Prior-based facade rectification for AR in urban environment

Antoine Fond, Marie-Odile Berger, Gilles Simon

▶ **To cite this version:**

Antoine Fond, Marie-Odile Berger, Gilles Simon. Prior-based facade rectification for AR in urban environment. ISMAR workshop on Urban Augmented Reality, Sep 2015, Fukuoka, Japan. hal-01235842

HAL Id: hal-01235842
https://hal.science/hal-01235842

Submitted on 30 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prior-based facade rectification for AR in urban environment

Antoine Fond* Marie-Odile Berger† Gilles Simon‡

Université de Lorraine, Inria, LORIA

Abstract

We present a method for automatic facade rectification and detection in the Manhattan world scenario. A Bayesian inference approach is proposed to recover the Manhattan directions in camera coordinate system, based on a prior we derived from the analysis of urban datasets. In addition, a SVM-based procedure is used to identify right-angle corners in the rectified images. These corners are clustered in facade regions using a greedy rectangular min-cut technique. Experiments on a standard dataset show that our algorithm performs better or as well as state-of-the-art techniques while being much faster.

Index Terms: I.2.10 [Vision and Scene Understanding]: 3D/stereo scene analysis—; H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities—;

1 Introduction

In Augmented Reality, accurate pose computation is fundamental for seamless integration of virtual objects into the real scene. We are interested in applications which take place in man-made environments and we suppose that the camera intrinsic parameters are available. We focus in this paper on the initialization stage which is especially difficult in urban scenes due to the presence of repeated patterns. Another difficulty originates in the fact that a pedestrian is free of his motion in the scene and can therefore adopt uncontrolled viewpoints - close or distant views - with respect to the model (see Fig. 7 for various examples of images). As a result, the set of 2D/3D correspondence hypotheses may contain a high ratio of outliers which may lead to erroneous pose computation.

In this paper, we invoke the so-called "Manhattan world" assumption, which states that groups of lines are aligned with the cardinal axes of a global frame. Past works have investigated rectification based on the detection of orthogonal vanishing points (VPs) to facilitate wide-baseline matching and reconstruction[20, 14, 3]. Such methods allow to cope with the limitations of affine invariant descriptors which are unable to match points when large projective deformations occur. However, identifying areas in correspondence after this rectification step can still be difficult. In the context of extracting dominant rectangular structures, [14] rely heavily on the the strong assumption that the boundaries or the corners of the rectangle can be extracted. With the goal to match street-level facades to airborne images, [3] propose a descriptor that captures the structure of repetition of patterns and attempt to characterize facades by clustering these descriptors. Preliminary results are promising but manual marking of buildings is required to initialize the clustering.

In the continuity of these past works, we investigate how facade rectification and delimitation can be improved by considering prior information about the scene and the camera relevant to AR in urban context. Note that our goal is not to identify accurately the facades but only to provide regions of interest where facades features are likely to be found, in order to improve the robustness and speed of subsequent recognition tasks. Fig. 1 illustrates the interest of such a strategy by using a very common matching procedure between a roughly textured polyhedral model of the Hôtel de Ville of Nancy and a picture of this building. Rectifying the image and using the top-ranked rectangle provided by our algorithm (see Fig. 5) leads to a significant increase in both the number and ratio of inliers.

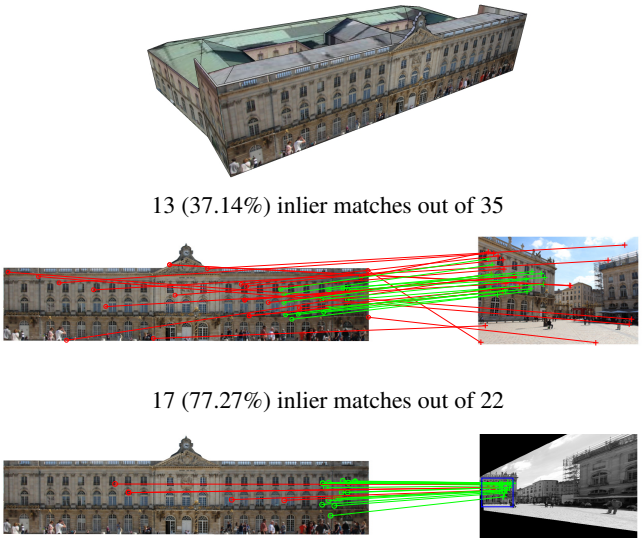

Figure 1: RANSAC-based matching of SIFT features between a textured 3D model of the Hôtel de Ville of Nancy and a photo of this building, before and after having rectified and detected the facade.

Our contribution are twofold. First we provide a Bayesian framework for detecting VPs in Manhattan worlds which incorporate prior about the Manhattan frame by imposing a near-vertical direction as well as orthogonality constraints. Second, we propose to use machine learning and cutting graph techniques to formulate facade hypotheses which will be used subsequently to guide the matching between the model and the considered image. Instead of attempting to detect repetitive patterns in the image as in [3], we propose to detect right-angle corners due to windows or doors using a SVM-based machine learning technique. Rectangular facade hypotheses are then generated through min-cuts techniques with the idea to identify rectangles with high density of right-angle corners.

The paper is organized as follows. Related work about orthogonal VP detection is described in section 2. The prior distribution is provided in section 3. Our Bayesian framework for VP detection is described in section 4 and the facade detection algorithm is presented in section 5. Extensive comparisons of our method with state of the art techniques [19, 15] are presented in section 6 along with some results of facade detection.

*e-mail: antoine.fond@loria.fr
†e-mail: marie-odile.berger@inria.fr
‡e-mail: gilles.simon@loria.fr

2 Related Work

There is a vast literature on the problem of VP detection. Early methods used the Hough transform (HT) to detect VPs on the Gaussian sphere [16]. However, such approaches are sensitive to the quantization level of the bins and can produce false VP. Some methods use HT as an initialization stage and Expectation Maximization (EM) iterations to get more accurate and confident results [2, 13, 14]. EM performs both classification and estimation tasks by iterating between two steps. However, a reasonable initialization is required and the number of models in the mixture formulation has to be fixed, which does not guarantee that the Manhattan directions are finally obtained. Several attempts have been made to tackle these problems. For instance, [19] estimate VP hypotheses in the image plane using pairs of edges and compute consensus sets using the J-linkage algorithm. In [15], the problem is solved in the dual domain where converging lines become aligned points. The use of a robust point alignment detector leads to candidate VPs. Both [19] and [15] provide a RANSAC-like procedure to find the three Manhattan directions once the set of candidate VPs has been obtained, assuming the internal camera parameters are known. However, these procedures do not enforce orthogonality between the Manhattan VPs, and fail if one of the Manhattan VPs is missing in the candidate set.

Another category of techniques directly estimate the Manhattan directions (or, equivalently, the camera orientation) from image data. In [21], a minimal solution for computing three orthogonal VPs and focal length from four line segments is used to maximize a consensus set using RANSAC. In [4], the number of clustered lines is globally maximized over the rotation search space, using a branch-and-bound procedure based on the Interval Analysis theory. This kind of techniques may be optimal in general case, but improved performance in terms of efficiency and robustness may be obtained when some prior information is available. Our method is thus more in line with some works such as [6, 7, 9], where the Manhattan directions are estimated using Bayesian inference. In the early work of Coughlan and Yuille [6], the camera is assumed oriented in the horizontal plane. A posterior distribution on the compass direction is derived at each pixel by combining knowledge of the geometry of the Manhattan world with statistical knowledge of edges in images. The image data at each pixel is explained by one of five models: edge due to one of the three orthogonal VPs, random edge or off-edge. The prior probability of each of the edge models was estimated empirically. The maximum a posteriori (MAP) estimate is obtained by evaluating the log posterior for the compass direction in the range -45° to $+45^\circ$, in increments of 1°. The horizontal camera orientation assumption is relaxed in [7, 9], though at the expense of high combinatorial search over discretized Euler angles in [7], or Quasi-Newton or EM optimizations which both require reasonable initial guesses in [9].

In this work, we use a prior on the distribution of the Manhattan directions, that was derived from real data. Such a distribution has, to our knowledge, never been provided before and is in itself a contribution of this paper. Moreover, in order to reduce the complexity of MAP estimation, we divided the problem into three steps: in the first step, our prior is used to provide posterior probabilities of VPs sampled on the Gaussian sphere. In the second step, local maxima of these probabilities are extracted using a spherical weighted mean shift. Finally, the Manhattan frame is obtained by solving the MAP among a discrete set of candidate VP triplets.

3 Manhattan Frame Prior Distribution

A histogram of 648 ground truth Manhattan directions obtained from the York Urban Line Segment [9] (102 images) and the Toulouse Vanishing Points [1] (114 images) datasets is shown in Fig. 2(left). The VPs are expressed on the Gaussian sphere S_2, where x, y, z represent, respectively, the horizontal axis, the vertical axis and the principal axis of the camera. An examination of this figure leads to the following observations:

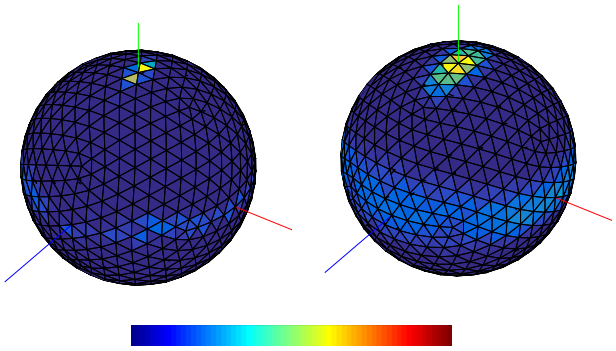

Figure 2: Histograms of VPs on the Gaussian sphere, extracted from the York Urban and Toulouse datasets (left) and sampled from our prior distribution p_Y (right). x, y, z camera axes are colored, respectively, in red, green, blue. The histogram values are colored using the Matlab Jet colormap shown at the bottom of the figure. The same color conventions are used in all figures of this paper.

Observation 1 The vertical Manhattan directions Y are nearly vertical in the camera frame and mainly constrained in the $y-z$ plane: this reflects that camera rotations around x-axis (pitch) are often performed, while the x-axis keeps horizontal in the Manhattan frame (roll angles are generally very small). The pitch angle has a limited range and is centered around 0.

Observation 2 Due to the orthogonality between the Manhattan VPs, a consequence of Observation 1 is that the horizontal directions X and Z are concentrated in a narrow range around the equator. Moreover, we observe that these directions are distributed all around the sphere.

According to Observation 1, we use the Kent distribution [12] to model the prior distribution of the vertical Manhattan direction:

$$p_Y(Y) = \frac{1}{c(\kappa_Y, \beta)} \exp\left(\kappa_Y y^T Y + \beta\left((z^T Y)^2 - (x^T Y)^2\right)\right), \quad (1)$$

where $\kappa_Y > 0$ determines the spread of the distribution, β determines the ellipticity of the contours of equal probability and $c(\kappa_Y, \beta)$ is a normalizing constant. The parameter β is set to $\frac{2}{5}\kappa_Y$ so that the major axis of the confidence ellipses is aligned with the principal axis of the camera (Fig. 3(left)).

Knowing the Y-direction and considering Observation 2, the X-direction can be obtained using a Watson distribution [17]:

$$p_{X|Y}(X, Y) = \frac{1}{M(\frac{1}{2}, \frac{3}{2}, -\kappa_X)} \exp\left(-\kappa_X (Y^T X)^2\right), \quad (2)$$

where the normalizing constant M is the Kummer function (Fig. 3, middle).

The third Manhattan direction Z is likely to set near the cross product of directions X and Y, leading to the von-Mises-Fisher distribution [17] (Fig. 3(right)):

$$p_{Z|X,Y}(X, Y, Z) = \frac{\kappa_Z}{4\pi \sinh \kappa_Z} \exp\left(\kappa_Z (X \times Y)^T Z\right). \quad (3)$$

Finally, the joint probability of a triplet X, Y, Z can be inferred from equations (1) to (3):

$$p_{X,Y,Z}(X, Y, Z) = p_{Z|X,Y}(X, Y, Z) p_{X|Y}(X, Y) p_Y(Y). \quad (4)$$

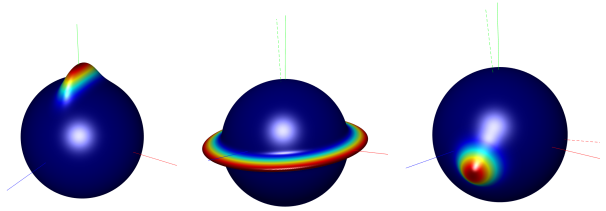

Figure 3: Prior / conditional probability distributions of the Manhattan directions in the camera frame. From left to right: p_Y, $p_{X|Y}(Y_0)$ and $p_{Z|X,Y}(X_0,Y_0)$. X_0, Y_0 are shown in dashed lines. $\kappa_Y = 50, \kappa_X = \kappa_Z = 30$.

4 Bayesian estimation of the Manhattan VPs

Most VP estimation algorithms rely on segment lines extraction in the image. Now that we have a prior distribution for the Manhattan directions, we could define a likelihood using the line segments L as measures, and solve a MAP $p_{L|X,Y,Z}p_{X,Y,Z}$ in $(S_2)^3$. However, the high dimensionality of the prior would render this method computationally infeasible. In order to simplify the problem, we first estimate the local maxima in S_2 of the posterior distribution $p_{L|V}p_V$, defined for any VP, using a spherical weighted mean shift. Then the local maxima are considered as candidate Manhattan directions and the MAP is solved on a discrete set of VP triplets.

4.1 Computation of candidate VPs

Line segments are detected in the image plane using LSD [10] and divided into equal-length segments. When a set of line segments l_i are converging to the same VP in the image plane, the normal vectors n_i of their great circle on S_2 are laying in the same plane. The normal vector of that plane is the VP direction V. We thus can define the likelihood $p_{L|V}$ as:

$$p_{L|V}(L,V) = \frac{1}{C}\sum_{l_i \in L} \exp\left(-\frac{(n_i^T V)^2}{2\sigma^2}\right), \tag{5}$$

where C is a normalizing term. A VP on S_2 can be one of the three Manhattan VPs or a non-Manhattan VP generated by the background structure. The prior probability of a VP V can therefore be seeing as a mixture from all four causes:

$$p_V(V) = \pi_X p_X(V) + \pi_Y p_Y(V) + \pi_Z p_Z(V) + \pi_N p_N(V), \tag{6}$$

where $p_X(V)$, $p_Y(V)$ and $p_Z(V)$ are the marginal probabilities of the Manhattan frame prior distribution $p_{X,Y,Z}$ defined in equation (4) and $p_N(V)$ is a probability distribution on S_2 that models the non-Manhattan VPs. Following e.g. [6] and [9], we take $\pi_X = \pi_Y = \pi_Z = \frac{1-\pi_N}{3}$. In images where the Manhattan world assumption is valid, non-Manhattan VPs are due to extraneous structures such as striped awnings, rows of posts, etc., which are generally much rarer than building structures. For that reason, we used $\pi_N = 0$ in our implementation. Fig. 2(right) shows a histogram of VPs sampled from our prior distribution (6): as we can see, this histogram is close to the one generated from ground-truth data (Fig. 2(left)) though a bit more spread out, which allows us to handle a slightly larger variability of VPs than the one obtained in the datasets.

To find the local maxima of the posterior distribution $p_{L|V}p_V$ we use a spherical weighted mean shift. V is sampled from the prior distribution p_V and P seeds are selected from that sampling. For each seed V_j we apply a mean shift on the sphere S_2 over the previous sampling weighted by the likelihood. In a certain neighborhood $N_\varepsilon(V_j)$ we compute the weighted Karcher Mean μ_j on S_2 using the Newton-like algorithm from [5] in the following minimization:

$$\mu_j = \underset{\mu}{\operatorname{argmax}} \sum_{V \in N_\varepsilon(V_j)} w_V d_g(\mu, V) \tag{7}$$

with d_g the geodesic distance on the sphere and the likelihood weights w_V

$$w_V = \frac{p_{L|V}(L,V)}{\sum_{X \in N_\varepsilon(V_j)} p_{L|V}(L,X)}. \tag{8}$$

If the distance $d_g(\mu_j, V_j)$ is not too small, V_j becomes μ_j and we repeat the procedure until convergence. Mean shift has been proven to perform a gradient ascent. Thus at the end of the mean shift we get P maxima of the posterior distribution which are our candidate VPs $\mathscr{V} = \{V_j\}_{1 \le j \le P}$ (Fig. 4).

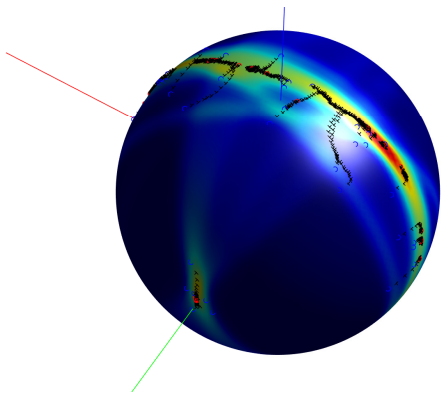

Figure 4: Mean shift paths obtained with the image of the Hôtel de Ville of Nancy: blue circles show the seeds, black crosses the steps and red circles the convergence points.

4.2 Discrete resolution of the MAP

We now only have to find the MAP estimate over the discrete set $\mathscr{V}$ of guesses:

$$\max_{(X,Y,Z) \in \mathscr{V}^3} p_{L|X,Y,Z}(X,Y,Z) p_{X,Y,Z}(X,Y,Z), \tag{9}$$

where the prior $p_{X,Y,Z}(X,Y,Z)$ is given in equation (4) and the likelihood $p_{L|X,Y,Z}(X,Y,Z)$ is obtained using the independence of the line segments $l_i \in L$:

$$p_{L|X,Y,Z} = \prod_{l_i \in L} p_{n_i|X,Y,Z} \tag{10}$$

$$= \prod_{l_i \in L} \left(\pi_X p_{n_i|X} + \pi_Y p_{n_i|Y} + \pi_Z p_{n_i|Z}\right) \tag{11}$$

where $p_{n_i|V}(n_i,V) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(n_i^T V)^2}{2\sigma^2})$.

Triplets of VPs are selected in $\mathscr{V}$ and the one maximizing the posterior probability $p_{X,Y,Z|L} \propto p_{L|X,Y,Z} p_{X,Y,Z}$ is considered to be the estimate $\tilde{X}, \tilde{Y}, \tilde{Z}$ of the Manhattan frame in camera coordinate system. In order to both reduce the combinatorial complexity of the search and favor orthogonal triplets, we proceed as follow: first, we select VPs from $\mathscr{V}$ that are inside a confidence region of the Kent distribution $p_Y(Y)$ (1). These VPs are guesses for the vertical Manhattan direction Y. Then, for each guess Y_i, all candidate VPs $\{X_j\}$ inside a confidence region of $p_{X|Y}(X,Y_i)$ (2) are selected. Finally, for each guess X_j, candidate VPs $\{Z_k\}$ inside a confidence

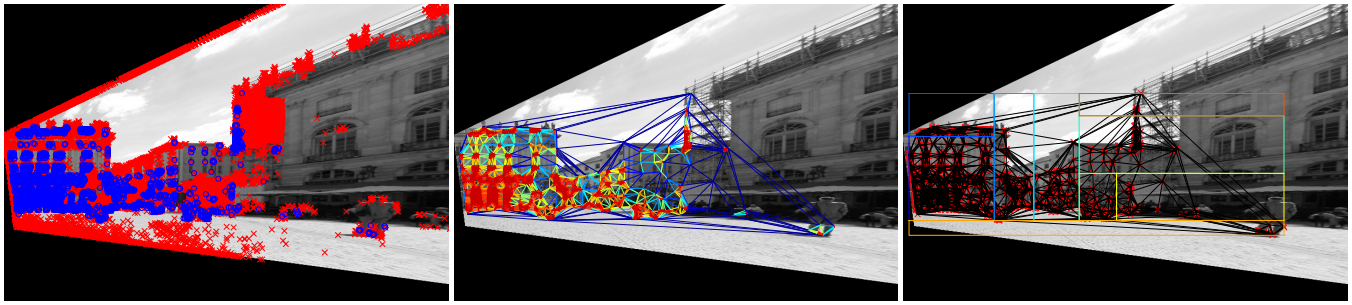

Figure 5: Main steps of our facade detection algorithm illustrated with the rectified image of the Hôtel de Ville of Nancy. From left to right: Corners classification (right-angle corners are in blue, non-right-angle corners in red). Delaunay triangulation of the right-angle corners (weights on edges are mapped to the Jet colormap). Greedy rectangular min-cut (ranks of the rectangles are mapped to the Jet colormap).

region of $p_{Z|X,Y}(X_j,Y_i,Z)$ (3) are selected and the posterior probability is assessed for all triplets $\{X_j,Y_i,Z_k\}$. Note that specific triplets $X_j,Y_i,X_j\times Y_i$ are added to the set of assessed triplets, which allows us to handle images where only one horizontal Manhattan direction is represented.

At the end of this procedure, the 3x3 matrix $(\tilde{X}|\tilde{Y}|\tilde{Z})$ is generally not in SO(3), which may produce visually poor results in the rectified image and compromise the detection of the right-angle corners whose algorithm is presented in the next section. For that reason, we eventually perform an iterative optimization of the expectation of the log-likelihood:

$$\max_{R=(X|Y|Z)\in SO(3)} E\left(\log p_{L|X,Y,Z}(X,Y,Z)\right). \tag{12}$$

This procedure is initialized with the rotation matrix $R_0=UV^T$, where $U\Sigma V^T$ is the SVD of $(\tilde{X}|\tilde{Y}|\tilde{Z})$. The cost function is parameterized with Euler angles and a quasi-Newton method is used to perform the optimization. As R_0 is generally close to the solution, the convergence is very fast. Finally, using the intrinsic camera projection matrix K we can compute the homographies $H_1=K(X|Y|Z)K^{-1}$ and $H_2=K(Z|Y|-X)K^{-1}$ which rectify the building facades aligned with resp. (X,Y) and (Z,Y) planes.

5 Facade detection

To detect a coarse bounding box of the rectified facade we rely on the fact that most facades are composed of right-angle architectural features. Doors, windows, bricks, etc., share strong vertical and horizontal components on their visual appearance. As it is difficult to precisely quantify that vertical and horizontal edge distribution for a right-angle feature, we learned the appearance of such features using supervised classification.

To that purpose, a training set was built as follows. First, a set of images coming from the York Urban database were rectified using the ground truth VPs. Then, corners where detected in these images using Shi & Tomasi algorithm [18]. Histograms of Oriented Gradient of 16 bins were used as descriptor of these corners. These descriptors were computed by locally summing the gradient values for a certain orientation rather than counting the edges [8]. A manual labeling step enabled supervised classification between right-angle and non-right-angle corners using SVM classification. About 5000 corners were labeled in almost equal proportion. SVM performed fast and accurate classification, with a rate of 86% of good classification on cross validation.

This classifier can be used to extract right-angle corners from rectified facades (Fig. 5(left)). As facades of interest often appear as rectangles in the rectified images, we want to enforce this geometrical constraint in the clustering process. Therefore, we need a measure to evaluate the clustering relevance of a rectangle over the right-angle corners. For that purpose, we first perform a Delaunay triangulation of the right-angle corners. That triangulation embeds a graph structure which enables us to use min-cut cost as a clustering measure. The weights $w_{i,j}$ of the edges $e_{i,j}$ are function of the distance between corners C_i and C_j (Fig. 5(middle)):

$$w_{i,j}=\exp\left(\frac{-\|C_i-C_j\|^2}{\sigma^2}\right) \tag{13}$$

The choice of the Delaunay triangulation is motivated by the speed of computation and the regularity of the faces generated from regular data. To find the best rectangle partition we start from the bounding box of the triangulation and we split it recursively using a greedy approach based on the min-cut cost. The cost of a split $S(R,x)$ cutting a rectangle R through axis x into two subrectangles R_X and $R_{\overline{X}}$ relies on the edges cut and the density of edges in the subrectangles.

$$S(R,x)=\frac{\sum_{e_{i,j}\in cut(R_X,R_{\overline{X}})} w_{i,j}}{\sum_{e_{i,j}\in R_X} w_{i,j}}+\frac{\sum_{e_{i,j}\in cut(R_X,R_{\overline{X}})} w_{i,j}}{\sum_{e_{i,j}\in R_{\overline{X}}} w_{i,j}} \tag{14}$$

The idea is to scan the vertical axis x and the horizontal axis y to find the minimum cost $\min\left(\min_x S(R,x),\min_y S(R,y)\right)$ where to split the rectangle R. Then we repeat that procedure recursively on the two subrectangles R_X and $R_{\overline{X}}$ until the min-cut cost $\mathscr{S}$ of the whole partition $\{R_k\}_{1\le k\le P}$ is small enough, with

$$\mathscr{S}=\sum_{k=1}^{P} S(R_k)=\sum_{k=1}^{P}\frac{\sum_{e_{i,j}\in cut(R_k,\overline{R_k})} w_{i,j}}{\sum_{e_{i,j}\in R_k} w_{i,j}}. \tag{15}$$

Finally rectangles R_k are set to the bounding box of the corners inside and ranked with respect to their partition score $S(R_k)$ (Fig. 5(right)).

6 Experiments

In this section, we present experimental results of our method on real data. We first compare our algorithm for detecting the Manhattan VPs with two state-of-the-art methods:

- Tardif's method for detecting VPs with J-Linkage and LSD [19], using the Matlab implementation available at [1],
- Lezama et al.'s method for finding VPs via points alignment in primal and dual domains [15], using code available at [2].

[1] https://code.google.com/p/vpdetection/

[2] http://dev.ipol.im/~jlezama/vanishing_points/

We ran the three algorithms on the York Urban Database [9]. This database is composed of 102 images of indoor and outdoor urban environments. In order not to bias the results, we used the same set of line segments extracted with LSD [10] for each of the three methods. In most of the images the 3 Manhattan directions are visible but in some images there are only 2. The three methods are compared using two different metrics for measuring the distance between the expected VPs and the ground-truth VPs. The first metric $\mathscr{M}_1$ is the average of the geodesic distance on the sphere from each ground truth direction to the expected direction. The second metric $\mathscr{M}_2$ is the geodesic distance on $SO(3)$ [11]. As $\mathscr{M}_2$ is measured in the rotation manifold, it also embeds a measure of the orthogonality of the solution. The cumulative error histograms obtained for these two metrics are shown in Fig. 6.

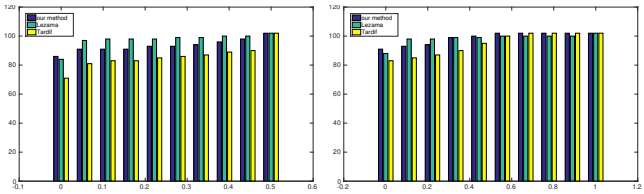

Figure 6: Cumulative error histograms for the York Urban DB using $\mathscr{M}_1$ metric (left) and $\mathscr{M}_2$ metric (right).

Our prior information on the VPs regularizes the data. So even when the set of line segments is noisy it can help the algorithm not to fall into a meaningless local maximum and find the correct solution. That allows our algorithm to be more accurate than Tardif's method for both metrics. Lezama algorithm obtains the best results when the $\mathscr{M}_1$ metric is used. However, their results are very similar to ours when using $\mathscr{M}_2$. This is due to the fact that a triplet resulting in a small error with $\mathscr{M}_1$ can have much higher value with $\mathscr{M}_2$. Each direction can be close to the ground truth direction but keep away from the orthogonality of the triplet. As our main purpose is image rectification, it is important that the solution remains on $SO(3)$. In our experiments, we noticed that an image looks visually rectified roughly up to an error of 0.1 in $\mathscr{M}_2$ metric. We get $91/102$ images under that threshold against $88/102$ for Lezama and $83/102$ for Tardif.

While in terms of accuracy our algorithm does not significantly improve Lezama's results, it is much faster (Tab. 1). This is mainly due to the fact that we reduced both dimensionality and search space (due to the prior) in our MAP solving. In our method the execution time is dominated by the spherical mean-shift step. We implemented this part of our Matlab code in C. However, it is important to notice that both Lezama and Tardif implementations also use C code with Matlab wrapping for the time-consuming parts of their algorithms.

Methods	Lezama	Tardif	our
Mean time in seconds	11.30	2.66	1.49

Table 1: Comparison of the mean time per image on a Intel Xeon W3565 Quadcore 3.2 Ghz with 8Go RAM

Results of our HOG+SVM corner classifier are shown on Fig.7. For a building where both facades are visible, the non-rectified one is almost completely devoid of right-angle corners. Rectangle clustering generally fits a coarse bounding box of the facade. However, it can be noticed that spurious right-angle corners are often detected around the horizon line. This is not surprising, as any horizontal line in the scene at the height h of the camera is projected to the horizon line. As all vertical lines in the scene are orthogonal to the horizon line in the rectified image, corners of the scene at the intersection of a horizontal segment at height h (regardless of its compass orientation) and a vertical segment are indeed right-angle corners on the horizon line. Clusters of such corners generally have a weak ranking in the partition, but can still lead to spurious facade detections. As the main purpose is to limit matching hypothesis oversegmentation of the facade is not a really a problem. However a further merging step could find the biggest rectangle and discard the false detection due to the horizon line.

7 Conclusion

We presented a method for facade rectification and detection in urban environment. A Bayesian inference approach was proposed to recover the Manhattan directions in the camera frame. Our algorithm performs better or as well as state-of-the-art techniques and is much faster, mainly as a result of using a suitable prior. In addition, a SVM was used to identify right angle-corners in rectified images. These corners were clustered into rectangular regions in order to identify facades aligned with the Manhattan frame. This approach performed very well in a large variety of frames.

Several improvements could be made to our algorithm. For instance, in this work, the MAP estimate of our model is retained as the VP triplet used for image rectification. However, as a result of our algorithm, several candidate triplets are obtained associated with probability measures. These candidate may be evaluated with regard to criteria measured in the rectified image. For instance, the ratio between the number of right-angle and non-right angle corners may be such a criterion.

Now that we are able to automatically rectify and detect facades in images, our future work will focus on feature matching between faces of a 3D model and a new image, leading to a facade recognition and pose computation procedure. Of course, once a facade is rectified, a much more appropriate strategy than the basic one presented in Fig. 1 can be found to match the facade with the model.

References

[1] V. Angladon, S. Gasparini, and V. Charvillat. The Toulouse Vanishing Points Dataset. In *Multimedia Systems Conference*, 2015.

[2] M. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *CVPR*, 2000.

[3] M. Bansal, K. Daniilidis, and H. Sawhney. Ultra-wide baseline facade matching for geo-localization. In *ECCV Workshop on Visual Analysis and Geo-Localization of Large-Scale Imagery*, 2012.

[4] J.-C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys. Globally optimal line clustering and vanishing point estimation in manhattan world. In *CVPR*, 2012.

[5] S. R. Buss and J. P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *Transaction on Graphics*, 20(2):95–126, 2001.

[6] J. M. Coughlan and A. L. Yuille. Manhattan world: compass direction from a single image by bayesian inference. In *ICCV*, 1999.

[7] J. M. Coughlan and A. L. Yuille. Manhattan world: Orientation and outlier detection by bayesian inference. *Neural Comp.*, 15(5), 2003.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[9] P. Denis, J. H. Elder, and F. J. Estrada. Efficient edge-based methods for estimating manhattan frames in urban imagery. In *ECCV*, 2008.

[10] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: a Line Segment Detector. *Image Proc. On Line*, 2:35–55, 2012.

[11] D. Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *J. Math. Imaging Vis.*, 35(2):155–164, Oct. 2009.

[12] J. T. Kent. The fisher-bingham distribution on the sphere. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1), 1982.

[13] J. Kosecka and W. Zhang. Video compass. In *ECCV*, 2002.

[14] J. Kosecka and W. Zhang. Extraction, matching and pose recovery based on dominant rectangular structures. *CVIU*, 100, 2005.

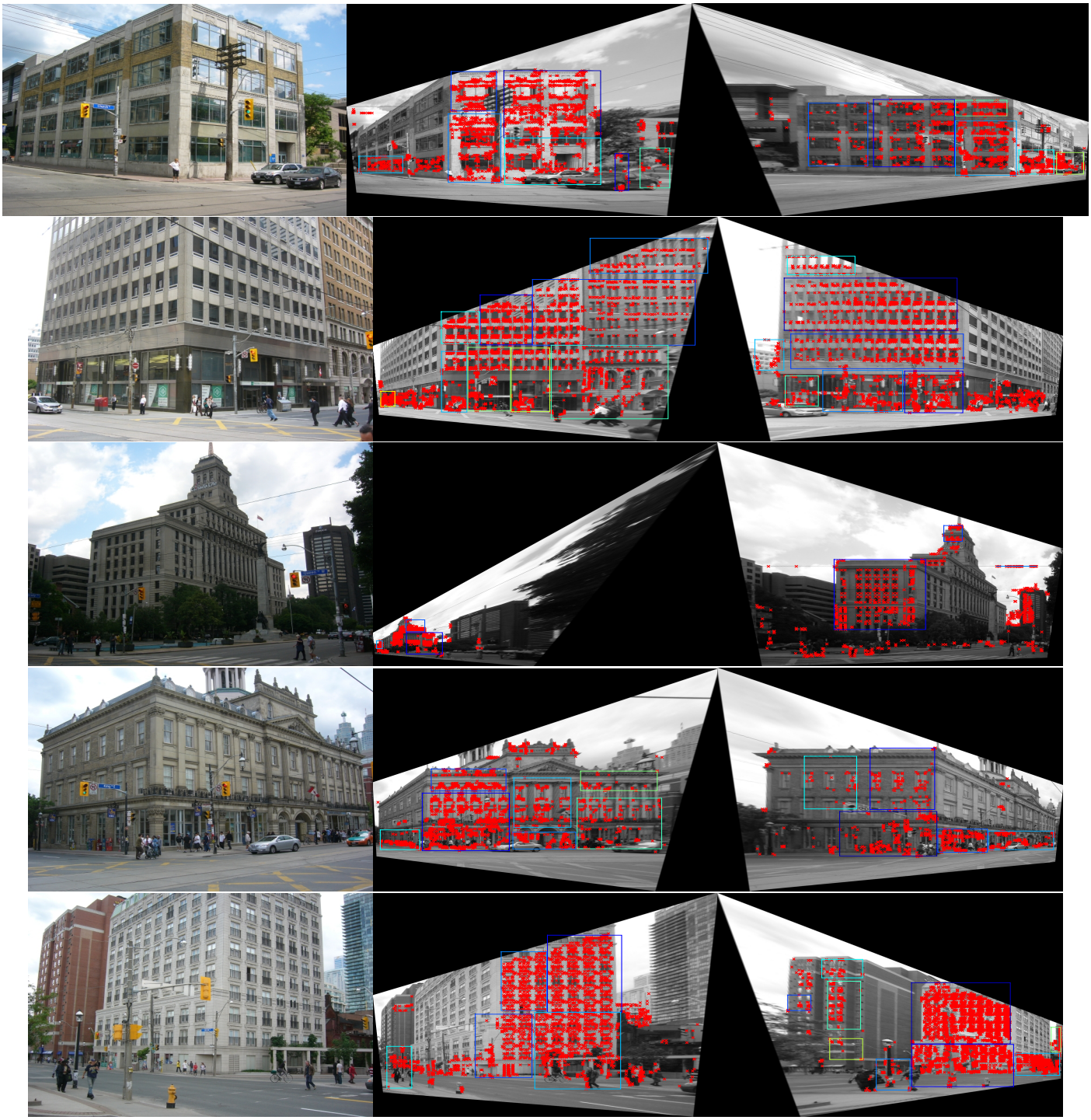

Figure 7: Facade rectification and detection for 5 images of the York Urban DB. First column shows the original images, second and third column show the rectified images according to H_1 and H_2. Only corners that were classified as 'right-angle' (red crosses) and rectangles with partition score below a threshold are shown. Rectangle colors are mapped to the score ranking using the Jet colormap.

[15] J. Lezama, R. Grompone von Gioi, G. Randall, and J.-M. Morel. Finding vanishing points via point alignments in image primal and dual domains. In *CVPR*, 2014.

[16] M. Magee and J. Aggarwal. Determining vanishing points from perspective images. *Comp. Vis., Graphics, and Image Proc.*, 26(2), 1984.

[17] K. V. Mardia and P. E. Jupp. *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2000.

[18] J. Shi and C. Tomasi. Good features to track. In *CVPR*, 1994.

[19] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, 2009.

[20] T. Werner and A. Zisserman. New techniques for automated architectural reconstruction from photographs. In *ECCV*, 2002.

[21] H. Wildenauer and A. Hanbury. Robust camera self-calibration from monocular images of Manhattan worlds. In *CVPR*, 2012.