Detection and matching of rectilinear structures

Conference Paper *in* Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition · July 2008

DOI: 10.1109/CVPR.2008.4587488 · Source: IEEE Xplore

3 authors:

Branislav Micusik
AIT Austrian Institute of Technology
58 PUBLICATIONS **1,468** CITATIONS
SEE PROFILE

Horst Wildenauer
TU Wien
34 PUBLICATIONS **563** CITATIONS
SEE PROFILE

J. Kosecka
George Mason University
47 PUBLICATIONS **2,324** CITATIONS
SEE PROFILE

Detection and Matching of Rectilinear Structures*

Branislav Mičušík[1] Horst Wildenauer[2] Jana Košecká[1]
[1]George Mason University, USA [2]Vienna University of Technology, Austria

Abstract

Indoor and outdoor urban environments posses many regularities which can be efficiently exploited and used for general image parsing tasks. We present a novel approach for detecting rectilinear structures and demonstrate their use for wide baseline stereo matching, planar 3D reconstruction, and computation of geometric context. Assuming a presence of dominant orthogonal vanishing directions, we proceed by formulating the detection of the rectilinear structures as a labeling problem on detected line segments. The line segment labels, respecting the proposed grammar rules, are established as the MAP assignment of the corresponding MRF. The proposed framework allows to detect both full as well as partial rectangles, rectangle-in-rectangle structures, and rectangles sharing edges. The use of detected rectangles is demonstrated in the context of difficult wide baseline matching tasks in the presence of repetitive structures and large appearance changes.

1. Introduction

Rectilinear geometric structures are one of the most commonly encountered structures in man-made indoor and outdoor environments. In many instances one can directly associate semantic labels with detected regions of rectangular shape such as doors, windows, posters, tables, building facades, *etc*. As planar structures, they can also be viewed as large support regions of co-planar points and hence provide an effective alternative for difficult wide baseline matching tasks and subsequent piecewise planar 3D reconstruction. These are just few reasons why efficient and reliable detection of the rectangular structures is of importance.

The two main contributions of the presented work are in a new method for detection of full or partial rectangles as well as a novel approach for their matching and pose recovery in a difficult wide baseline setting. We will also briefly demonstrate how the rectangles can serve as support regions for geometric context computation in case of poorly textured environments in the spirit of [7], see Fig. 1.

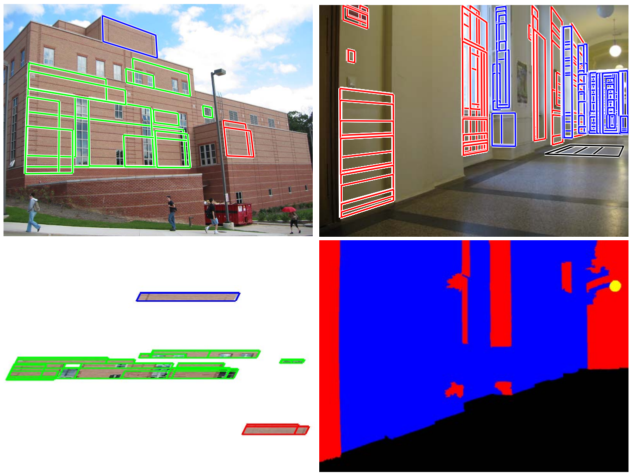


Figure 1. The detection of rectangles and their applications. *Left column:* Wide baseline stereo matching. An outdoor image with detected and matched rectangles (only large ones are shown) with the first image from Fig. 8 and top view of the two-view piecewise planar 3D reconstruction. *Right column:* The MRF-based orthogonal plane detection. An indoor image with rectangles detected separately in each orthogonal plane and a resulting plane segmentation. Each plane is depicted in a different color.

We formulate the rectangle detection as a labeling problem in the Markov Random Field (MRF) framework where the underlying graph structure is obtained by detected line segments connected through Constrained Delaunay Triangulation. The proposed approach allows to detect partial rectangles, facilitates edge sharing between neighboring rectangles, and rectangle-in-rectangle configurations. Furthermore, we demonstrate their effective matching in the presence of large changes in viewpoint by endowing them with the Discrete Cosine Transform based descriptor. Finally, their consequent piecewise planar 3D reconstruction, as shown in Fig. 1, is obtained by plane sweeping.

Related work. Several approaches for localization of rectangular structures have been proposed in the past. With the exception of few, they typically start with the detection

*The research has received funding from US National Science Foundation Grant No. IIS-0347774 and the European Community's Sixth Framework Programme (FP6/2003-2006) under grant agreement No. FP6-2006-IST-6-045350 (robots@home) .

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

of line segments and a subsequent estimation of orthogonal vanishing directions. The approach of [15] proceeded with instantiation of planar hypothesis in a bottom up manner by linking and grouping detected line segments to form initial rectangular hypothesis. The rectangular regions obtained in such a manner have a small extent and are prone to mismatching especially in the presence of repetitive structures. In the work on symmetry-based 3D reconstruction [18] rectangles are extracted by using color segmentation allowing only for structures with uniform appearance. In [10] multiple rectangle hypotheses are found by exhaustive computation of intersections between line segments coming from different orthogonal vanishing directions. Subsequently the hypotheses are verified by checking gradient consistency in the regions with the vanishing directions. This method is computationally very expensive generating many superfluous hypotheses. Moreover, the hypothesis verification uses strong assumptions about appearance which is not valid in many environments. The approach of [6] utilizes both geometric and luminance constraints and the rectangle detection is formulated as a search for most likely associations of pairs of line segments. All possible pairs of lines are considered and the most compatible assignments are obtained in the relaxation framework. The applicability in outdoor scenarios is hampered by the luminance dependent components of the compatibility term used in the relaxation.

The approach most similar to ours in the sense of using the MRF and context sensitive grammar rules for rectangle detection is work of [4]. Our work focuses on the simpler task of the detection of individual rectangles instead of modeling their mutual alignment resulting in lower computational requirements. We formulate the detection of the rectangles on a restricted neighborhood structure given by Delaunay triangulation to keep the problem tractable and efficient, while solvable on a global level.

The detected rectangles as salient regions offer nice properties in context of wide baseline stereo matching (WBS) where notable progress, reviewed in [11, 13], has been made utilizing interest points or regions. Although the descriptors of these primitives have favorable invariance properties, their applicability is still limited to relatively small out of plane rotations and affine distortions. Difficult cases still arise when the viewpoint change is very large and perspective distortion becomes significant. The importance of the rectangles in this context is underlined by their ability to tackle the perspective distortion directly.

Difficulties in successfully matching points and line segments across multiple-views with large baseline have compelled researchers to employ assumption on piecewise planarity of observed urban environments to produce a reasonable 3D reconstruction [2, 17, 1, 3]. Building on the advantageous properties of the detected rectangles, we demonstrate how to use the plane sweeping idea to instantiate additional rectangles lying on a non-dominant plane and hence facilitate piecewise planar 3D reconstruction.

The structure of the paper is the following. We explain the MRF-based detection of rectangles in Sec. 2 and their favorable use for the WBS matching in Sec. 3 demonstrated on some example images in Sec. 4. We will also briefly show in Sec. 4 how to use the rectangles as features of intermediate support to compute the geometric context from a single image.

2. Detection of rectangles

We aim at the detection of those rectangles in an image of man-made environment which are perspectively projected into the image as quadrilaterals aligned with two out of three dominant scene directions. These dominant directions are captured in the image via vanishing points. To tackle the quadrilateral detection in an effective way we formulate the problem as a search for the Maximum Aposteriori Probability (MAP) solution of the MRF defined on lines consistent with the vanishing points. Such formulation allows to avoid an exhaustive search over rectangle hypotheses coming from all possible intersections of detected lines in the image [10]. The proposed strategy restricts the space of accepted rectangles directly in the inference stage by imposed grammatical rules.

The main steps of our method are the following. *i)* Line segments and vanishing points are localized and used for camera auto-calibration. Each line segment is assigned to its corresponding vanishing direction. *ii)* A graph representing the MRF is constructed from the detected line segments respecting vanishing direction assignment and geometric properties between pairs of the neighboring lines; encoded via data and smoothness terms. *iii)* The MAP is computed yielding a unique label, representing one of four rectangle edges, assigned to each line segment such that meaningful rectangles are established.

2.1. MRF formulation

We formulate the MAP of the MRF via the equivalent MAX-SUM labeling problem. First, we give its formal definition and later, in the next section, we explain meaning of the introduced symbols in connection to the particular problem been solved in this paper.

Let us define a triplet $(\mathcal{G}, \mathcal{X}, \mathbf{g})$ as an instance of the MAX-SUM problem where the symbol $\mathcal{G} = \langle \mathcal{T}, \mathcal{E} \rangle$ denotes a graph consisting of a discrete set $\mathcal{T}$ of vertices and a set $\mathcal{E} \subseteq \binom{|\mathcal{T}|}{2}$ of pairs of those vertices. Each vertex $t \in \mathcal{T}$ is assigned a label $x_t \in \mathcal{X}$ where $\mathcal{X}$ is a discrete set of nodes. Let the elements $g_t(x_t)$ and $g_{tt'}(x_t, x_{t'})$ express qualities given to node x_t in vertex t and pairwise qualities on edges between nodes x_t, $x_{t'}$ between two vertices t and t', respectively. A MAX-SUM *labeling* is a mapping that assigns a single label x_t to each vertex, represented by a $|\mathcal{T}|$-tuple $\mathbf{x}$, which maximizes the following sum of unary and binary functions of discrete variables

$$\mathbf{x}^* = \operatorname*{argmax}_{\mathbf{x}\in\mathcal{X}^{|\mathcal{T}|}} \Big[\sum_t g_t(x_t) + \sum_{\{t,t'\}} g_{tt'}(x_t, x_{t'}) \Big]. \tag{1}$$

For better understanding of the symbols in the MAX-SUM formulation, we refer the reader to Fig. 1 in [16].

Recently, very efficient and fast algorithms for solving the MAX-SUM problem through linear programming relaxation and its Lagrangian dual have been reviewed in [9, 16]. Although, finding a global optimum of Eq. (1) is not guaranteed, as the problem is NP-hard, it has been shown that often the optimal solution or one very close to it can be reliably achieved.

2.2. Graph construction

In this section we describe the construction of the graph $\mathcal{G}$, introduced in the previous section, and design the functions $g_t(x_t)$ and $g_{tt'}(x_t, x_{t'})$ from Eq. (1). The graph $\mathcal{G}$ is built from the line segments belonging to two vanishing points. For a typical image of a man-made environment, where three orthogonal vanishing points are detected, three independent graphs can be constructed. The problem of the detection of rectangles separates then into searching for the rectangles in three mutually orthogonal planes.

Line segments and vanishing points. We extract line segments by polygonalisation of edges obtained by the Canny edge detector and split them at high curvature points followed by total least squares line fitting. The detection of vanishing points is carried out adopting the EM technique proposed in [10] followed by the Kanatani's renormalization-based refinement and camera calibration [8]. In principle, the proposed method does not depend on the strategy used for localization of the line segments and the vanishing points.

Graph structure. The line segments belonging to two vanishing points represent the set $\mathcal{T}$ of graph vertices. The line pairs, *i.e.* the set $\mathcal{E}$, are established from all pairs of line segments connected by the Constrained Delaunay Triangulation (CDT). An example of the constructed graph $\mathcal{G}$ is shown in Fig. 3.

Adding more pairs into the graph or using connectivity other than given by the CDT would not affect following formulations. Of course, if more pairs are established the problem becomes more complex and there is a higher chance of the MAX-SUM solver to converge to a local optimum.

Labels. We define $|\mathcal{X}| = 7$ labels. The basic labels $\{1, 2, 3, 4\}$ correspond to four single edges of a rectangle respecting the position of both vanishing points; the extra labels $\{5, 6\}$ denote shared edges between two rectangles; and $\{7\}$ is the label for "non-rectangle" line segments not consistent with any rectangle, see Fig. 2. The line segments marked by the extra labels $\{5, 6\}$ represent edges shared by two rectangles. Such lines can be considered as having two basic labels simultaneously, *i.e.* $\{5\}$ is $\{3, 4\}$, and $\{6\}$ is $\{1, 2\}$, see Fig. 2.

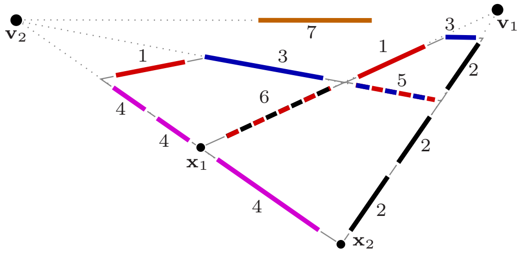


Figure 2. Visual meaning of the proposed labels.

Data term. The data term or the unary function $g_t(x_t)$ from Eq. (1) captures the membership of each line segment to a particular label. The line segments consistent with the vanishing point $\mathbf{v}_1$, resp. $\mathbf{v}_2$, may get labels $\{1, 2, 6, 7\}$, resp. $\{3, 4, 5, 7\}$, see Fig. 2. This is expressed in the following data term

$$g_t(x) = \begin{cases} [\ 0 \quad 0 \quad a \quad a \quad a \quad b \quad c\ ]^\top & \text{line } t \rightarrow \mathbf{v}_1 \\ [\ a \quad a \quad 0 \quad 0 \quad b \quad a \quad c\ ]^\top & \text{line } t \rightarrow \mathbf{v}_2. \end{cases} \tag{2}$$

The constant $a < 0$ is the cost that a line segment aligned with one vanishing point is consistent also with the second vanishing point. This double assignment is not allowed and therefore a is set to a very small number, meaning high cost. The constant b is the cost of a line segment to have shared labels $\{5, 6\}$. It is set to $a < b \leq 0$ as we allow these labels while preferring the basic ones. The number $c < 0$ is a cost for the "non-rectangle" line assignment.

Smoothness term. The smoothness term or the binary function $g_{tt'}(x_t, x_{t'})$ from Eq. (1) between each established pair of line segments $\{t, t'\}$ captures cost of all possible label combinations of that pair. Two scenarios can happen that two line segments t and t' are assigned to

1. the same vanishing point, either $\mathbf{v}_1$ or $\mathbf{v}_2$. A line passing through a particular vanishing point is fitted to the end points of two line segments t and t' in a least sum of square manner with an error ϵ. The line segments can be either parallel or one is a continuation of the other. If they are parallel, the line segment projections onto the fitted line overlap and the pair $\{t, t'\}$ is removed from the graph $\mathcal{G}$. The removal of all such pairs allows the rectangle-in-rectangle structure to be detected and makes

the optimization problem in Eq. (1) more tractable. The removal of some pairs can be seen in Fig. 3. If the pair is preserved, the following edges are created

$$g_{tt'}(i,j) = -\epsilon, \qquad \forall i,j \in \mathcal{X}\colon\ i = j. \tag{3}$$

2. two different vanishing points. Let $\mathbf{x} = \mathbf{l}_t \times \mathbf{l}_{t'}$ be an intersection point of two lines $\mathbf{l}_t$, $\mathbf{l}_{t'}$ and let d_t, resp. $d_{t'}$, be the distance between $\mathbf{x}$ and the closest end point of the line segment t, resp. t'. Denote $\gamma = d_t + d_{t'}$, then if $\gamma > const$ remove the pair $\{t, t'\}$ from the graph $\mathcal{G}$, otherwise

$$g_{tt'}(i,j) = -\gamma, \qquad \forall (i,j) \in \mathcal{A} \times \mathcal{A}'. \tag{4}$$

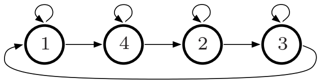
   The label sets $\mathcal{A}$ and $\mathcal{A}'$ depend on position of the point $\mathbf{x}$ w.r.t. both vanishing points and the line segments t, t'. In general, four different cases can occur, *e.g.* for the intersection point $\mathbf{x}_1$ in Fig. 2 and $\mathbf{l}_t \rightarrow \mathbf{v}_1$, $\mathbf{l}_{t'} \rightarrow \mathbf{v}_2$, $\mathcal{A} = \{1, 6\}$ and $\mathcal{A}' = \{4, 5\}$, for the point $\mathbf{x}_2$, $\mathcal{A} = \{2, 6\}$ and $\mathcal{A}' = \{4, 5\}$, and analogously for other two cases.

To properly handle the "non-rectangle" label $\{7\}$, in each pair $\{t, t'\}$ we need to create edges between that label and all other labels, *i.e.* the edges $g_{tt'}(i, 7)$, $g_{tt'}(7, i)$, $\forall i \in \mathcal{T}$ are added and set to δ. In our case, $\delta = 5\,\gamma$ as we want them to be more expensive than other edges. Not established edges $g_{tt'}(i, j)$ stand for edges set to $-\infty$ as such labels are not allowed to occur close to each other.

The smoothness term could possibly be weighted by the confidence of prolongations of two considered line segments to a possible rectangle corner, *e.g.* by checking gradients along the prolongations, or testing for the presence of Harris corners.

2.3. Parsing of rectangles

After establishing and appropriately setting the data and smoothness terms, the publicly available MAX-SUM solver[1] is used to solve Eq. (1). As a result, a unique label is assigned to each line segment, see Fig. 3. To get final rectangles a recursive parsing is performed. It starts from each line segment in the graph $\mathcal{G}$, respecting the graph connectivity, and searches for possible paths following the topological structure of the state machine



where the numbers stand for the labels. If a line segment is labeled by any of the extra labels $\{5, 6\}$, the segment is considered to have both basic labels $\{3, 4\}$, resp. $\{1, 2\}$. At this stage, we can decide for the type of rectilinear structure we want to parse from the labeled line segments. This is controlled by a required number of different labels in each path in the above state machine, *i.e.* 2 for L-shapes, 3 for U-shapes, 4 (without the final loop) for incomplete, and 4 for complete rectangles, see Fig. 5.

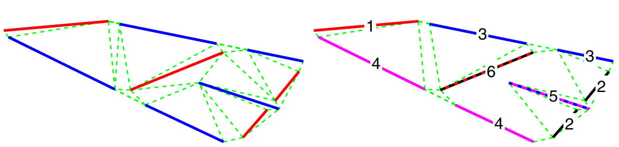


Figure 3. A toy example of the detection of rectangles. *Left:* An input image with two sets of line segments (bold), where each set is consistent with one vanishing point, connected by the CDT (dashed). *Right:* Result after the labeling. Each line segment is labeled by a unique label. Note that some of the connections were removed during the procedure.

3. Wide baseline stereo matching

Motivated by the class of matching problems mentioned in the introduction, we demonstrate how the detected quadrilaterals in two views along with a suitable descriptor offer an alternative way to find their matches. This is especially useful when point/region based matching techniques cannot reliably establish correspondences. The primary advantage is that the robustness w.r.t. viewpoint change and perspective distortion is naturally handled by a homography warping of the quadrilaterals to their canonical rectangular patches. In the next we assume that the vanishing points detected in two views separately are already mutually matched.

3.1. Region Descriptor

Each detected quadrilateral region can be warped to a $s \times s$ square patch through a homography [5]. The parameter s in pixels controls the resolution of the canonical patch (we use $s = 50$). The canonical square patch can be efficiently represented by low frequency coefficients of the Discrete Cosine Transformation (DCT)[2]. This has been shown in [14] to be an efficient and slightly superior way of obtaining a patch compared to the SIFT descriptor [11]. The main reason we adopted the DCT is that the descriptor captures low frequencies as opposed to the SIFT relying on high frequencies like edges which are often not present or weak in many rectangles.

Before computing the DCT, the patch is photometrically normalized by transforming the R, G, B color channels to have zero mean and unit variance. Then the DCT is computed on the patch for each channel separately. The descriptor is composed of the first low frequency coefficients

[1] http://cmp.felk.cvut.cz/cmp/software/maxsum/code

[2] usually hardware supported because of widespread use in image compression (JPEG, MPEG, *etc.*)

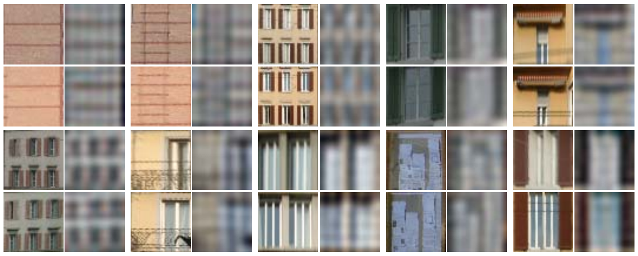


Figure 4. Examples of matched rectangles mapped to 50×50 square patches and their reconstructions from low pass DCT coefficients.

(reverted diagonals) as proposed in [14]. Moreover, three additional numbers are stored, two for the chromacity vector computed over three color channels and one for the rectangle height/width ratio α explained in the next section.

3.2. Single-view geometry

Given the detected quadrilaterals and assuming that we are viewing rectangles we can estimate their dimensions and their relative 3D poses. The basic element of the pose estimation is the homography matrix $\mathtt{H}_{0i} \in \Re^{3\times3}$ mapping points $\mathbf{x}_i$ from i-th view of the quadrilateral to the points $\mathbf{x}_\pi$ of its canonical rectangle, *i.e.*

$$\mathbf{x}_i \simeq \mathtt{K}[\mathtt{R} \mid \mathbf{t}]\,\mathbf{X} = \mathtt{K}[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}]\,\mathbf{x}_\pi = \mathtt{H}_{0i}\,\mathbf{x}_\pi. \tag{5}$$

The rectangle is assumed to lie on a plane π, its points are $\mathbf{X} = [X\ Y\ 0\ 1]^\top$ as the origin of the world coordinate system is in the plane and the z axis is perpendicular to that plane. $\mathtt{K}_{3\times3}$ is a known calibration matrix, $\mathtt{R} = [\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3]$ is a rotation matrix, and $\mathbf{t}$ is a translation vector. For more detail see [5].

Let us consider a canonical rectangular region $\mathcal{S}$ specified by 4 corner points stacked as columns in the matrix

$$\mathtt{S}_\pi = \begin{bmatrix} 0 & 0 & \alpha\, s & \alpha\, s \\ 0 & s & s & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \tag{6}$$

where s is the height of that rectangle in 3D and α is an ratio between its height and width. It has been shown in [10] that one can estimate the unknown parameters α, s, and 3D position $\mathtt{R}$, $\mathbf{t}$ (up to scale) for each quadrilateral. The parameter α is part of the descriptor vector and the 3D pose is later used in the plane sweeping algorithm.

3.3. Matching

Given the descriptor vectors of all detected quadrilaterals in two views, we establish their tentative matches. The matching score is computed as a Euclidean distance on those matches passing a pre-test on having a similar ratio α and chromacity vector. For each quadrilateral the k-nearest matches are stored (we use $k = 3$). We assume one dominant plane in the scene and therefore the tentative matching is followed by a homography-based inlier selection. The quadrilaterals bring significant advantage over interest point matching as only *one* match is needed to compute the homography matrix in contrast to *four* points in the general case. Standard RANSAC-based estimation can be avoided and a simple exhaustive search can be employed. To find more matches on the dominant plane the estimated homography is employed in the guided matching respecting both the Sampson error [5] and the descriptor distance.

3.4. Plane sweeping

Once the dominant plane with supporting quadrilateral matches is found, one can search for additional matches on parallel planes by the sweeping strategy proposed in the following.

First, we estimate the position of the dominant plane and both cameras. The estimated rotation matrices, Eq. (5), of all matched quadrilaterals on a dominant plane in each view must be the same. Therefore, we compute one mean rotation matrix for each view, $\hat{\mathtt{R}} = [\hat{\mathbf{r}}_1\ \hat{\mathbf{r}}_2\ \hat{\mathbf{r}}_3]$ and $\hat{\mathtt{R}}'$, as a mean over all rotations $\mathtt{R}$, $\mathtt{R}'$ of the matched quadrilaterals, projected back to the manifold of rotation matrices ($\hat{\mathtt{R}} = \mathtt{U}\,\mathtt{V}^\top$, where $[\mathtt{U}\ \mathtt{D}\ \mathtt{V}] = \mathrm{svd}(\bar{\mathtt{R}})$ [5]). Prime symbols stand for estimates in the second view. The translation vectors $\hat{\mathbf{t}}$, $\hat{\mathbf{t}}'$ are taken from the match giving the smallest Sampson error.

Second, the sweeping is done by sliding a plane, parallel to the dominant plane, in 3D forward and backwards in h increments along the z-axis. At each height a shifted homography is computed and quadrilaterals consistent (small descriptor distance) with that homography are established as new matches. The composite shifted homography which maps a point from the first image to the second, assuming that the corresponding 3D point lies on a parallel plane to the dominant one, can be computed as

$$\mathtt{H}^h_{12} = \mathtt{H}^h_{02}\,(\mathtt{H}^h_{01})^{-1}, \text{where} \tag{7}$$
$$\mathtt{H}^h_{01} = \mathtt{K}[\hat{\mathbf{r}}_1\ \hat{\mathbf{r}}_2\ h\hat{\mathbf{r}}_3 + \hat{\mathbf{t}}], \quad \mathtt{H}^h_{02} = \mathtt{K}'[\hat{\mathbf{r}}'_1\ \hat{\mathbf{r}}'_2\ h\hat{\mathbf{r}}'_3 + \hat{\mathbf{t}}']$$

are shifted homographies mapping points from the shifted canonical plane to image planes. The plane sweeping declares more matches which otherwise would not be found using only appearance based matching and allows direct 3D reconstruction of the parallel planes. It is especially useful for images with repetitive structures containing parallel planes, see Fig. 8.

4. Experiments

An example of the detection of full rectangles, incomplete rectangles, and U-shapes can be seen in Fig. 5. Each U-shape is closed to two rectangles by completing the

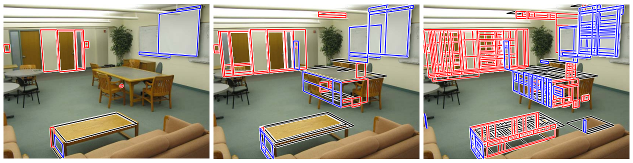


Figure 5. Types of rectilinear structures. *Left:* Full rectangles. Middle: Incomplete rectangles. *Right:* U-shapes completed to rectangles.

fourth missing edge by a line passing through one of the U-shape end points and a corresponding vanishing point. In our remaining experiments we use the closed U-shapes providing more hypotheses for subsequent stages. Examples of extracted U-shapes in indoor images are depicted in Fig. 1 and Fig. 7. In the rest of the paper we show only matched rectangles to avoid too cluttered images.

We evaluated the proposed method for the detection of quadrilaterals followed by two-view matching on a large variety of images. Some representative ones are shown in Fig. 6. The first four are from the ZuBuD[3] database consisting of 640×480 images, the last 1126×844 image was taken by ourselves. Overall, their quality varies since they were taken by different, to us unknown, cameras under different illumination conditions. Despite that and moreover suffering from considerable wide baseline, light reflections, shadows, jpg-artifacts, occlusions, and repetitive structures the matching results show feasible and stable performance. For descriptor vector we used 10 reverted diagonals from the DCT on a 50×50 square resulting in 66 numbers for each color channel.

The plane sweeping is demonstrated in Fig. 8. Once a dominant plane with supporting quadrilateral matches was found, the plane sweeping from Sec. 3.4, was performed and new matches on different but parallel planes were established, see Fig. 8. The advantage of the sweeping is evident as two additional planes where just few rectangles are present were still found. Standard sequential search for homography consistent sets of inliers [5], where at each stage already found inliers are dropped and a new plane is searched for, was not able to find the few matches out of the dominant plane. Due to large spatial support of the quadrilaterals the accuracy of their localization is not so crucial compared to other salient point detectors. They are often followed by a noise sensitive point triangulation to obtain a 3D reconstruction. I our case, we already have the 3D poses of the matched quadrilaterals, parsed from the homographies in Eq. (7), and therefore the 3D model can directly be built, shown in Fig. 1.

The last briefly mentioned application is the use of rectangles as support regions for the computation of geometric context from a single image. In this case the goal is to label individual superpixels as belonging into one of the three orthogonal planes. The quadrilaterals impose constrains on overlapped superpixels and significantly help to increase the stability of this ill-posed problem [12]. An example with comparison to [7] can be seen in Fig. 7 and in Fig. 1.

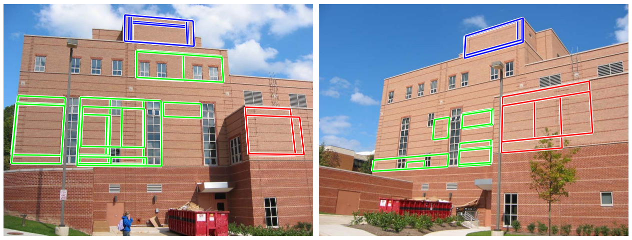


Figure 8. The plane sweeping. Each image is matched with the image from Fig. 1. The resulting matched rectangles are shown in different color associated to the plane they belong to.

From computational complexity point of view the presented method for detecting rectangles is comprised of efficient steps, *i.e.* searching for lines, vanishing points, computing the CDT, running the MAX-SUM solver, and recursive parsing of the rectangles. The running time of the first four stages is under 5 secs, the last stage takes from few to tens of secs on a modest notebook using Matlab / C implementation. However, it could be speeded up significantly.

5. Conclusions

We have presented a method for the detection of perspectively distorted rectangles, exploiting solely geometric cues. Compared to previously published work, our method is computationally efficient and we can effectively detect partially occluded structures, such as U-shapes,and L-shapes, and rectangle-in-rectangle configurations. Furthermore, we have demonstrated how the rectangles can serve as features of intermediate complexity, in the context of wide baseline matching tasks. The capability of detected regions of large image support is advantageous, as they are less ambiguous compared to single points and can be matched more reliably.

The approach has been verified by extensive experiments in indoor and outdoor environments presenting results of both the detection and the matching stages. The quality of the detected structures largely depends on the quality of the detected line segments which could be improved by using more elaborated line detection techniques.

References

[1] H. Bay, A. Ess, A. Neubeck, and L. Van Gool. 3D from line segments in two poorly-textured, uncalibrated images. In *Proc. Int. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, June 2006.

[2] R. Collins. A space-sweep approach to true multi-image matching. In *Proc. of CVPR*, pages 358–363, 1996.

[3] http://www.vision.ee.ethz.ch/showroom/zubud

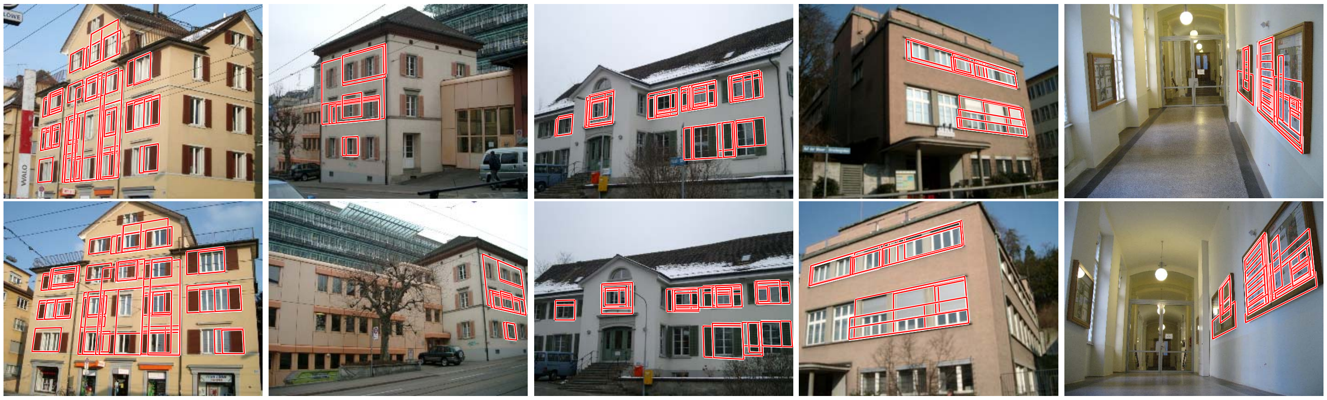


Figure 6. WBS matching examples. Each column shows two stereo images with matched rectangles. From left: *i)* `object0039` - 79 matches (645 / 686), *ii)* `object0008` - 10 matches (345 / 257), *iii)* `object0084` - 59 matches (362 / 318), *iv)* `object0184` - 29 matches (540 / 478), *v)* `00439-00441` - 59 matches (506 / 396). The numbers in the brackets express total number of detected rectangles in the first and the second image respectively.

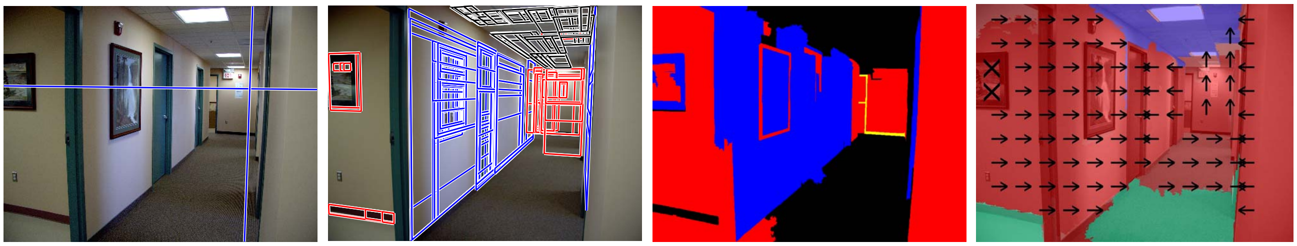


Figure 7. An example of the segmentation of orthogonal planes in a single indoor image using the extracted rectangles. From left: *i)* Input image with in-plotted ideal lines. *ii)* The detected quadrilaterals by the proposed method. *iii)* An MRF based method [12] utilizing the quadrilaterals and segmenting an image into three orthogonal planes. Each plane is depicted in a different color, where yellow color denotes "undecided" pixels. *iv)* The method by Hoiem *et al*. [7] segmenting images into ground plane and vertical planes. Arrows stand for plane orientations to the left/up/right, markers 'o' and 'x' for porous and solid materials respectively. Notice better result of our method.

[3] D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys. Real-time plane-sweeping stereo with multiple sweeping directions. In *Proc. of CVPR*, 2007.

[4] F. Han and S. Zhu. Bottom-up/top-down image parsing by attribute graph grammar. In *Proc. of ICCV*, 2005.

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[6] J. Hayet, F. Lerasle, and M. Devy. Visual landmarks detection and recognition for mobile robot navigation. In *Proc. of CVPR*, pages II: 313–318, 2003.

[7] D. Hoiem, A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 75(1):151–172, 2007.

[8] K. Kanatani and Y. Sugaya. Statistical optimization for 3D reconstruction from a single view. *IEICE Trans. on Information and Systems*, E88-D(10):2260–2268, 2005.

[9] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.

[10] J. Košecká and W. Zhang. Extraction, matching and pose recovery based on dominant rectangular structures. *CVIU*, 100(3):174–293, 2005.

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[12] B. Mičušík, H. Wildenauer, and M. Vincze. Towards detection of orthogonal planes in monocular images of indoor enviroments. In *Proc. of Int. Conf. on Robotics and Automation (ICRA)*, 2007.

[13] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[14] Š. Obdržálek and J. Matas. Object recognition using local affine frames on maximally stable extremal regions. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *LNCS*, pages 83–104. Springer, 2006.

[15] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. of ICCV*, pages 767–774, 1998.

[16] T. Werner. A linear programming approach to Max-sum problem: A review. *PAMI*, 29(7):1165–1179, 2007.

[17] T. Werner and A. Zisserman. New techniques for automated reconstruction from photographs. In *Proc. of ECCV*, pages 541 – 555, 2002.

[18] A. Yang, K. Huang, S. Rao, and Y. Ma. Symmetry-based 3D reconstruction from perspective images. *CVIU*, 99:210–240, 2005.