Video Compass

Jana Košecká and Wei Zhang[*]

Department of Computer Science,
George Mason University, Fairfax, VA 22030, USA
{kosecka,wzhang2}@cs.gmu.edu,
www home page: http://cs.gmu.edu/~kosecka

Abstract. In this paper we describe a flexible approach for determining the relative orientation of the camera with respect to the scene. The main premise of the approach is the fact that in man-made environments, the majority of lines is aligned with the principal orthogonal directions of the world coordinate frame. We exploit this observation towards efficient detection and estimation of vanishing points, which provide strong constraints on camera parameters and relative orientation of the camera with respect to the scene.

By combining efficient image processing techniques in the line detection and initialization stage we demonstrate that simultaneous grouping and estimation of vanishing directions can be achieved in the absence of internal parameters of the camera. Constraints between vanishing points are then used for partial calibration and relative rotation estimation. The algorithm has been tested in a variety of indoors and outdoors scenes and its efficiency and automation makes it amenable for implementation on robotic platforms.

Key words: Vanishing point estimation, relative orientation, calibration using vanishing points, vision guided mobile and aerial robots.

1 Introduction

The problem of recovering relative orientation of the camera with respect to the scene is of importance in a variety of applications which require some knowledge of the environment's geometry or relative pose of the camera with respect to the scene. These range from basic structure and motion or pose recovery problems from single or multiple views, autonomous robotic navigation, manipulation and human computer interaction tasks. Recent efforts in building large city models as well as basic surveillance and monitoring applications often encounter the alignment problem of registering current view to the model or previously stored view. The structural regularities of man-made environments, such as presence of sets of parallel and orthogonal lines and planes can be exploited

[*] This work is supported by NSF grant IIS-0118732 and George Mason University Research Initiative Grant.

towards determining the relative orientation of the camera using the information about vanishing points and vanishing lines. The problem of vanishing point detection and estimation have been addressed numerous times in the past. The proposed approaches vary in the level of automation, computational complexity, assumptions about camera calibration and initialization and grouping stage. The geometric constraints imposed by vanishing directions on camera intrinsic parameters and rotation estimation as well as associated estimation techniques are well understood and have been used previously in the context of structure and motion recovery problems in the uncalibrated case. However the grouping of the line segments into vanishing directions has been often considered separately from the geometric estimation problems, or it has been studied in the case of calibrated camera.

In this paper we will advocate an integrated approach, where the constraints of man-made environments are exploited in different stages of the algorithm pipeline. This and the assumption of uncalibrated camera yields an efficient and flexible approach for estimation of relative orientation of the camera with respect to the scene. We believe that the presented approach is superior to the previously suggested techniques and due to its efficiency it is amenable to implementation on robotic platforms.

1.1 Related Work

The subproblems and techniques used in our approach fall into two broad categories: vanishing point estimation and geometric constraints of uncalibrated single view. We will review more recent representatives of these works and point out the commonalities and differences between our approach. The starting point common to all techniques is the line detection and line fitting stage. The traditional textbook approach suggests the edge detection, followed by edge chaining and line fitting. The constraints of man-made environments, where the majority of line segments is aligned with three principal orthogonal directions, can make this stage more efficient. In cases when the camera is calibrated, the image line segments are represented as unit vectors on the Gaussian sphere and several techniques for both grouping and initialization stage on the Gaussian sphere exist [1, 2, 4]. The main advantage of the Gaussian sphere representation is the equal treatment of all possible vanishing directions, including those at infinity. In [1] the authors demonstrated efficient technique for automated grouping and vanishing point estimation using expectation maximization (EM) algorithm, using a wide field of view camera, which was calibrated ahead of time. While the insensitivity of the Gaussian sphere representation with respect to the focal length has been demonstrated by [2], we show that with proper normalization, identical grouping and estimation problem can be formulated in the absence of camera intrinsic parameters. Similarly as in [1] we formulate the simultaneous grouping and estimation stage using Expectation Maximization (EM) algorithm on the unit sphere, assuming uncalibrated cameras and proposing more efficient initialization scheme.

The initialization and grouping are the determining factors of the efficiency. Previous techniques vary in the choice of the accumulator space, where the peaks correspond to the dominant clusters of line segments; most common alternatives are the Gaussian sphere and Hough space [1, 4, 12, 14]. In some cases all pairwise intersections of the detected line segments are considered for initialization, yielding dominant peak detection [13, 2]. While this strategy has been shown to lead to more accurate detection, the running time is quadratic in terms of the number of line segments. By exploiting the constraints of man-made environments, we suggest the initialization stage which is linear in the number of detected line segments. The errors in the initialization stage are reconciled in the grouping and estimation stage using EM.

Given the detected line segments, the MAP estimates of vanishing points can be obtained by minimizing the distance of the line end points from the ideal lines passing through the vanishing point and leads to a nonlinear optimization problem [6]. An alternative to the nonlinear minimization is a covariance weighted linear least squares formulation suggested first in [9], which tries to minimize the algebraic errors. The EM iterations are in spirit similar to this approach.

In man-made environments the dominant vanishing directions are usually those belonging to three orthogonal directions associated with the reference world coordinate frame. These orthogonality constraints and partial assumptions about camera parameters can be used towards estimation of remaining internal camera parameters and relative orientation of the camera with respect to the scene. In the case of zero skew and known aspect ratio Caprile and Torre [3] have shown that the principal point can be recovered by a geometric construction of the orthocenter. Thorough study of the structure constraints, such as parallelism, orthogonality, known distances and angles and their role in recovering metric properties of the scene can be found in [10]. The use of partially calibrated camera for recovery of pose and structure from single view has been also explored by [7].

We present an integrated approach to the problem of estimation of relative orientation, where the constraints of man-made environments are exploited at different stages of the algorithm pipeline. The approach demonstrates that simultaneous grouping and estimation of vanishing directions can be effectively addressed using Expectation Maximization algorithm in the presence of uncalibrated camera.

2 Approach

Line Detection Stage The first step of the line detection stage is the computation of image derivatives followed by the non-maximum suppression using Canny edge detector. The line fitting stage follows an approach suggested by [8]. The gradient direction is quantized into a set of k ranges, where all the edge pixels having an orientation within the certain range fall into the corresponding bin and are assigned a particular label. In our case $k = 16$. The edge pixels having the same label are then grouped together using the connected components algorithm. The artifacts caused by the fact that the pixels belonging to the same line segment fall into different bins, due to the differences in the orientation are reconciled in the connected components computation stage, where the grouping considers also pixels whose gradient orientation extends beyond bin boundaries. For the purpose of calculation of vanishing points we only consider dominant connected components, whose length is beyond certain threshold, which depends on the size of the image, Each obtained connected component with the length above certain threshold, in our case 5% of the image size, is represented as a list of edge pixels (x_i, y_i) which form the line support region. The line parameters are directly determined from the eigenvalues λ_1 and λ_2 and eigenvectors e_1 and e_2 of matrix D associated with the line support region:

$$D = \begin{bmatrix} \sum_i \tilde{x}_i^2 & \sum_i \tilde{x}_i \tilde{y}_i \\ \sum_i \tilde{x}_i \tilde{y}_i & \sum_i \tilde{y}_i^2 \end{bmatrix} \tag{1}$$

where $\tilde{x} = x_i - \bar{x}$ and $\tilde{y} = y_i - \bar{y}$ are the mean corrected pixel coordinates belonging to a particular connected component where $\bar{x} = \frac{1}{n}\sum_i x_i$ and $\bar{y} = \frac{1}{n}\sum_i y_i$. The quality of the line fit is characterized by the ratio of the two eigenvalues $\frac{\lambda_1}{\lambda_2}$ and the line parameters (ρ, θ) are determined from the eigenvector associated with the largest eigenvalue e_1, as follows:

$$\begin{aligned} \theta &= \mathrm{atan2}(e_1(2), e_1(1)) \\ \rho &= \bar{x}\cos\theta + \bar{y}\sin\theta \end{aligned} \tag{2}$$

where $\bar{\mathbf{x}}(\bar{x}, \bar{y})$ is the mid-point of the line segment. The end points $\mathbf{x}_1$ and $\mathbf{x}_2$ of line $\mathbf{l}$ are determined from the line equation (2) and the line extent. In practice many detected line segments do not come from the actual environment edge segments lines and are due to either shadow or shading effects. These spurious line segments are inevitable and effectively depend on the choice of the threshold, which in this stage is selected globally for the entire image. One possible way how to avoid the commitment to the particular threshold in this early stage is to pursue probabilistic techniques as suggested in [5], which directly relate the line slope to gradient orientation in terms of likelihood function. This strategy is applicable only in the case of calibrated camera. In our approach we will revise our commitment in the later grouping stage, where the spurious line segments will be classified as outliers.

2.1 Vanishing points

Due to the effects of perspective projection, the line segments parallel in the 3D world intersect in the image. Depending on the orientation of the lines the intersection point can be finite or infinite and is referred to as vanishing point. Consider the perspective camera projection model, where 3D coordinates of points $\mathbf{X} = [X, Y, Z, 1]^T$ are related to their image projections $\mathbf{x} = [x, y, 1]^T$ in a following way:

$$\lambda \mathbf{x} = P g \mathbf{X}$$

where $P = [I_{3\times3}, 0] \in \mathbb{R}^{3\times4}$ is the projection matrix, $g = (R, T) \in SE(3)$ is a rigid body transformation represented by 4×4 matrix using homogeneous coordinates and λ is the unknown scale corresponding to the depth Z of the point $\mathbf{X}$. Given two image points $\mathbf{x}_1$ and $\mathbf{x}_2$, the line segment passing through the two endpoints is represented by a plane normal of a plane passing through the center of projection and intersecting the image in a line l, such that $\mathbf{l} = \mathbf{x}_1 \times \mathbf{x}_2 = \widehat{\mathbf{x}}_1\mathbf{x}_2$[1]. The unit vectors corresponding to the plane normals $\mathbf{l}_i$ can be viewed as points on a unit sphere. The vectors $\mathbf{l}_i$ corresponding to the parallel lines in 3D world all lie in the plane, whose intersection with the Gaussian sphere forms a great circle. The vanishing direction then corresponds to the plane normal where all these lines lie. Given two lines the common normal is determined by $\mathbf{v} = \mathbf{l}_1 \times \mathbf{l}_2 = \widehat{\mathbf{l}}_1\mathbf{l}_2$. Hence given a set of line segments belonging to the lines parallel in 3D, the common vanishing direction $\mathbf{v}$ can be obtained by solving the following linear least squares estimation problem:

$$\min_{\mathbf{v}} \sum_{i=1}^{n} (\mathbf{l}_i^T \mathbf{v})^2$$

This corresponds to $\min_{\mathbf{v}} \|A\mathbf{v}\|^2$, where the rows of matrix $A \in \mathbb{R}^{n\times3}$ are the lines segments $\mathbf{l}_i$ belonging to the same vanishing direction. This particular least squares estimation problem has be studied in [4] assuming the unit vectors on the sphere are distributed according to Binghman distribution. The optimal solution to this type of orthogonal least squares problems is also described in [9].

Uncalibrated camera Given a set of line segments sharing the same vanishing direction, the orthogonal least squares solution is applicable regardless of the camera being calibrated. We would like to address the problem of estimation of vanishing points and grouping the lines into vanishing directions simultaneously. In order to be able to determine and adjust along the way the number of groups present in the image some notion of a distance between the line and vanishing direction or two vanishing directions is necessary. Such distance is in the calibrated setting captured by the notion of an inner product between two vectors $u^T v$ with $u, v \in \mathbb{R}^3$. In the case of uncalibrated camera the image coordinates undergo an additional transformation K which depends on the internal camera parameters:

$$\mathbf{x}' = K\mathbf{x} \text{ with } K = \begin{bmatrix} \alpha_x & \alpha_\theta & o_x \\ 0 & \alpha_y & o_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f & \alpha_\theta & o_x \\ 0 & kf & o_y \\ 0 & 0 & 1 \end{bmatrix}.$$

where f is the focal length of the camera in pixel units, k is the aspect ratio and $[o_x, o_y, 1]^T$ is the principal point of the camera. The inner product $u^T v$ in the uncalibrated setting becomes:

$$u^T v = u^{'T} K^{-T} K^{-1} v'$$

[1] $\hat{\mathbf{x}}$ is a skew symmetric matrix associated with $\mathbf{x} = [x_1, x_2, x_3]^T$.

where $u' = Ku$ and $v' = Kv$. The inner product now depends on an unknown matrix $S = K^{-T}K^{-1}$, which can be interpreted as a metric of the uncalibrated space and hence causes a distortion of the original space.

In following we will seek a normalizing transformation which preserves the vanishing directions and makes the process of simultaneous grouping and vanishing point estimation well conditioned. First we will demonstrate that by transforming the measurements (image coordinates) by an arbitrary nonsingular transformation A has no effect on the computation of the vanishing points.

Lemma 1. *If $v \in \mathbb{R}^3$ and $A \in SL(3)$, then $A^T \widehat{v} A = \widehat{A^{-1}v}$.*

Proof. Note that both $A^T \widehat{(\cdot)} A$ and $\widehat{A^{-1}(\cdot)}$ are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^{3\times 3}$, using the fact that $\det(A) = 1$, one may directly verify that these two linear maps are equal on the bases: $(1,0,0)^T, (0,1,0)^T$ or $(0,0,1)^T$.

Suppose that the lines end points are $\mathbf{x}'_1, \mathbf{x}'_2$ and $\mathbf{x}'_3, \mathbf{x}'_4$, such that $\mathbf{l}'_1 = \mathbf{x}'_1 \times \mathbf{x}'_2$ and $\mathbf{l}'_2 = \mathbf{x}'_3 \times \mathbf{x}'_4$, where the measurements $\mathbf{x}'_i = A\mathbf{x}_i$ are related to the calibrated image coordinates by some unknown nonsingular transformation A. We wish to show that vanishing point $\mathbf{v}'$ computed as normal to the plane spanned by vectors $\mathbf{l}'_1$ and $\mathbf{l}'_2$; $\mathbf{v}' = \mathbf{l}'_1 \times \mathbf{l}'_2$ is related to the actual vanishing direction in the calibrated space by the unknown transformation A, namely $\mathbf{v} = A^{-1}\mathbf{v}'$ is the same vanishing direction as the one recovered in the calibrated case. Hence using the above lemma in the context of our problem we have:

$$\begin{gathered}\mathbf{v}' = \mathbf{l}'_1 \times \mathbf{l}'_2 = (\widehat{A\mathbf{x}_1} A\mathbf{x}_2) \times (\widehat{A\mathbf{x}_3} A\mathbf{x}_4) = (A^{-T}\widehat{\mathbf{x}_1}\mathbf{x}_2) \times (A^{-T}\widehat{\mathbf{x}_3}\mathbf{x}_4) = \\ (A^{-T}\mathbf{l}_1) \times (A^{-T}\mathbf{l}_2) = A\widehat{\mathbf{l}_1}\mathbf{l}_2 = A\mathbf{v}\end{gathered} \tag{3}$$

From now on we will drop the $'$ superscript while considering image measurements and assume the uncalibrated case. The above fact demonstrates that in the context of vanishing point estimation, transforming the image measurements by an arbitrary nonsingular transformation A and then transforming the result back, does not affect the final estimate. The particular choice of A is described in the paragraph of the following section. We will use this fact in the normalization step of the least squares estimation in the context of EM algorithm.

2.2 Initialization and grouping

Prior to the vanishing point estimation the line segments obtained in the previous stage need to be grouped into the dominant vanishing directions. In general such grouping is a difficult problem, since any two parallel lines intersect in a vanishing point, possibly yielding a large set of vanishing points. In the case of man-made environments we will exploit the fact that the dominant vanishing directions are aligned with the principal orthogonal axes e_i, e_j, e_k of the world reference frame. Hence there will be only few principal directions, where the majority of the lines belong. We address the grouping stage and vanishing point estimation stage simultaneously as a problem of probabilistic inference with an unknown model. In such instances the algorithm of choice is the Expectation Maximization algorithm (EM), which estimates the coordinates of vanishing points as well as the probabilities of individual line segments belonging to a particular vanishing directions. This approach has been suggested previously by [1], assuming calibrated camera and Gaussian Sphere representation. We will demonstrate that with proper normalization, the identical technique can be applied in case of an uncalibrated camera and present more efficient initialization stage.

The posterior distribution of the vanishing points given line segments can be expressed using Bayes rule in terms of the conditional distribution and prior probability of the vanishing points:

$$p(\mathbf{v}_k \mid \mathbf{l}_i) = \frac{p(\mathbf{l}_i \mid \mathbf{v}_k)p(\mathbf{v}_k)}{p(\mathbf{l}_i)} \tag{4}$$

where $p(\mathbf{l}_i \mid \mathbf{v}_k)$ is the likelihood of the line segment belonging to a particular vanishing direction $\mathbf{v}_k$. Hence for a particular line segment, $p(\mathbf{l}_i)$ can be expressed using the conditional mixture model representation:

$$p(\mathbf{l}_i) = \sum_{k=1}^{m} p(\mathbf{v}_k)p(\mathbf{l}_i \mid \mathbf{v}_k) \tag{5}$$

The number of possible vanishing directions m, will vary depending on the image, but in general we will assume that there are at most four significant models, three corresponding to the dominant vanishing directions and an additional one modeling the outlier process. Line segments which do not belong to the vanishing direction aligned with the principal axes e_i, e_j, e_k are considered to be outliers. The choice of the likelihood term $p(\mathbf{l}_i \mid \mathbf{v})$ depends on the form of the objective being minimized as well as the error model. In the noise free case $\mathbf{l}_i^T\mathbf{v}_k = 0$. In the case of noisy estimates we assume that the error represented by $\mathbf{l}_i^T\mathbf{v}_k$ is a normally distributed random variable with $N(0, \sigma_1^2)$. Then the likelihood term is given as:

$$p(\mathbf{l}_i \mid \mathbf{v}_k) \propto exp\left(\frac{-(\mathbf{l}_i^T\mathbf{v}_k)^2}{2\sigma_1^2}\right) \tag{6}$$

Given a set of line segments we would like to find the most likely estimates of vanishing points as well as probabilities of each line belonging to a particular vanishing direction. Given initial estimates of the vanishing points $\mathbf{v}_k, k = 1, \ldots m$, the membership probabilities of a line segment $\mathbf{l}_i$ belonging to the k-th vanishing direction are computed in the following way:

$$p(\mathbf{v}_k \mid \mathbf{l}_i) = \frac{p(\mathbf{l}_i \mid \mathbf{v}_k)p(\mathbf{v}_k)}{\sum_{k=1}^{m} p(\mathbf{l}_i \mid \mathbf{v}_k)p(\mathbf{v}_k)}, \qquad k = 1, \ldots, m \tag{7}$$

The posterior probability terms $p(\mathbf{v}_k \mid \mathbf{l}_i)$ represent so called membership probabilities, denoted by w_{ik} and capture the probability of a line segment $\mathbf{l}_i$ belonging to k-th vanishing direction $\mathbf{v}_k$. Initially we assume that the prior probabilities of all vanishing directions are equally likely and hence do not affect the posterior conditional probability. The prior probabilities of the vanishing directions can be estimated from the likelihoods and can affect favorably the convergence process as demonstrated in [1]. In the following paragraph we describe the main ingredients of the EM algorithm for simultaneous grouping and estimation of vanishing directions.

Normalization Prior starting the estimation process and determining the line segment representation, we first transform all the endpoint measurements by A^{-1}, in order to make the line segments and vanishing directions well separated on the unit sphere and consequently similar to the calibrated setting:

$$\mathbf{x} = A^{-1}\mathbf{x}' = \begin{bmatrix} \frac{1}{f^*} & 0 & -\frac{o_x^*}{f^*} \\ 0 & \frac{1}{f^*} & -\frac{o_y^*}{f^*} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}'$$

Given an image of size $s = [nrows, ncols]$ the choice of the transformation A is determined by the size of the image and captures the assumption that the optical center is in the center of the image and the aspect ratio $k = 1$. The focal length in the pixel units is $f^* = nrows$, $o_x^* = \frac{nrows}{2}$ and $o_y = \frac{ncols}{2}$. Given the assumptions about optical center and aspect ratio, the chosen focal length f^* is related to the actual focal length by a scale factor.

The first phase of the EM algorithm, the E-step, amounts to computation of posterior probabilities $p(\mathbf{v}_k \mid \mathbf{l}_i)$ given the currently available vanishing points estimates. The goal is to estimate the coordinates of all vanishing points so as to maximize the likelihood of the vanishing point estimates given a set of line segments. The M-step of the algorithm involves maximization of the expected complete log likelihood with respect to the unknown parameters $\mathbf{v}_k$ [11]. This step yields a maximization of the following objective function:

$$\max_{\mathbf{v}_k} \prod_{i=1}^{n} p(\mathbf{l}_i) = \sum_{i=1}^{n} \log p(\mathbf{l}_i) \qquad (8)$$

where $p(\mathbf{l}_i \mid \mathbf{v}_k)$ is the likelihood term defined in (6). The above objective function in case of linear log-likelihood model yields a solution to a weighted least squares problem; one for each model. Each line has an associated weight w_{ik} determined by posterior probabilities computed in the E step. In such case the vanishing points are estimated by solving the following linear least-squares problem:

$$J(\mathbf{v}_k) \;=\; \min_{\mathbf{v}_k} \sum_i w_{ik} (\mathbf{l}_i^T \mathbf{v}_k)^2 \;=\; \min_{\mathbf{v}_k} \|WA\mathbf{v}_k)\|^2 \qquad (9)$$

Where $W \in \mathbb{R}^{n\times n}$ is a diagonal matrix associated with the weights and rows of $A \in \mathbb{R}^{3\times n}$ are the detected line segments. Closed form solution corresponds

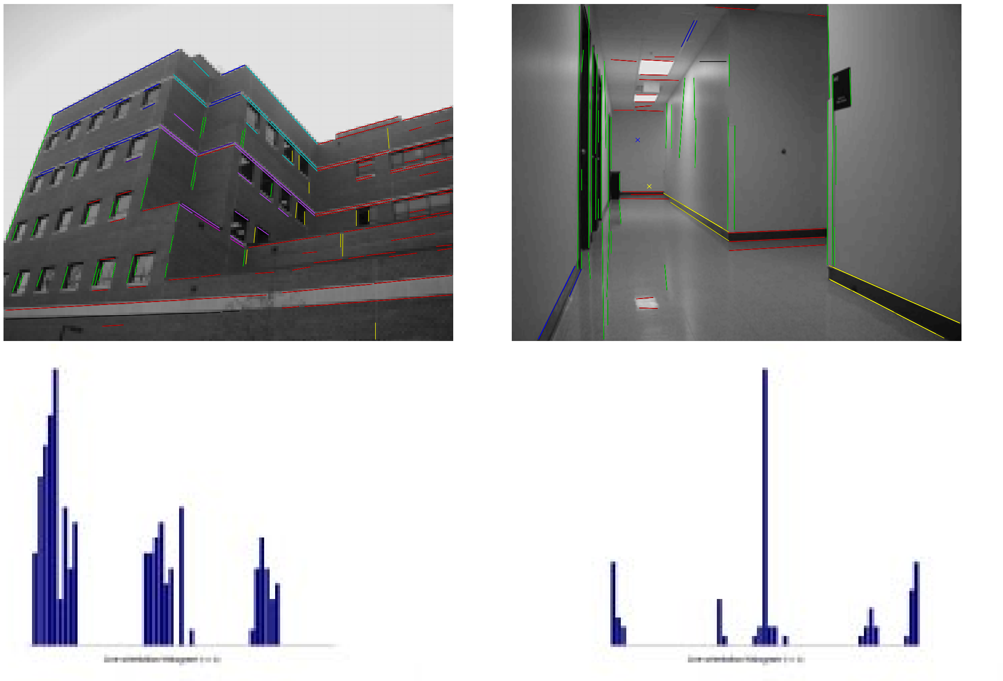

Fig. 1. Detected line segments and associated orientation histograms of the lines. The color coding corresponds to the initial membership assignment of the individual line segments. The initial number of groups exceeds 4.

to the eigenvector associated with the smallest eigenvalue of A^TW^TWA and yields the new estimate of $\mathbf{v}_k$. The EM algorithm is an iterative technique guaranteed to increase the likelihood of the available measurements. We terminate the iteration procedure once the vanishing point estimates reach an equilibrium, *i.e.* $\mathbf{v}^{jT}\mathbf{v}^{(j-1)} < \epsilon_a$. The iterations of the EM algorithm are depicted in Figures 2 and 3. The initially large number of vanishing point estimates, got reduced through the merging process to three dominant directions.

In order to account for the line segments which do not belong to any of the vanishing directions we add an additional *outlier process*. The initial probability of a line segment belonging to the mixture is then determined by the highest possible residual angle $\theta_i = \pi/4$, which a line segment can have with respect to one of the dominant orientations. The outlier probability is then modeled by a Gaussian distribution with a large variance, approximating the uniform distribution. During the course of iteration of the EM algorithm, we also adjust the standard deviation of the individual mixtures, to reflect the increasing confidence in the model parameters and memberships of the individual segments.

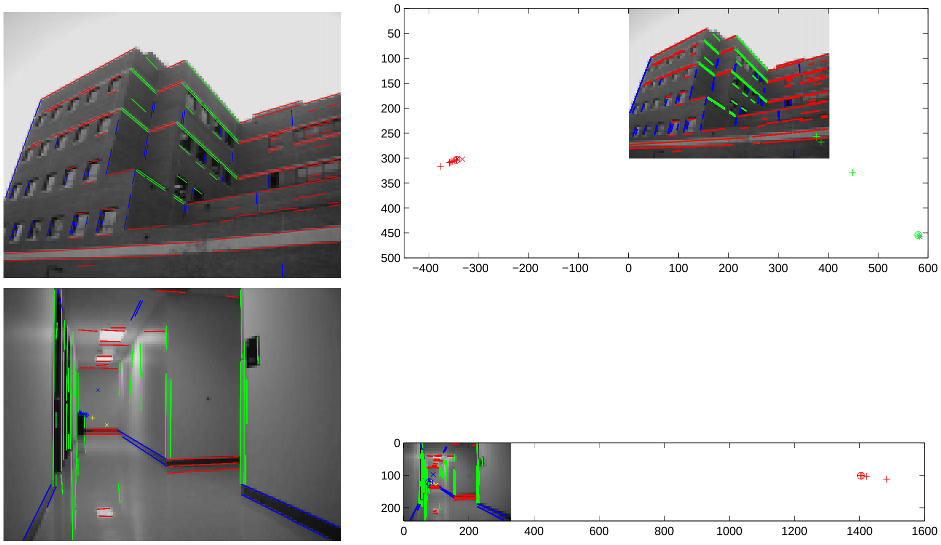

Fig. 2. The final assignment of the lines to three vanishing directions (left). The intermediate estimates of the vanishing points during EM iterations (right). The line segments are color coded based on their membership to the estimated vanishing directions. The vanishing points far from the image center (at infinity) are not plotted. The final vanishing point estimate is marked by 'o'. In the case of the corridor $\hat{\mathbf{v}}_k$ is in the image plane.

Initialization While the EM algorithm is known in each step to increase the likelihood of the measurements given the model, it is often very sensitive to initialization and can converge to a local minimum. Hence it is very important that the algorithm is correctly initialized. The initialization scheme adopted here is based on the observation that the lines that belong to a particular vanishing directions have similar orientation in the image. This is not the case for the line segments whose vanishing point is close to the center of the image plane. In such situation the lines will be initially assigned to a separate groups and merged later in the EM stage. Given a set of detected line segments we form a histogram h_θ of their orientations and search for the peaks in the histogram (see Figure 1). The peaks are detected by first computing the curvature $\mathcal{C}(k)$ of the histogram, followed by a search for zero crossings, which separate the dominant peaks. The curvature is computed by subtracting the local histogram mean:

$$\mathcal{C}(k) = h_\theta(k) - \frac{1}{s}\sum_{i=k-\frac{s}{2}}^{k+\frac{s}{2}+1} h_\theta(i) \tag{10}$$

The total number of line orientation bins in our case is 60 and the size of the integration window $s = 9$. Histogram h_θ is smoothed prior to the curvature computation. The typical number of detected peaks ranges between $2-6$ and

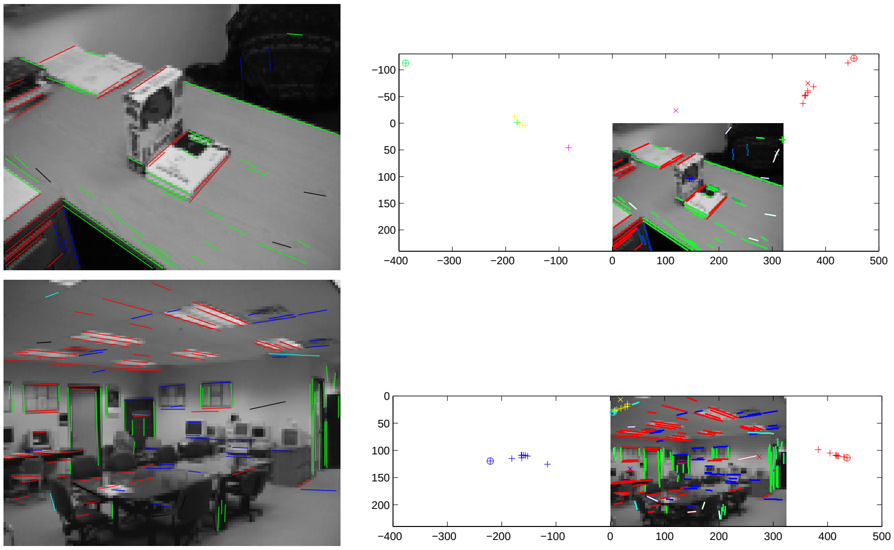

Fig. 3. The final assignment of the lines to three vanishing directions (left). The intermediate estimates of the vanishing points during EM iterations (right) are marked by '+' and the final vanishing point estimate is marked by 'o'. The line segments are color coded based on their membership to the estimated vanishing directions. The white line segments in the left figures correspond to the outliers.

determines the initial number of models in the mixture formulation. The number of models is reconciled in the expectation maximization process. The decision to adjust the number of models is made based of the distance between currently available estimates. If the distance between two vanishing directions $\mathbf{v}_k^T \mathbf{v}_l > \epsilon_b$ has decreased, *i.e.* their inner product is close to 1, the two mixtures are merged and the number of vanishing directions is adjusted. This is usually very easy to detect because thanks to the normalization, the dominant vanishing directions e_i, e_j, e_k are well separated in the image. In Figure 1 of a building there are 6 peaks detected after initialization, three of which were merged during the EM iteration process yielding the assignment of line into three dominant directions depicted in Figure 2.

The performance of the algorithm for the images in Figures 2 and 3 is summarized in Table 1. The vanishing points are reported in the image coordinates with the origin in the upper left corner of the image. The figures also depict the iterations of the EM algorithm. The accuracy of the vanishing point estimation depends on the position of the true vanishing point estimate. We have observed that in case the vanishing point lies in the image plane the standard deviation of the errors was around 5 pixels from the true location, which was determined by hand.

	vanishing point estimate				*vanishing point estimate*		
	$\hat{\mathbf{v}}_i$	$\hat{\mathbf{v}}_j$	$\hat{\mathbf{v}}_k$		$\hat{\mathbf{v}}_i$	$\hat{\mathbf{v}}_j$	$\hat{\mathbf{v}}_k$
Fig. 3	-380.17	456.54	307.36	Fig. 2	1404.23	524.73	78.76
table	-109.62	-122.71	1052.13	corridor	100.97	1.3179×10^5	120.73
Fig. 3.	-220.95	308.36	435.75	Fig. 2	-333.95	306.7	581.31
room	119.65	1069.14	113.65	building	300.14	-591.15	455.79

The above EM algorithm has been tested on a variety of outdoors and indoors scenes and successfully converged in 2-5 iterations, depending on the number of initial vanishing point estimates.

3 Calibration

The following section will demonstrate how to exploit the vanishing point constraints in order to partially determine camera calibration and relative orientation of the camera with respect to the scene. In the calibrated and uncalibrated case the relationship between image coordinates of a point and its 3D counterpart is as follows:

$$\lambda\mathbf{x} = R\mathbf{X} + T \;\text{ and }\; \lambda\mathbf{x}' = KR\mathbf{X} + KT \tag{11}$$

where $\mathbf{x}'$ denotes a pixel coordinate of $\mathbf{X}$ and K is the matrix of internal parameters of the camera. Let's denote the unit vectors associated with the world coordinate frame to be: $e_i = [1,0,0]^T$, $e_j = [0,1,0]^T$, $e_k = [0,1,0]^T$. The vanishing points corresponding to 3D lines parallel to either of these directions are $\mathbf{v}_i = KRe_i, \mathbf{v}_j = KRe_j, \mathbf{v}_k = KRe_k$. Hence the coordinates of vanishing points depend on rotation and internal parameters of the camera. The orthogonality relations between e_i, e_j, e_k readily provide constraints on the calibration matrix K. In particular we have:

$$e_i^T e_j = 0 \Rightarrow \mathbf{v}_i^T K^{-T} R R^T K^{-1} \mathbf{v}_j = \mathbf{v}_i^T K^{-T} K^{-1} \mathbf{v}_j = \mathbf{v}_i^T S \mathbf{v}_j = 0 \tag{12}$$

where S is the metric associated with the uncalibrated camera:

$$S = K^{-T}K^{-1} = \begin{pmatrix} s_1 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{pmatrix}$$

When three finite vanishing points are detected, they provide three independent constraints on matrix S:

$$\begin{aligned} \mathbf{v}_i^T S \mathbf{v}_j &= 0 \\ \mathbf{v}_i^T S \mathbf{v}_k &= 0 \\ \mathbf{v}_j^T S \mathbf{v}_k &= 0 \end{aligned} \tag{13}$$

In general matrix symmetric matrix $S_{3\times3}$ has six degrees of freedom and can be recovered up to a scale, so without additional constraints we can recover the S

only up to a two parameter family. Other commonly assumed assumption of zero skew and known aspect ratio can provide additional constraints and can also be expressed in terms of constraints on the metric S [10]. The zero skew constraint expresses the fact that the image axes are orthogonal can be written as:

$$[1,0,0]^T S[0,1,0] = 0$$

In the presence of zero skew assumption, the known aspect ratio constraint can be expressed as $s_1 = s_4$. By imposing these two additional constraints we obtain a sufficient number of constraints. The solution for $\mathbf{s} = [s_1, s_2, s_3, s_4, s_5, s_6]^T$ can be obtained by minimizing $\|B\mathbf{s}\|^2$ and corresponds to the eigenvector associated with the smallest eigenvalue of $B^T B$. The calibration matrix K^{-1} can be obtained from S by Cholesky decomposition. In the case the vanishing directions arise from a set of lines parallel to the image plane, the associated vanishing point lies at infinity. In practical situations, when one of the vanishing directions is close to infinity one of the constraints becomes degenerate and recovered S fails to be positive definite. This situation can be also noticed by checking the condition number of B. In such case we assume that the principal point lies in the center of the image and hence S is parameterized by the focal length only. The focal length can be then recovered in closed form, from the remaining constraint. The recovered calibration matrices for the examples outlined in Figures 2 and 3 are below:

$$K_{building} = \begin{bmatrix} 409.33 & -0.0000 & 177.46 \\ 0 & 409.33 & 165.75 \\ 0 & 0 & 1 \end{bmatrix} \quad K_{table} = \begin{bmatrix} 322.16 & -0.0000 & 289.4949 \\ 0 & 322.16 & -23.5644 \\ 0 & 0 & 1 \end{bmatrix}$$

$$K_{room} = \begin{bmatrix} 361.133 & -0.0000 & 263.99 \\ 0 & 361.133 & 129.038 \\ 0 & 0 & 1 \end{bmatrix} \quad K_{true} = \begin{bmatrix} 408.79 & -0.0000 & 199.50 \\ 0 & 408.79 & 149.5 \\ 0 & 0 & 1 \end{bmatrix}$$

At this stage we carried out only qualitative evaluation of the obtained estimates. The focal length error in all cases was below 5%. Note that in the above examples the difference in the focal length is due to the difference in the image size. While the sub-sampling affects also the position of the principal point, the above statement assumes that the focal length of the sub-sampled images is related to the original focal length by the sub-sampling factor. The quality of the estimates depends on the accuracy of the estimated vanishing points. As the vanishing points approach infinity their estimates become less accurate. This affects in particular the estimate of principal point, which in case one of the vanishing points is at infinity cannot be uniquely determined unless additional constraints are introduced [10]. In such case we assume that the principal point of the camera lies in the center of the image and estimate the focal length in the closed form, using a single orthogonality constraint between vanishing directions. The estimate of the focal length obtained in this manner has larger error compared to the case when all the constraints are used simultaneously (if available) and the principal point is estimated as well.

3.1 Relative orientation

Once the vanishing points have been detected and the unknown camera parameters determined by the above procedure, the relative orientation of the camera with respect to the scene can be computed. Note first that since the vanishing directions are projections of the vectors associated with three orthogonal directions i, j, k, they depend on rotation only. In particular we can write that:

$$K^{-1}\mathbf{v}_i = Re_i \quad K^{-1}\mathbf{v}_j = Re_j \quad K^{-1}\mathbf{v}_k = Re_k$$

with each vanishing direction being proportional to the column of the rotation matrix $R = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$. Choosing the two best vanishing directions, properly normalizing them, the third row can be obtained as $\mathbf{r}_3 = \widehat{\mathbf{r}}_1\mathbf{r}_2$ by enforcing the orthogonality constraints. There is a four way ambiguity in R due to the sign ambiguity in $\mathbf{r}_1$ and $\mathbf{r}_2$. Additional solutions can be eliminated by considering relative orientation or structure constraints.

4 Summary and Conclusions

We presented an efficient, completely automated approach for detection of vanishing points from a single view assuming an uncalibrated camera. Along the way the assumptions about the structure of man-made environments, were used in the initialization and grouping stage. In particular it was the fact that the majority of the line segments belongs to one of the dominant vanishing directions aligned with the axes of the world coordinate frame. The estimation and grouping problems were addressed simultaneously using the Expectation Maximization algorithm. We are currently exploring the use of the prior information in the vanishing point estimation task and developing more quantitative assessment of the sensitivity of the estimation process. While this problem has been studied frequently in the past in the context of different computer vision applications, in most of those instances the speed and robustness were not the primary concerns. The capability of robustly detecting vanishing points in an automated manner will enable us to employ these types of systems in the context of mobile and aerial robots and facilitate partial calibration of the vision system as well as estimation of relative orientation with respect to the scene. This information has great utility in the context of basic navigation and exploration tasks in indoor and outdoor environments, where the alternative sensing strategies such as GPS or compass are known to under perform.

References

1. M. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *IEEE Proceedings of CVPR*, 2000.
2. B. Brillaut-O'Mahony. New method for vanishing point detection. *CVGIP: Image Understanding*, 54(2):289–300, September 1991.

3. B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 3:127–140, 1990.
4. R. Collins. Vanishing point calculation as statistical inference on the unit sphere. In *Proceedings of International Conference on Computer Vision*, pages 400–403, 1990.
5. J. Coughlan and A. Yuille. Manhattan world: Compass direction from a single image by bayesian inference. In *IEEE Proceedings International Conference on Computer Vision*, pages 941–7, 1999.
6. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
7. D. Jelinek and C.J. Taylor. Reconstruction of linearly parametrized models from single images with camera of unknown focal lenght. *IEEE Transactions of PAMI*, pages 767–774, July 2001.
8. P. Kahn, L. Kitchen, and E.M. Riseman. A fast line finder for vision-guided robot navigation. *IEEE Transactions on PAMI*, 12(11):1098–1102, 1990.
9. K. Kanatani. *Geometric Computation for Machine Vision*. Oxford Science Publications, 1993.
10. D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proceedings of International Conference on Computer Vision*, 1999.
11. G. J. McLachlan and K. E. Basford. *Mixture Models: Inference and Applications*. Marcel Dekker Inc., N.Y., 1988.
12. L. Quan and R. Mohr. Determining perspective structures using hierarchical hough transforms. *P.R. Letters*, 9:279–286, 1989.
13. C. Rother. A new approach for vanishing point detection in architectural environments. In *Proceedings of the British Machine Vision Conference*, 2000.
14. T. Tuytelaars, M. Proesmans, and L. Van Gool. The cascaded Hough transform. In *Proceedings of ICIP*, pages 736–739, 1998.