2 Images and its close-up. Left: input image; Middle: window-like regions with our gradient-based method; Right: Lighten regions generated by OSTU threshold. The gradient-based method relieves the interference caused by indoor intensity variation.

into a single stable region. Please note that the extraction method of lighting regions used here is based on the intensity variation and connectivity of light, which will be more robust than the methods based on threshold binarization. We show the difference between gradient-based method and threshold-based method in Fig. 2.

5 Lattice detection

5.1 Group candidates into façade

To generate lattice proposals, an exhaustive search is used to make sure all the possible proposals won't be missed. Given a start seed, we start from its centroid and iteratively expand it into two of its possible direction connected to its neighbors until the search satisfies the terminal condition. We treat the window centroids as the lowest location of basins, and employ watershed method to segment the image into fragments as same amount as the window candidates. The adjacency relation of the fragments is transferred to candidates, yielding possible connecting directions for every candidate. The experiment demonstrates that it is effective enough to find correct façades, including the quantity and structures.

For two close candidates, we use a patch matching method based on normalized cross correlation (NCC). On the binary image, we check three factors in a search direction every time, including measures of the texture similarity, angle error and patch location error. The score is defined as:

$$\begin{aligned} Score(i,i') = NCC(P(i),P(i')) \cdot (1-\theta(r(i,i'),p(i,i'))) \\ \cdot(1-sigmoid(\|r(i,i')-p(i,i')\|_2/\|p(i,i')\|_2)) \end{aligned} \tag{1}$$

i' is the predicted neighbor of candidate i. The $NCC(,)$ is used to check the texture similarity, while $P(i)$ is the image region at the candidate i. $r(i,i')$ denotes the direction vector between the real centers of patch i' and i, and $p(i,i')$ is the predicted one. $\theta(r(i,i'),p(i,i'))$ is the angle distance between this two vectors. The sigmoid function is used to punish the Euclidean distance between the predict center and the real center. We repeat the search procedure for every candidate in the image, resulting a set of overlapped lattices. And when the score is lower than a threshold T_s, the search in this direction will end.

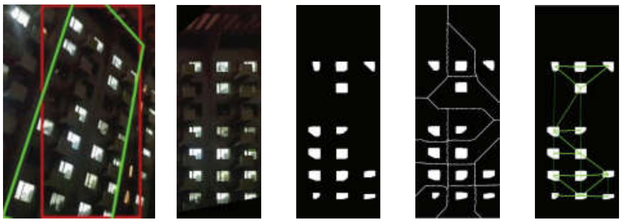

Fig. 3 The steps of window adjacency relationships generation. (a) Input image (b) rectified image(not necessary, helpful for expressing idea) (c) window label (white for window, black for non-window). (d) shows watershed segmentation result, and (e) their adjacency relationships give search directions for iterative expansion.

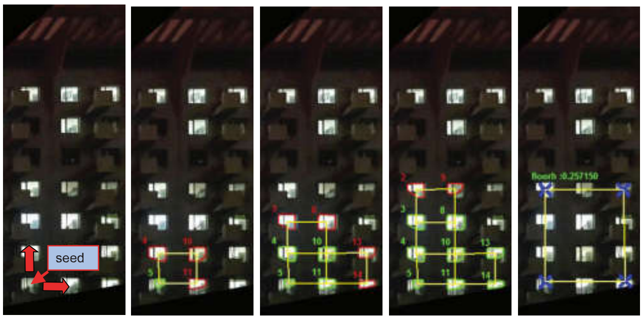

Fig. 4 Lattice proposal generation. We iteratively expand seed windows (a) into the two suggested direction. After three steps of expanding (b-d), (e) shows final parsing result.

5.2 Lattice energy representation

Since the relative position of features is important for characterizing lattice structures, we model this relationship with energy representation. We choose the start candidate to represent its lattices. The output is a set of seed windows and their lattices. We use the regular linear error to measure the quality of lattices.

We denote the set of window candidates as $\mathbf{C} = \{c_1, \cdots, c_N\}$ and the corresponding set of labels as $\mathbf{L} = \{l_1, \cdots, l_N\}$. c_i is the centroid of ith window candidate in the image. I denotes the correct window set within the lattice proposal. O denotes the outlier set. $L \in I \cup O$ represents either an outlier or a window in the lattice. The energy function is written as below.

$$E_{lattice} = E_{data} + E_{struct} + E_{outlier} \tag{2}$$

We try to explain as many candidates as possible by using a regular lattice structure. $E_{data} = -|\{l_i | l_i \in I\}|$ is the negative number of explained lattice elements. $E_{outlier} = \lambda_o \cdot |\{l_i | l_i \in O\}|$ represents the number of outlier candidates in lattice's convex bounding hull. And E_{struct} measures how well candidate i can be predicted by its neighboring candidates k and j, weighted by the number of explained candidates:

$$E_{struct} = -|\{l_i | l_i \in I\}| \cdot \max_{i,j,k \in I} \frac{||c_j + c_k - 2c_i||_2}{||c_j - c_k||_2} \tag{3}$$

In more general situation, the direct neighboring j and k of candidate i may be missing at this lattice proposal. Then we use the candidates nearest to candidate i to replace it. E_{struct} can be rewritten as below:

$$E_{struct} = -|\{l_i | l_i \in I\}| \cdot \max_{i,k \in I} \frac{||cPre_{ik} - cReal_{ik}||_2}{||cPre_{ik}||_2} \tag{4}$$

$$cPre_{ik} = \begin{vmatrix} c_m - c_n \\ c_m - c_k \end{vmatrix} \setminus \begin{vmatrix} x_m - x_n & y_m - y_n \\ x_m - x_k & y_m - y_k \end{vmatrix} \cdot \begin{vmatrix} x_k - x_i \\ y_k - y_i \end{vmatrix} \tag{5}$$

$$cReal_{ik} = c_k - c_i \tag{6}$$

In Eq. 5, $k, m, n \in I$ are three non-collinear nearest neighboring candidates from the lattice. And (x, y) is their locations in the lattice structure. We predict a vector from jth candidate to the ith, denoted as $cPre_{ij}$, while $cReal_{ij}$ is the real one.

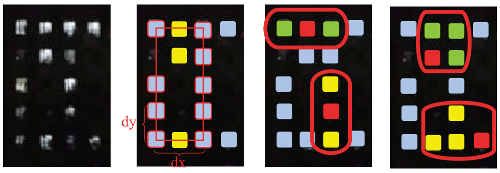

Fig. 5 (a) Input image. (b) Three outliers (yellow regions) found in this red lattice proposal. (c) Linear condition of three neighbor candidates, yellow and green candidates is used to represent the red candidates. (d) when its directly neighboring candidates are partly missing, the red candidate should be represented by its nearest neighbors.

5.3 Overlapping removal

Different lattice configurations may overlap each other, which means that a candidate belongs to two or more lattices. However, this situation is impossible to happen in real world. We must select a set of feasible non-overlapped proposals which covers as many images as possible. A greedy method is used to remove duplicates by keeping only the top scoring proposal first. After overlapping removal, we get a set of seed windows and its responding planar lattices.

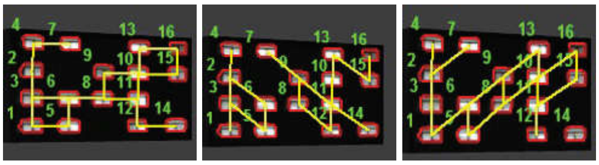

Fig. 6 Three proposals with same window candidates but different directions. Left shows a right direction.

6 Orientation estimation

6.1 Horizon constraint

If we do not use edge features to decide the search direction, we will yield a group of possible ambiguous proposals, as shown in Fig. 6. These proposals have same window candidates but different directions. With the horizon constraint, we can solve this problem. Actually, we refine both low-level locations of lattice elements and high-level lattice orientations in one equation.

6.2 Global energy model

We now explain our global energy model of the scene in our method. The input is several non-overlapped clusters of lattice configurations. And a lattice decides two principle direction: one for horizontal vanishing point and another for zenith. The lattice is denoted as $\mathbf{T} = \{t_i\}_{i=1,...,T}$, and its possible horizontal VP h_i, vertical VP z_i (computed by the parallel lines in lattice structure). All the vertical VP is equal, and we have $z_1 = z_2 = \cdots = z_T = \mathbf{z}$. $\mathbf{z}$ is the zenith, corresponding to the parallel line family in the vertical direction.

The energy function in our method includes an individual energy E_{ind} for every lattice t_i, which defined as:

$$E_{ind}(t_i | h_i, z_i) = \min\{\lambda_t \cdot E_{lattice,i}, \lambda_c\} \tag{7}$$

In Eq. 7, E_{ind} is decided by the lattice energy $E_{lattice,i}$, which is the ith rectified lattice in the image and will vary from h_i and z_i. λ_t is a weighted constant to balance this term and other terms. λ_c is a large value when l_i cannot be represented by h_i and z_i.

As we mentioned in section 3, all the local CCSs share the same vertical vanishing point. And their horizontal VPs lie closely to the horizon line in the image plane. We enforce this by measuring the angle distance between horizon line and connection line of two lattices' horizontal VPs. Then the energy term under the horizon constraint is defined as:

$$E_{horizon}(h_i, h_j|\mathbf{z}) = \lambda_h \cdot f(\varphi(h_i - h_j, u - z)) \tag{8}$$

where φ is the absolute angle between $u - \mathbf{z}$ and $h_i - h_j$. z is the zenith, and u is the principle point of the camera which is usually the center of this image. $f(\cdot)$ is a monotonic increasing function, and it should penalize (up to $+1$) on strong non-orthogonality, since $u - \mathbf{z}$ is perpendicular to the horizon line. We choose the tangent function as this function f. λ_h is weighted constant to balance this term and other terms.

Then the final energy is defined as:

$$\begin{gathered} E_{total}(\mathbf{h}, \mathbf{z}|\mathbf{T}) = \sum_{i=1\cdots T} (t_i|h_i, z_i)+ \\ \sum_{1\leq i\leq j\leq T} (h_i, h_j|\mathbf{z}) + E_{smooth}(\mathbf{h}) \end{gathered} \tag{9}$$

where $E_{smooth} = \lambda_s \cdot ||\mathbf{h}||_0 = \lambda_s \cdot T$ is a smooth term penalizing the number of lattice number. λ_s is the constant regulating the strength of this term. The energy E_{total} combines the two stage components into one equation. And the optimized lattice result can be obtained by minimizing this global model.

6.3 Approximation solution

The minimization of Eq. 9 is a hard computation problem when the lattices number T is large and the locations of vanishing point vary randomly in the image near the true values, so its solution necessitates the use of approximations. We use its discrete approximation method to solve the minimization of the global energy:

$$E_{discrete}(\mathbf{x}, \mathbf{z}|\mathbf{T}) \equiv E_{total}(\{\hat{h}_k\}_{k:x_k=1}, \mathbf{z}|\mathbf{T}) \tag{10}$$

The variable x is binary and decides whether a horizontal vanishing point hi is present ($x_i = 1$) or absent ($x_i = 0$) in the image. The discrete energy is defined as the continuous energy of the appropriate subsets of vanishing points.

The discrete energy defined in Eq. 10 can be written as:

$$\begin{gathered} E_{discrete}(\mathbf{x}, \mathbf{z}|\mathbf{T}) = \sum_{i=1\cdots T} E_{ind}(t_i|\{\hat{h}_i\}_{i:x_i=1}, z_i)+ \\ \sum_{1\leq i\leq j\leq T} x_i \cdot x_j \cdot (\hat{h}_i, \hat{h}_j|\mathbf{z}) + \lambda_s \cdot \sum_{i=1\cdots T} x_i \end{gathered} \tag{11}$$

We vote the zenith $\mathbf{z}$ from the whole lattices based on the evident that the vertical vanishing point is on the extreme top of image. Given the fixed $\mathbf{z}$, we then perform optimization over the binary variables x through the Iterated Conditional Modes (ICM) algorithm with the randomized node visiting order.

6.4 Expand and merge lattices

After affirming that the principle directions of all the lattices are correct, we expand lattices and make them contain more window regions as many as possible. We relax the terminal condition and expand the lattice into its two principle directions once more, to verify its max possible size and update its orientation. If two lattices share same principle directions and adjacent windows, we merge them into one, as shown in Fig. 7. These operations are designed for reducing the error situation that split happens among the detected lighten windows.

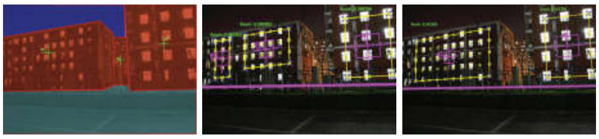

Fig. 7 Illustrations of how to merge façade. Left is the ground truth.

7 Results

7.1 Datasets for Evaluation

We collect a night dataset Night campus with manual annotation of façade segmentation and surface orientation. The surface orientation is calculated with the Samsung GALAXY S4's K330 gyroscope sensor and GPS sensor clinging to the camera. The dataset consists of 82 images captured by ourselves, and there are more than two lattice-like façades per image on average in this dataset. To our knowledge, this dataset is the first public benchmark for night façade parsing and providing planar annotations. It is a challenging dataset for testing our planar lattice detection method.

7.2 Façades recovery performance

We show several qualitative and quantitative results in Fig. 8. Fig. 8(a) illustrates successful separation of lighting windows in planar façades from street lamps in non-planar regions, while Fig. 8(c) shows an effective grouping of salient regions which belong to different lattices. We also draw the estimation of the horizon on test images. Compared to the ground truth, the estimated horizon line is located within a reasonable range.

The runtime of our algorithm depends on the number of repeated patches detected. The average runtime on a 2.6GHz quad-core CPU is about 2 min. Some images in the urban scenes category take longer than 5 min due to the large number (≥ 400) of repetitive patterns detected.

7.3 3D orientation error estimation

We obtain an orientation error for all façades of 12.1 ± 4.5 degree in 82 nighttime images. We believe the mean orientation accuracy of 12.1 to be very reasonable since the method makes no explicit use of geometric information, such as line segments. In addition, the angle error may be lower if we consider the measurement error of the hardware. The gyroscope sensor we used is not a strict high-precision attitude sensor, and in our test, its orientation error will reach ±2° when the phone keeps a large pitch angle.

As shown in Fig. 9, we plot the distribution of orientation errors. A small amount of outlying lattices have large errors, and a significant number is under 15° (70%) and under 20° (86%).

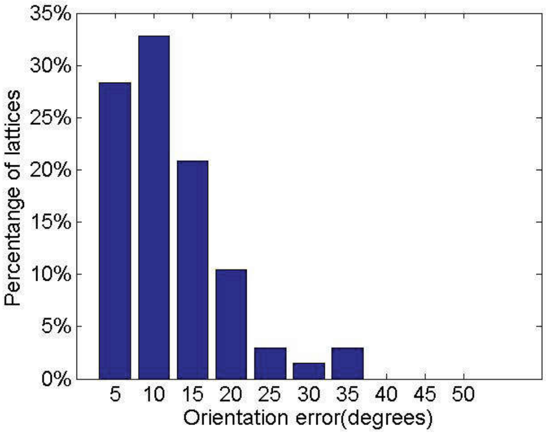

Fig. 9 Histogram of orientation errors in the night dataset, showing that the majority of regions are given an orientation estimate with low error.

7.4 Comparison with state-of-the-art lattice detection methods

We have compared the proposed algorithm, against Park et al. at PAMI 2009 [13], and their another work at ACCV 2010 [14]. We tested three algorithms on the night campus dataset with our self-labeled ground-truth. For a fair comparison, we improve PAMI 2009 to support multiple lattice detection. We first run PAMI 2009, then remove the convex hull of image parts where the 2D lattice is found, and repeat until no more lattices are found. For ACCV 2010, we also keep their feature aggregation from a variety of interest point detectors, which contains the same detector MSER we used.

As can be seen in Fig. 10, we give several typical results. Comparing columns 2 3and 4, we can see that the grouping window-like regions in our method has low constraint to the color of regions, and doesn't enforce regions in same lattices have consistent texture, so PAMI 2009 and ACCV 2010 have better texture regularity than ours. But exactly because of this reason, our method has other strength that we can group light-variant windows into one lattice. In row 1 of Fig. 10, our method detects both dark window and lighting window in all the three lattices, and our lattice is more complete than PAMI 2009 and ACCV 2010. In row 2, our method detects one lattice on the right of image, and other two methods do not find it. On the other hand, due to the dark light of the night, the method is likely to find un-meaningful lattices. In row 3 of column 1, PAMI 2009 find a lattice(the pink one) which don't fit the frontal facets of the building, in other words, the basic vector pairs of lattice aren't aligned with vertical or horizontal direction of the building. Although ACCV 2010 can has a correct direction by using symmetry [14], but it doesn't always work in nighttime image due to the dark vision in the night. Our method uses global horizon constraint, and has better trend to the direction of façade.

7.5 Failure modes and discussions

We report some modes where our algorithm fails. Fig. 11(a) shows an example where one façade is completely missed caused by incorrect extraction of lighting regions where outdoor illumination exists. And Fig. 11(b) shows another window-missed result because of the excessive corrupted situation in real scenes. Above all, our primary failure mode is the lack of enough easy-detected lighting windows. This problem stems from the nature of input nighttime images and cannot been simply overcome on the algorithm level. If we indeed need a solution, we thought that a statistical image from long term sequence at night might give this algorithm a well-formed input.

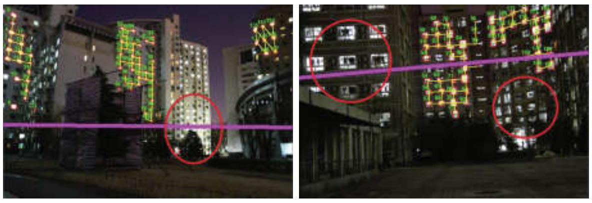

Fig. 11 Failure results. The error region is marked in red circles.

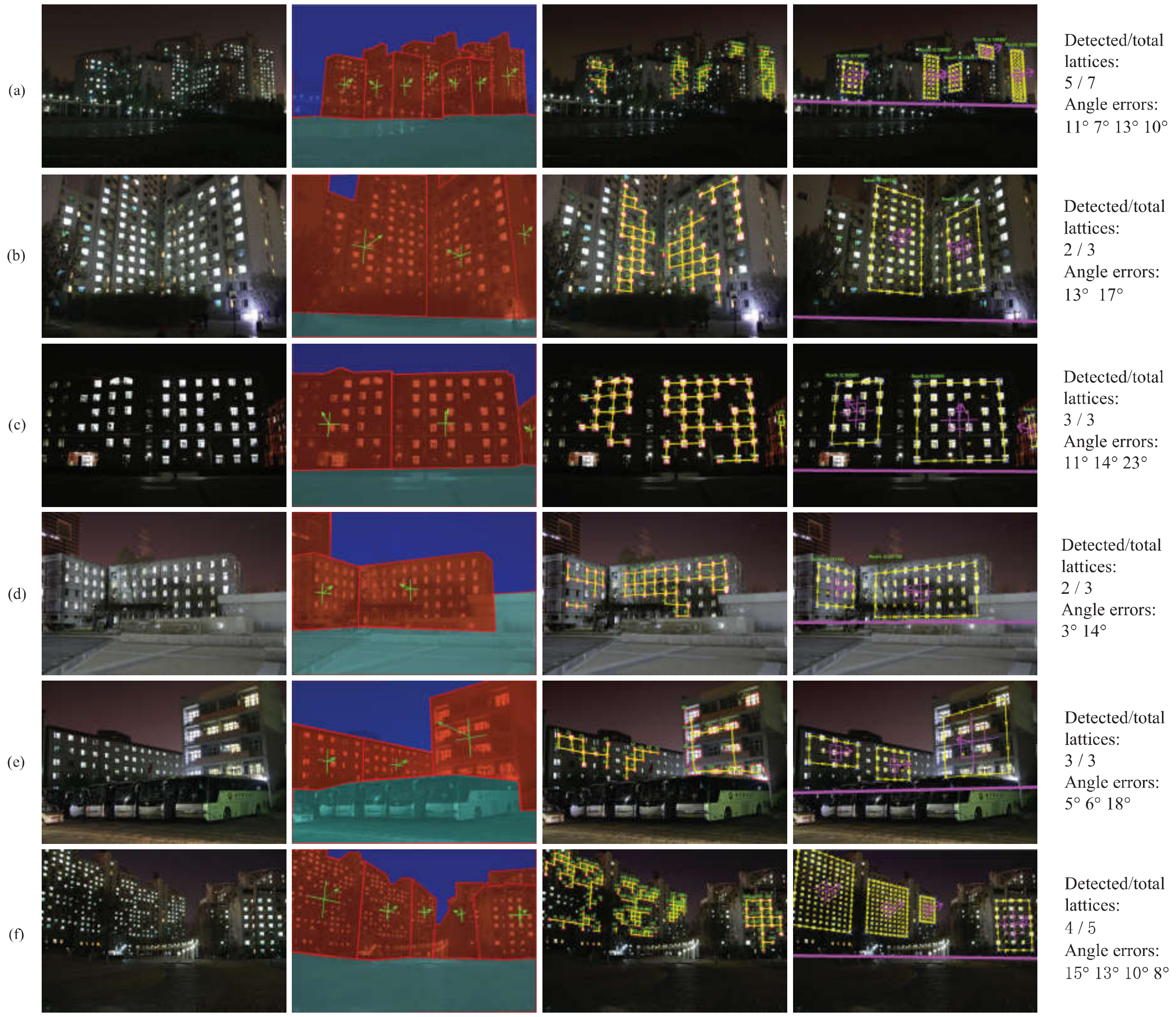
(a)
Detected/total lattices: 5 / 7
Angle errors: 11° 7° 13° 10°
(b)
Detected/total lattices: 2 / 3
Angle errors: 13° 17°
(c)
Detected/total lattices: 3 / 3
Angle errors: 11° 14° 23°
(d)
Detected/total lattices: 2 / 3
Angle errors: 3° 14°
(e)
Detected/total lattices: 3 / 3
Angle errors: 5° 6° 18°
(f)
Detected/total lattices: 4 / 5
Angle errors: 15° 13° 10° 8°

Fig. 8 Exemplar results on the night campus dataset. Column 2 shows the hand-labeled façades segmentation and its ground truth orientation. Columns 3-4 show our results, including a lattice detection result and an orientation result (pink line denotes the horizon line).

8 Applications

8.1 3D reconstruction

We all know that the same kind of buildings in 3D world all possess an approximately consistent storey height, which fits the architectural specification. If we know a façade's height in world coordinate system, other façades' height in the real world can be computed. And we can estimate relative depth via 3D orientation.

We check the orthogonality between each two horizontal VPs and unify all the LMW, which means every façade shares same focus length. Given two orthogonal VPs, focus length can be estimated by the method mentioned in [21]. We use other horizontal VPs to check the focus length estimated by the first horizontal VP and zenith. After unifying the focus length, the relative depth can be estimated. We use 3D billboards to model façades. Fig. 12 shows two results of 3D reconstruction.

8.2 Surface layout enhancement for daytime images

Another kind of works can be done with the nighttime image and its associated daytime image. We treated the lattice structures from nighttime images as a kind of accuracy auxiliary information to enhance the rough surface orientation estimation from daytime images [22], which aims at arranging each pixel in vertical regions a possibility vector for discrete orientation proposals (planar left, front or right). The key point here is that the lattice stores the similarity of nodes in a global description, which will benefit the scene geometry un-

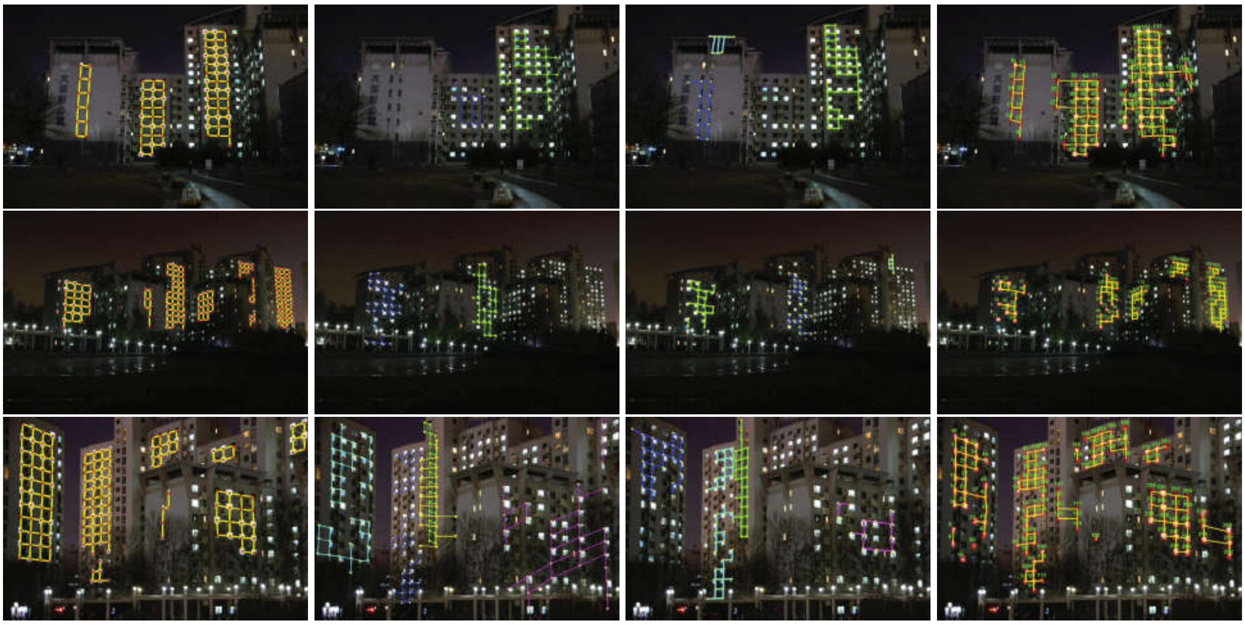

Fig. 10 Comparison with Park [13, 14] on a set of examples of night campus dataset. Column 1 is the ground-truth, Columns 2 and 3 is the results of PAMI 2009 and ACCV 2010, and column 4 is our results.

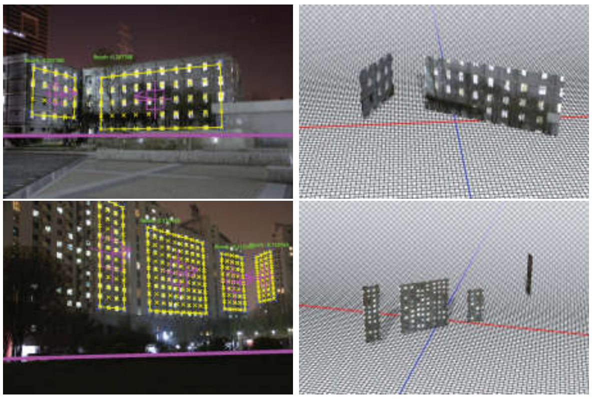

Fig. 12 Left is the result of planar detection, and the right is its façade models based on the consistency assumption of storey height. LeftBottom displays a subtle oriented distinction between three neighboring façades.

derstanding from another perspective. Thus, we can use this similarity to adjust the probability vector of orientations.

To enhance the surface layout estimation, we firstly adopt the Generalized Dual-Bootstrap ICP (GDB-ICP) proposed by Yang G. et. al. [23] to align day-night image pairs. Then the lattice in night can be transformed to daytime image. For each pixel located in the lattice regions, we build a possibility vector, in which component is computed as the cosine distance between lattice normal and its component direction. And we complement the original probability vector from local pixel by adding this weighted vector from lattices. The final surface orientation is decided by the component with the maximum possibility. The Fig. 13 shows an aligned instance of complex image pairs, and the Fig. 14 compares the results of surface layout estimation before enhancing and after enhancing.

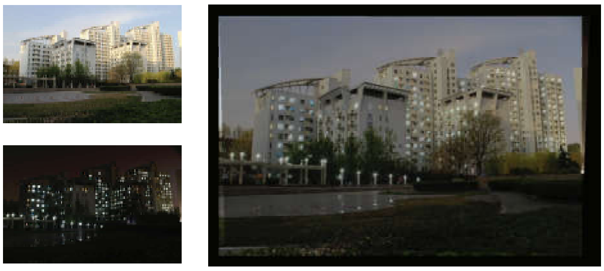

Fig. 13 Day and night image pairs; and its aligned result.

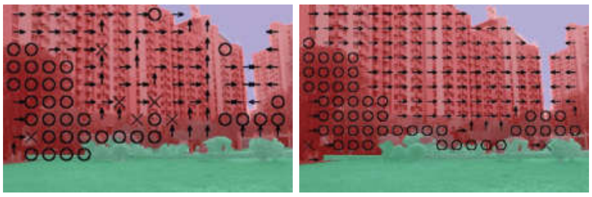

Fig. 14 Surface layout estimation. We label a planar left region with a left arrow, and a planar right region with a right arrow. The hollow circle denotes the region is porous, while the cross denotes the region is solid.

9 Conclusion

We have presented an automatic façade recovery method for a single night image by detecting lattice structure. And it can group the salient regions of an image into planar surfaces, and get their lattice structures and orientation estimates. We demonstrate its high detection rate and low-error orientation estimation on real scenes data.

Our method cannot be used to deal with the situation in which only a small majority of windows are lighted because of the predominance of the corrupted or missing data. A direction of future work, therefore, is to incorporate time-lapsed video or daytime images, which can be used to collect multiple information from human activities and synthesize ‘all light open’ images. What’s more, the method is not relied on line segments or other explicit geometric information, so it can easily be extended to integrate other monocular cues. It will be interesting to combine the planar information with geometric context labels provided by a daytime image [22] more closely, and could achieve a more reasonable spatial relationship and complete 3D model.

Acknowledgements We gratefully acknowledge the proofreading of Doctor Yuehua Wang. This work is supported by the National 863 Program of China under Grant No.2015AA016403, the Natural Science Foundation of China under Grant No.61572061, 61472020, 61502020, and the China Postdoctoral Science Foundation under Grant No. 2013M540039.

References

1. Gonzalez R. C, Woods R. E. Digital Image Processing. New Jersey: Prentice Hall, 2008
2. Durand F, Dorsey J. Fast bilateral filtering for the display of high dynamic range images. In: Proceedings of the 29th International Conference and Exhibition on Computer Graphics and Interactive Techniques. 2002, 257–266
3. Rao Y, Chen L. A survey of video enhancement techniques. Journal of Information Hiding and Multimedia Signal Processing, 2012, 3(1): 71–99
4. Raskar R, llie A, Yu J. Image fusion for context enhancement and video surrealism. In: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering(NPAR). 2004, 85–152
5. Cai Y, Huang K, Tan T, Wang Y. In: Proceedings of 13rd International Conference on Pattern Recognition. 2006, 980–983
6. Dong X, Wang G, Pang Y, Li W, Wen J, Meng W, Lu Y. Fast efficient algorithm for enhancement of low lighting video. In: Proceedings of International Conference on Multimedia and Expo. 2011, 1–6
7. Micusik B, Wildenauer H, Kosecka J. Detection and matching of rectilinear structures. In: Proceedings of Computer Vision and Pattern Recognition. 2008, 1–7
8. Kosecka J, Zhang W. Extraction, matching, and pose recovery based on dominant rectangular structures. In: Proceedings of International Workshop on Higher-Level Knowledge in 3d Modeling and Motion Analysis. 2003, 83
9. David L, Martial H, Takeo K. Geometric Reasoning for Single Image Structure Recovery. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009, 2136–2143
10. Wu C, Frahm J, Pollefeys M. Detecting large repetitive structures with salient boundaries. In: Proceedings of European Conference on Computer Vision, 2010, 142–155
11. Schindler G, Krishnamurthy P, Lublinerman R, Liu Y, Dellaert F. Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2008, 1–7
12. Hays J, Leordeanu M, Efros A. A, Liu Y. Discovering texture regularity as a higher-order correspondence problem. In: Proceedings of European Conference on Computer Vision. 2006, 522–535
13. Park M, Brocklehurst K, Collins R T, Liu Y. Deformed Lattice Detection in Real-World Images Using Mean-Shift Belief Propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(10): 1804
14. Park M, Brocklehurst K, Collins R T, Liu Y. Translation-Symmetry-Based Perceptual Grouping with Applications to Urban Scenes. In: Proceedings of Asian Conference on Computer Vision. 2010:329-342
15. Liu S, Ng T T, Sunkavalli K, Do M N, Shechtman E, Carr N. PatchMatch-Based Automatic Lattice Detection for Near-Regular Textures. In: Proceedings of IEEE International Conference on Computer Vision. 2015, 181–189
16. Mobahi H, Zhou Z, Yang A Y, Ma Y. Holistic 3d reconstruction of urban structures from low-rank textures. In: Proceedings of IEEE International Conference on Computer Vision Workshops. 2011, 593–600
17. Zhang Z, Ganesh A, Liang X, Ma Y. TILT: transform invariant low-rank textures. International Journal of Computer Vision, 2012, 99(1): 1–24
18. Liu J, Liu Y. Local Regularity-Driven City-Scale Facade Detection from Aerial Images. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2014, 3778–3785
19. Matas J, Chum O, Urban M, Pajdla T. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. Image and Vision Computing, 2004, 22(10): 761–767
20. Forssen P E, Lowe D G. Shape Descriptors for Maximally Stable Extremal Regions. In: Proceedings of IEEE International Conference on Computer Vision. 2007, 1–8
21. Guillou E, Meneveaux D, Maisel E, Bouatouch K. Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. The Visual Computer, 2000, 16(7): 396–410
22. Hoiem D, Efros A A, Hebert M. Geometric context from a single image. In: Proceedings of IEEE International Conference on Computer Vision. 2005, 654–661

23. Yang G, Stewart C V, Sofka M, Tsai C. Registration of challenging image pairs: initialization, estimation, and decision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(11):1973-1989
24. Pan Z, Zhang Y, Kwong S. Efficient motion and disparity estimation optimization for low complexity multiview video coding. IEEE Transactions on Broadcasting, 2015, 61(2) 166-176

Yi Zhou was born in 1988. He received the Bachelor degree in Computer Science and Technology from Beihang University, in 2010. Currently, he is a Ph.D. candidate at State Key Lab of Virtual Reality Technology and Systems. His research interests include augmented virtual reality and time-lapsed video understanding. He is a student member of CCF.

Qichuan Geng was born in 1989. He received the Bachelor degree in School of Automation Science and Electrical Engineering from Beihang University, in 2012. Now, he is a Ph.D. candidate at State Key Lab of Virtual Reality Technology and Systems. His research interests include augmented virtual reality and semantic segmentation understanding. He is a student member of CCF.

Zhong Zhou is a professor at State Key Lab of Virtual Reality Technology and Systems, Beihang University, Beijing, China. He received his BS degree from Nanjing University and PhD degree from Beihang University in 1999 and 2005, respectively. His main research interests include natural phenomena simulation, augmented virtual reality, and Internet-based virtual reality technologies. He is a member of China Computer Federation and Institute of Electrical and Electronics Engineers.

Wei Wu is a professor in the School of Computer Science and Engineering at Beihang University, currently the chair of the Technical Committee on Virtual Reality and Visualization (TCVRV) of the China Computer Federation (CCF). He received the PhD degree from Harbin Institute of Technology, China, in 1995. He has published more than 90 papers, 33 issued patents, and one book. His current research interests involve real-time 3D reconstruction, remote immersive system, and augmented reality.