Orientation in Manhattan: Equiprojective Classes and Sequential Estimation

André T. Martins,
Pedro M.Q. Aguiar, *Member*, *IEEE*, and
Mário A.T. Figueiredo, *Sr. Member*, *IEEE*

Abstract—The problem of inferring 3D orientation of a camera from video sequences has been mostly addressed by first computing correspondences of image features. This intermediate step is now seen as the main bottleneck of those approaches. In this paper, we propose a new 3D orientation estimation method for urban (indoor and outdoor) environments, which avoids correspondences between frames. The scene property exploited by our method is that many edges are oriented along three orthogonal directions; this is the recently introduced *Manhattan world* (MW) assumption. The main contributions of this paper are: the definition of equivalence classes of equiprojective orientations, the introduction of a new *small rotation* model, formalizing the fact that the camera moves smoothly, and the decoupling of elevation and twist angle estimation from that of the compass angle. We build a probabilistic sequential orientation estimation method, based on an MW likelihood model, with the above-listed contributions allowing a drastic reduction of the search space for each orientation estimate. We demonstrate the performance of our method using real video sequences.

Index Terms—Camera orientation, sequential estimation, Manhattan world assumption, camera calibration.

1 Introduction

Applications in areas such as digital video, virtual reality, mobile robotics, and visual aids for blind people require efficient methods to estimate the 3D pose of a video camera from the images it captures.

The most popular approaches to 3D pose estimation are feature-based. In the multiview case, this requires finding correspondences between features [2], [3], [4]. In the single-image case, typical methods involve feature grouping [5], [6], [7]. Naturally, in both cases, feature detection (e.g., corners, edges) is an indispensable first step. However, it is widely accepted that automatic feature matching or grouping are serious bottlenecks. Moreover, by basing all inference on a usually small feature set (relative to the whole image), potentially useful information may be prematurely discarded.

In the multiview case, methods that estimate the 3D structure directly from the image intensity values, i.e., without involving feature detection and matching, have been proposed [8], [9]. These approaches lead to complex time-consuming algorithms and strongly rely on the assumption that the brightness pattern remains (approximately) constant from view to view.

Recently, a very different approach has been proposed which avoids dealing with features in the single-image case by using prior knowledge about the structure of the scene. Specifically, in typical indoor and outdoor urban scenes, many edges are aligned with one of the three directions defining an orthogonal coordinate system. Under this so-called *Manhattan world* (MW) assumption, Coughlan and Yuille [10], [11] used Bayesian inference to estimate the rotational component of the 3D pose (i.e., 3D orientation) of the camera, with respect to this coordinate system, from a single image. The MW assumption was also used in [12] for camera calibration and extended in [13] to more general urban environments.

In this paper, we propose a new method for 3D orientation estimation from image sequences in MW environments. The novelties in our method are the following:

- While, in [10], [11], the MW prior is used to perform 3D orientation estimation from a *single* image, we extend its use for *sequences* of images.
- We introduce a new *small rotation* (SR) model that expresses the fact that the video camera undergoes a smooth 3D motion.
- By defining the 3D orientation in terms of the equivalence classes of equiprojective orientations, we reduce the space in which the solution has to be searched.
- We show how the estimate of the elevation and twist angles can be computed independently of the compass angle, thus reducing the computational load.

The paper is organized as follows: In Section 2, we review the geometry of camera orientation. The concept of equiprojective orientations and the small rotation (SR) model are introduced in Sections 3 and 4, respectively. Section 5 describes the sequential estimation method. Experimental results are shown in Section 6 and Section 7 concludes the paper.

2 Camera Orientation and Vanishing Points

Let $(\mathbf{x},\mathbf{y},\mathbf{z})$ and $(\mathbf{n},\mathbf{h},\mathbf{v})$ be the Cartesian coordinate systems of the MW and the camera, respectively. These are related through the equation $(\mathbf{n},\mathbf{h},\mathbf{v})^T = \mathrm{O}\cdot(\mathbf{x},\mathbf{y},\mathbf{z})^T$, where $\mathrm{O}\in SO(3)$ is the orientation matrix, i.e., the *camera orientation*. In the following text, we often denote orientation as $\mathrm{O}\equiv\mathrm{O}(\alpha,\beta,\gamma)$, expressing the fact that it is parameterized with three angles: α, the *compass* (azimuth) angle, corresponding to rotation about the $\mathbf{z}$ axis; β, the *elevation* angle above the $\mathbf{xy}$ plane; and γ, the *twist* about the principal axis (see Fig. 1).

The principal point $\mathbf{P}$ lies on the sphere with center at the optical center $\mathbf{0}$ (chosen as the origin of the MW reference frame) and radius equal to the focal length f. Its 3D coordinates are related with the compass and elevation angles via

$$\mathbf{P} = (P_x, P_y, P_z)^T = f(\cos\alpha\cos\beta, \sin\alpha\cos\beta, \sin\beta)^T. \tag{1}$$

The orientation $\mathrm{O}(\alpha,\beta,\gamma)$ can be determined by finding where the vanishing points (VPs) of the MW axes project on the image plane [2], [3]. In fact, let $(\mathbf{h},\mathbf{v})$ be the reference frame of this plane and let the 2D principal point $\mathbf{p}$ be its origin, i.e., $\mathbf{p}=(0,0)^T$. Assuming a pinhole and radial-distortion-free camera, the 2D coordinates, $\mathbf{v_x}$, $\mathbf{v_y}$, $\mathbf{v_z}$, of the VP projections are related with $\mathrm{O}(\alpha,\beta,\gamma)$ via

$$\begin{aligned}\mathbf{v_x} &= f\,\mathrm{R}_\gamma\Big(-\tfrac{\tan\alpha}{\cos\beta}, -\tan\beta\Big)^T,\\ \mathbf{v_y} &= f\,\mathrm{R}_\gamma\Big(\tfrac{\cot\alpha}{\cos\beta}, -\tan\beta\Big)^T,\\ \mathbf{v_z} &= f\,\mathrm{R}_\gamma(0, \cot\beta)^T,\end{aligned} \tag{2}$$

where R_γ is the *twist matrix*,

$$\mathrm{R}_\gamma = \begin{bmatrix}\cos\gamma & \sin\gamma\\ -\sin\gamma & \cos\gamma\end{bmatrix}. \tag{3}$$

In the above, Cartesian coordinates are used only for simplicity; vanishing points at infinity can be handled by using homogeneous coordinates.

- A.T. Martins is with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisboa, Portugal. E-mail: jah@clix.pt.
- P.M.Q. Aguiar is with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisboa, Portugal, and the Institute for Systems and Robotics. E-mail: aguiar@isr.ist.utl.pt.
- M.A.T. Figueiredo is with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisboa, Portugal, and the Institute of Telecommunications. E-mail: mtf@lx.it.pt.

Manuscript received 21 May 2004; revised 20 Oct. 2004; accepted 4 Nov. 2004; published online 11 Mar. 2005.
Recommended for acceptance by L. Quan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0255-0504.

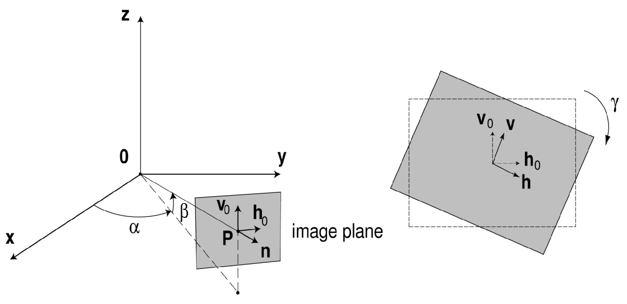

Fig. 1. Parameterization of the camera orientation. Left: Compass angle α and elevation angle β. Right: Twist angle γ represented on the image plane. (Note: The image plane is placed in front of the optical center.)

3 Equiprojective Orientations

Consider the problem of determining the camera orientation from the set of three VPs on a single image. Since it is not known which VP corresponds to which MW axis, the problem has multiple solutions. This ambiguity motivates the concept of *equiprojectivity*.

Definition 1 (equiprojective orientations). *Denote by* $\mathcal{V}(\mathrm{O}) = \{\mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z}\}$ *the set of VPs determined by an orientation* O. *Two orientations* O *and* O^* *are termed equiprojective iff they have identical sets of VPs, i.e., iff* $\mathcal{V}(\mathrm{O}) = \mathcal{V}(\mathrm{O}^*)$.

Equiprojectivity, as just defined, is reflexive, symmetric, and transitive; therefore, it is an equivalence relation. The following result provides a way to find the complete equivalence class of a given orientation, i.e., the set of all orientations which are equiprojective with it.

Proposition 2. *Let* O *be an orientation and* $\mathbf{P} = (P_x, P_y, P_z)^T$ *the corresponding principal point. The equivalence class of* O *always has* 24 *elements. Each* $\mathrm{O}^{(n)} = \mathrm{O}(\alpha_n, \beta_n, \gamma_n)$, *for* $n = 1, \ldots, 24$, *corresponds to a principal point* $\mathbf{P}^{(n)}$ *related to* $\mathbf{P}$ *through* $\mathbf{P}^{(n)} = \mathrm{M}_n\mathbf{P}$, *where* M_n *is a* 3×3 *signed permutation matrix (i.e., entries in* $\{-1, 0, 1\}$, *with one nonzero entry per row and per column) with* $\det \mathrm{M}_n = 1$. *The angles* α_n *and* β_n *are obtainable from* $\mathbf{P}^{(n)}$ *according to (1); the twist angles* γ_n *depend on* $\mathrm{O}(\alpha, \beta, \gamma)$ *and* $\mathbf{P}^{(n)}$ *as follows:*

$$\gamma_n = \begin{cases} \gamma & \Leftarrow \mathrm{M}_n^T\mathbf{z} = (0,0,1)^T & (P_z^{(n)} = P_z) \\ \gamma \pm \pi & \Leftarrow \mathrm{M}_n^T\mathbf{z} = (0,0,-1)^T & (P_z^{(n)} = -P_z) \\ \gamma + \operatorname{atan}\frac{\tan\alpha}{\sin\beta} \pm \pi & \Leftarrow \mathrm{M}_n^T\mathbf{z} = (1,0,0)^T & (P_z^{(n)} = P_x) \\ \gamma + \operatorname{atan}\frac{\tan\alpha}{\sin\beta} & \Leftarrow \mathrm{M}_n^T\mathbf{z} = (-1,0,0)^T & (P_z^{(n)} = -P_x) \\ \gamma - \operatorname{atan}\frac{\cot\alpha}{\sin\beta} \pm \pi & \Leftarrow \mathrm{M}_n^T\mathbf{z} = (0,1,0)^T & (P_z^{(n)} = P_y) \\ \gamma - \operatorname{atan}\frac{\cot\alpha}{\sin\beta} & \Leftarrow \mathrm{M}_n^T\mathbf{z} = (0,-1,0)^T & (P_z^{(n)} = -P_y). \end{cases}$$

Proof. Given an orientation O, the corresponding image plane can be seen as the plane that is tangent to the sphere $\{\mathbf{w} : ||\mathbf{w}|| = f\}$ in $\mathbf{P}$. The intersection of each MW axis $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ with the image plane defines its respective VP. Hence, a necessary condition for an orientation $\mathrm{O}^{(n)}$ to be equiprojective with O is that their corresponding principal points (respectively, $\mathbf{P}^{(n)}$ and $\mathbf{P}$) have the same coordinates up to permutations and/or sign changes, which is equivalent to the existence of a signed permutation matrix M_n satisfying $\mathbf{P}^{(n)} = \mathrm{M}_n\mathbf{P}$. Any permutation matrix satisfies $\det \mathrm{M}_n = \pm 1$; however, not all matrices of this kind yield a solution. Particularly, if $\mathbf{P}$ and $\mathbf{P}^{(n)}$ differ by a single permutation or by a single sign change, the triangles formed by the VPs at each case have opposite orientations, i.e., they are "reflected." Since the composition of two reflections is the identity, the number of permutations plus the number of sign changes defined by any matrix M_n must be *even*; this is equivalent to imposing $\det \mathrm{M}_n = 1$. Because the number of possible permutations in a 3-vector is $3! = 6$ and the number of sign changes is $2^3 = 8$, we can combine permutations and sign changes in 48 different ways; since half of these correspond to "mirror images," the cardinality of the set $\{\mathrm{M}_n\}$ is 24 (see illustration in Fig. 2).

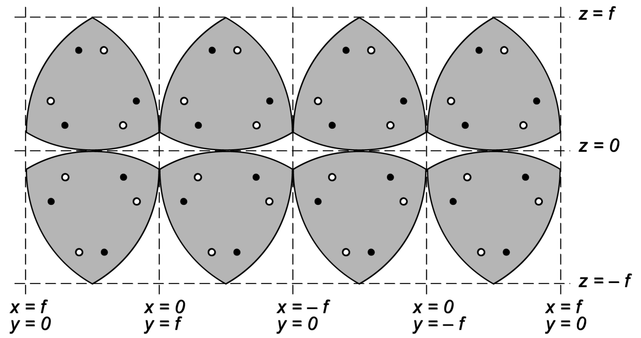

Fig. 2. Three-dimensional locations of the principal points of equiprojective orientations, on the octants of a sphere with radius f. Here, we have two equivalence classes: the white and the black points. Black points correspond to "mirror images" of white points.

For each M_n, we are able to know which VP in $\mathcal{V}(\mathrm{O})$ corresponds to which VP in $\mathcal{V}(\mathrm{O}^{(n)})$. Namely, for every $i, j \in \{x, y, z\}$, the VP $\mathbf{v}_i$ and $\mathbf{v}_j^{(n)}$ correspond iff $\mathbf{j}^T\mathrm{M}_n\mathbf{i} = \pm 1$, i.e., iff $P_i = \pm P_j^{(n)}$. Taking $j = z$, we have:

$$\gamma_n - \gamma = \begin{cases} \angle[\mathbf{v_z}\mathbf{p}\mathbf{v}_i] & \Leftarrow \mathbf{z}^T\mathrm{M}_n\mathbf{i} = 1 \quad (P_z^{(n)} = P_i) \\ \angle[\mathbf{v_z}\mathbf{p}\mathbf{v}_i] \pm \pi & \Leftarrow \mathbf{z}^T\mathrm{M}_n\mathbf{i} = -1 \quad (P_z^{(n)} = -P_i). \end{cases}$$

Finally, from (2)-(3), we obtain the expression for γ_n. □

The concept of equiprojectivity is useful in any problem of orientation estimation, or VP location since it allows reducing the search spaces. This was also pointed out in [12], where an algorithm was proposed to round a quaternion to a canonical value in $SO(3)/C$, where C is the octohedral group of cube symmetries. We formalize this search space reduction in the following proposition (proven in the Appendix).

Proposition 3. *Every orientation* O *has an equiprojective* $\mathrm{O}^* = \mathrm{O}(\alpha^*, \beta^*, \gamma^*)$ *such that:*

$$\alpha^* \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right], \;\; \beta^* \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right], \;\; \textit{and} \;\; \gamma^* \in \left]-\varphi, \varphi\right], \tag{4}$$

where $\varphi = \operatorname{atan}\sqrt{2} \approx 54.7°$. *An equivalent statement is: For any camera orientation* O, *there exists at least one VP inside the region of the image plane shown in Fig. 3.*

4 Small Rotations Model

Let us now assume that the camera is moving and acquiring a sequence of frames $\{\mathrm{I}_1, \ldots, \mathrm{I}_N\}$. We denote by $\mathrm{O}_k(\alpha_k, \beta_k, \gamma_k)$ the orientation at the kth frame. The sequence of orientations $\{\mathrm{O}_1, \ldots, \mathrm{O}_N\}$ depends only on the rotational component of the

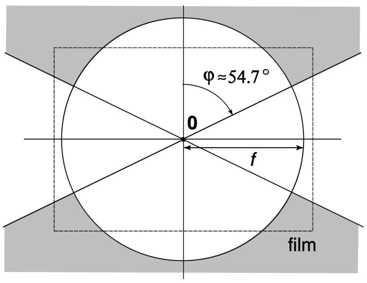

Fig. 3. Representation of the image plane. It is guaranteed that there exists at least one vanishing point in the shaded region.

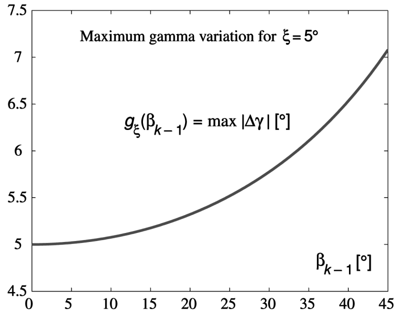

Fig. 4. Maximum variation for the twist angle as a function of the initial elevation angle, using an SR(5°) model.

motion. In typical video sequences, the camera orientation evolves in a smooth continuous way. We formalize this property by introducing the *small rotations* (SR) model, described next.

Definition 4. *Let* $\mathrm{R}_k(\rho_k, \mathbf{e}_k)$ *be the rotational component of the camera motion between the* $(k-1)\mathrm{th}$ *and* $k\mathrm{th}$ *frames, where* ρ_k *and* $\mathbf{e}_k$ *denote the angle and the axis of rotation, respectively. Independently of* $\mathbf{e}_k$*, we say that the camera is consistent with the SR(*ξ*) model iff there exists a small fixed angle* ξ *such that* $|\rho_k| \leq \xi$ *for any* k*.*

In our experiments, we have used an SR(5°) model, which implies that, for a sampling rate of 12.5 Hz, the rotation angle is always less than 62.5° in each second; this is an intuitively reasonable assumption.

The following proposition expresses how the variations of the compass, elevation and twist angles between consecutive frames are bounded due to the SR model.

Proposition 5. *If the camera motion is consistent with the SR(*ξ*) model, then, at any frame* k*, the following bounds hold:*

- *The elevation variation,* $\Delta\beta = \beta_k - \beta_{k-1}$*, satisfies*

$$|\Delta\beta| \leq \xi. \tag{5}$$

- *The compass variation,* $\Delta\alpha = \alpha_k - \alpha_{k-1}$*, satisfies*

$$|\Delta\alpha| \leq a_\xi(\beta_k, \beta_{k-1}) \equiv \begin{cases} \operatorname{acos}\Big(1 - \frac{\cos|\Delta\beta| - \cos\xi}{\cos\beta_{k-1}\cos\beta_k}\Big) & \Leftarrow |\beta_{k-1} + \beta_k| \leq \pi - \xi \\ \frac{\pi}{2} & \Leftarrow \textit{otherwise}. \end{cases} \tag{6}$$

If O_{k-1} *is in the region defined by (4), then, independently of* β_k *and* β_{k-1}*:*

$$|\Delta\alpha| \leq \operatorname{acos}(2\cos\xi - 1). \tag{7}$$

- *The twist variation,* $\Delta\gamma = \gamma_k - \gamma_{k-1}$*, satisfies*

$$|\Delta\gamma| \leq g_\xi(\beta_{k-1}), \tag{8}$$

where g_ξ *is an even function that increases in the subdomain* $[0, \frac{\pi}{2}]$ *from* $g_\xi(0) = \xi$ *to* $g_\xi(\frac{\pi}{2}) = \pi$*. If* O_{k-1} *is in the region defined by (4), then* $|\beta_{k-1}| \leq \frac{\pi}{4}$ *and*

$$|\Delta\gamma| \leq g_\xi\Big(\frac{\pi}{4}\Big), \tag{9}$$

Fig. 4 plots g_ξ *in the subdomain* $[0, \frac{\pi}{4}]$*, for* $\xi = 5°$*; this value of* ξ *leads to* $|\Delta\gamma| \leq 7.08°$*.*

Proof. $\mathrm{R}_k(\rho_k, \mathbf{e}_k)$ is the composition of two rotations: $\mathrm{R}_{k_1}(\rho_{k_1}, \mathbf{e}_{k_1})$ transforming the principal point $\mathbf{P}_{k-1}$ in $\mathbf{P}_k$, followed by $\mathrm{R}_{k_2}(\rho_{k_2}, \mathbf{e}_{k_2})$ that twists the camera through the principal axis. Composing these two rotations and taking into account that $\mathbf{e}_{k_1} \perp \mathbf{e}_{k_2}$, we obtain $\cos\frac{\rho_k}{2} = \cos\frac{\rho_{k_1}}{2}\cos\frac{\rho_{k_2}}{2}$. Therefore, the SR($\xi$) condition $|\rho_k| \leq \xi$ implies both $\cos\frac{\rho_{k_1}}{2} \geq \cos\frac{\xi}{2}$ and $\cos\frac{\rho_{k_2}}{2} \geq \cos\frac{\xi}{2}$, i.e., $|\rho_{k_1}| \leq \xi$ and $|\rho_{k_2}| \leq \xi$. Since $\cos\rho_{k_1} = f^{-2}\,\mathbf{P}_k^T\mathbf{P}_{k-1}$, from (1) we obtain $\cos\rho_{k_1} = \cos\beta_k\cos\beta_{k-1}\cos\Delta\alpha + \sin\beta_k\sin\beta_{k-1} \leq \cos\Delta\beta$. This suffices to prove (5). Now, rewriting the latest inequality for $\cos\Delta\alpha$ and simplifying leads to (6). If O_{k-1} is in the region defined by (4), then $|\beta_k + \beta_{k+1}| \leq \pi/4 + \pi/4 + \xi \leq \pi - \xi$. The maximum value of $\Delta\alpha$ occurs for $\beta_k = \beta_{k-1} = \frac{\pi}{4}$, which leads to (7).

For $\Delta\gamma$, we couldn't find a simple closed-form expression for $g_\xi(\beta_{k-1})$. Instead, since ρ_k is a function of β_{k-1}, β_k, $\Delta\alpha$, and $\Delta\gamma$, we can study g_ξ assuming that $\alpha_{k-1} = \gamma_{k-1} = 0$. Spherical symmetry implies that g_ξ is an even function; also, a simple geometric argument shows that $g_\xi(\beta_{k-1})$ increases with $|\beta_{k-1}|$. Writing R_k as a composition of the three individual compass, elevation and twist rotations, and using the formula for the product of quaternions, yields

$$|\Delta\gamma| = 2\operatorname{acos}\frac{AB - C\sqrt{B^2 + C^2 - A^2}}{B^2 + C^2}, \tag{10}$$

where $A = \cos\frac{\rho_k}{2}$, $B = \cos\frac{\Delta\alpha}{2}\cos\frac{\Delta\beta}{2}$, and $C = \sin\frac{\Delta\alpha}{2}(\cos\frac{\Delta\beta}{2}\sin\beta_k - \cos\beta_k\sin\Delta\beta)$. Numerical maximization of (10) w.r.t. $\Delta\alpha$ and β_k (for $\rho_k = \xi$) approximates g_ξ. □

If the orientation O_{k-1} lies in the minimal region defined by (4), the search space for O_k is significantly reduced by the bounds imposed by Proposition 5. In particular, with $\xi = 5°$, we have $|\Delta\alpha| \leq 7.08°$, $|\Delta\beta| \leq 5°$, and $|\Delta\gamma| \leq 7.08°$. If O_{k-1} does not lie in this minimal region, there is an equiprojective orientation that does. This shows how the SR model and the equiprojective orientations can be used together to reduce the search space.

5 Sequential Orientation Estimation

5.1 Estimation Criterion

To estimate the sequence of camera orientations $\{\mathrm{O}_1, \ldots, \mathrm{O}_N\}$ from the observed image sequence $\{\mathrm{I}_1, \ldots, \mathrm{I}_N\}$, we adopt a probabilistic sequential estimation framework, making use of the MW and SR assumptions.

The MW assumption states that the images contain many edges consistent with the $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ axes; hence, the statistics of the image intensity gradient $\nabla\mathrm{I}_k$ of each image carry information about the corresponding camera orientation O_k via a likelihood function $P(\nabla\mathrm{I}_k|\mathrm{O}_k)$ [10], [11]. In this paper, we embed this idea in a sequential estimation framework, using a *maximum a posteriori* (MAP) criterion:

$$\widehat{\mathrm{O}}_k = \arg\max_{\mathrm{O}_k}\Big\{\log P(\nabla\mathrm{I}_k|\mathrm{O}_k) + \log P(\mathrm{O}_k|\widehat{\mathrm{O}}_{k-1})\Big\}, \tag{11}$$

where the prior $P(\mathrm{O}_k|\widehat{\mathrm{O}}_{k-1})$ penalizes large changes between consecutive orientation estimates.

A fully Bayesian sequential estimation approach would require computationally expensive Monte Carlo methods [14], [15]. Our results show that the simplified criterion in (11) leads to good results and, by exploiting the equiprojectivity results and the SR assumption introduced in the previous section, can be implemented in near real time. An alternative scheme was proposed in [13], in which O_k is estimated via an iterative (EM) algorithm initialized with $\widehat{\mathrm{O}}_{k-1}$.

5.2 Likelihood Function

In this section, to simplify the notation, we will omit the time index k and derive the likelihood function $P(\nabla\mathrm{I}|\mathrm{O})$ for a generic image. Let $\mathbf{E}_\mathbf{u} = (E_\mathbf{u}, \phi_\mathbf{u})$ denote the element of the image gradient $\nabla\mathrm{I}$ at pixel $\mathbf{u}$, where $E_\mathbf{u}$ is the gradient magnitude and $\phi_\mathbf{u}$ the gradient direction. As in [10], [11], the likelihood function is derived as follows:

- Each pixel $\mathbf{u}$ has a class label $m_\mathbf{u} \in \{1, 2, 3, 4, 5\}$. Pixels in classes 1, 2, 3 belong to edges consistent with the $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$

axes, respectively. Pixels in class 4 are on edges not consistent with those axes. Nonedge pixels are in class 5. These classes have prior probabilities $\{P(m_{\mathbf{u}})\}$ (we adopt the values used in [10], [11]).

- The gradient magnitude and direction are conditionally independent, given the class label. Naturally, the gradient magnitude is also conditionally independent of the camera orientation and of the pixel location. Thus,

$$P(\mathbf{E_u}|m_{\mathbf{u}},\mathrm{O},\mathbf{u}) = P(E_{\mathbf{u}}|m_{\mathbf{u}})\,P(\phi_{\mathbf{u}}|m_{\mathbf{u}},\mathrm{O},\mathbf{u}), \tag{12}$$

where

$$P(E_{\mathbf{u}}|m_{\mathbf{u}}) = \begin{cases} P_{\mathrm{on}}(E_{\mathbf{u}}), & \text{if } m_{\mathbf{u}} \neq 5 \\ P_{\mathrm{off}}(E_{\mathbf{u}}), & \text{if } m_{\mathbf{u}} = 5, \end{cases} \tag{13}$$

and $P_{\mathrm{on}}(E_{\mathbf{u}})$ and $P_{\mathrm{off}}(E_{\mathbf{u}})$ are the probability mass functions of the quantized gradient magnitude, *conditioned* on whether pixel $\mathbf{u}$ is *on* or *off* an edge, respectively. These probabilities are learned offline.

- Let $\theta_x(\mathrm{O},\mathbf{u})$, $\theta_{\mathbf{y}}(\mathrm{O},\mathbf{u})$, $\theta_z(\mathrm{O},\mathbf{u})$ be the gradient directions that would be ideally observed at location $\mathbf{u}$ if $m_{\mathbf{u}} = 1, 2, 3$, respectively. The gradient direction probability function is

$$P(\phi_{\mathbf{u}}|m_{\mathbf{u}},\mathrm{O},\mathbf{u}) = \begin{cases} P_{ang}(\phi_{\mathbf{u}} - \theta_{\mathbf{x}}(\mathrm{O},\mathbf{u})) & \Leftarrow m_{\mathbf{u}} = 1 \\ P_{ang}(\phi_{\mathbf{u}} - \theta_{\mathbf{y}}(\mathrm{O},\mathbf{u})) & \Leftarrow m_{\mathbf{u}} = 2 \\ P_{ang}(\phi_{\mathbf{u}} - \theta_{\mathbf{z}}(\mathrm{O},\mathbf{u})) & \Leftarrow m_{\mathbf{u}} = 3 \\ U(\phi_{\mathbf{u}}) & \Leftarrow m_{\mathbf{u}} = 4, 5, \end{cases} \tag{14}$$

where

$$P_{ang}(t) = \begin{cases} \frac{1-\epsilon}{2\tau} & \Leftarrow t \in [-\tau,\tau] \\ \frac{\epsilon}{\pi-2\tau} & \Leftarrow t \in \,]-\pi/2,-\tau[\,\cup\,]\tau,\pi/2], \end{cases}$$

and $U(\cdot)$ is the uniform pdf on $]-\frac{\pi}{2},\frac{\pi}{2}]$. In our experiments, we use $\epsilon = 0.1$, and $\tau = 4^\circ$.

- Finally, the joint likelihood is obtained by marginalizing (summing) over all possible models at each pixel and assuming independence among different pixels:

$$P(\nabla\mathrm{I}|\mathrm{O}) = P(\{\mathbf{E_u}\}|\mathrm{O}) = \prod_{\mathbf{u}} \sum_{m_{\mathbf{u}}=1}^{5} P(E_{\mathbf{u}}|m_{\mathbf{u}})\;P(\phi_{\mathbf{u}}|m_{\mathbf{u}},\mathrm{O},\mathbf{u})\;P(m_{\mathbf{u}}). \tag{15}$$

5.3 Locating the Estimates

The maximization in (11), with the likelihood function (15), is a three-dimensional optimization problem with respect to α, β, and γ. We propose an approximate solution which decouples the problem into two simpler steps: a two-dimensional optimization w.r.t. β and γ, followed by a one-dimensional search w.r.t. α. This approximation is supported on the fact that the vanishing point $\mathbf{v_z}$ does not depend on the compass angle α, as is clear from (2).

In the first step, we estimate β and γ, for frame k, according to

$$\left(\widehat{\beta}_k,\widehat{\gamma}_k\right) = \arg\max_{\beta,\gamma}\Big\{\log P(\{\mathbf{E_u}\}_k|\beta,\gamma) + \log P(\beta,\gamma|\widehat{\beta}_{k-1},\widehat{\gamma}_{k-1})\Big\}, \tag{16}$$

where the likelihood $P(\{\mathbf{E_u}\}_k|\beta,\gamma)$ is a version of (15) which only models direction information of edges consistent with the $\mathbf{z}$ axis. More specifically, instead of (14), we use here

$$P(\phi_{\mathbf{u}}|m_{\mathbf{u}},\beta,\gamma,\mathbf{u}) = \begin{cases} P_{ang}(\phi_{\mathbf{u}} - \theta_{\mathbf{z}}(\beta,\gamma,\mathbf{u})) & \Leftarrow m_{\mathbf{u}} = 3 \\ U(\phi_{\mathbf{u}}) & \Leftarrow m_{\mathbf{u}} = 1,2,4,5. \end{cases} \tag{17}$$

Notice that the use of a uniform distribution is simply a way of ignoring angle information from all pixels but those corresponding to the $\mathbf{z}$ axis ($m_{\mathbf{u}} = 3$), when estimating β_k and γ_k; it doesn't mean that those angles are actually uniformly distributed.

$P(\beta,\gamma|\widehat{\beta}_{k-1},\widehat{\gamma}_{k-1})$ is a truncated bivariate Gaussian with mean $[\widehat{\beta}_{k-1},\widehat{\gamma}_{k-1}]^T$, defined over the region $\beta \in\,]\widehat{\beta}_{k-1} - \xi, \widehat{\beta}_{k-1} + \xi]$ and $\gamma \in\,]\widehat{\gamma}_{k-1} - g_\xi(\widehat{\beta}_{k-1}), \widehat{\gamma}_{k-1} + g_\xi(\widehat{\beta}_{k-1})]$. This prior formalizes the SR assumption (see (5) and (8)) as well as angle variation smoothness. The variance of this Gaussian controls the trade-off between the smoothness of the estimated sequence of angles and the accuracy of this estimates. In the first frame, the prior is flat over the entire domain $(\beta,\gamma) \in\,]-45^\circ,45^\circ] \times\,]-54.7^\circ,54.7^\circ]$, according to (4).

Given $\widehat{\beta}_k$ and $\widehat{\gamma}_k$, we then estimate the compass angle α_k using

$$\widehat{\alpha}_k = \arg\max_{\alpha}\Big\{\log P(\{\mathbf{E_u}\}|\alpha,\widehat{\beta}_k,\widehat{\gamma}_k) + \log P(\alpha|\widehat{\alpha}_{k-1},\widehat{\beta}_{k-1},\widehat{\beta}_k)\Big\}, \tag{18}$$

where the prior $P(\alpha|\widehat{\alpha}_{k-1},\widehat{\beta}_{k-1},\widehat{\beta}_k)$ is a truncated Gaussian with mean $\widehat{\alpha}_{k-1}$, defined over the interval $]\widehat{\alpha}_{k-1} - a_\xi(\widehat{\beta}_k,\widehat{\beta}_{k-1}), \widehat{\alpha}_{k-1} + a_\xi(\widehat{\beta}_k,\widehat{\beta}_{k-1})]$ (see (6)). For the first frame, the prior is flat over $]-45^\circ,45^\circ]$. The maximizations in (16) and (18) are carried out by exhaustive search.

If a given estimate $\widehat{\mathrm{O}}_k(\widehat{\alpha}_k,\widehat{\beta}_k,\widehat{\gamma}_k)$ is located outside of the minimal region defined in (4), we replace it by an equiprojective orientation inside that region. As explained in the last paragraph of Section 4, this allows $a_\xi(\widehat{\beta}_k,\widehat{\beta}_{k-1})$ to be less than 7.1°, hence keeping a small search space. As a final step, at each frame k, we select an orientation from the equivalence class of $\widehat{\mathrm{O}}_k$, such that the resulting sequence satisfies the SR model.

6 Experiments

The algorithm was tested with outdoor MPEG-4 video sequences, acquired with a hand-held camera. Although the sequences are of low quality due to radial distortion and several over and underexposed frames, our algorithm was able to successfully estimate the camera orientation, as illustrated in Fig. 5.

The images in Figs. 6 and 7 show frames from two other sequences. Notice that the algorithm is able to estimate the correct orientation, despite the many edges not aligned with the MW axes (e.g., people in Fig. 7). The plots in the same figures represent the estimates of the orientation angles, for these two sequences. Note that the estimates on the plot of Fig. 7 are slightly noisier than those in Fig. 6, due to the lower image quality. The smoothness of these estimates is controlled by the prior variances referred to in Section 5.3; here, these variances are the same for both sequences and the three angles. Of course, there is a trade-off between smoothness and ability to accurately follow fast camera rotations.

Typical processing time for each (288×360)-pixels frame is below one second, on a 3.0~GHz Pentium IV, using a MATLAB implementation. The only effort made to speed up the computation was the exclusion of nonrelevant pixels by nonmaxima suppression followed by thresholding of the gradient magnitude. We are currently working on a C implementation to achieve frame-rate.

7 Conclusion

We have proposed a probabilistic approach to estimating camera orientation from video sequences of urban scenes. The method avoids standard intermediate steps such as feature detection and correspondence or edge detection and linking. Experimental results show that the method is able to handle low-quality video sequences, even with many spurious edges.

Appendix

Here, we prove Proposition 3. From (1)-(2), we have (for $i, j \in \{x,y,z\}$):

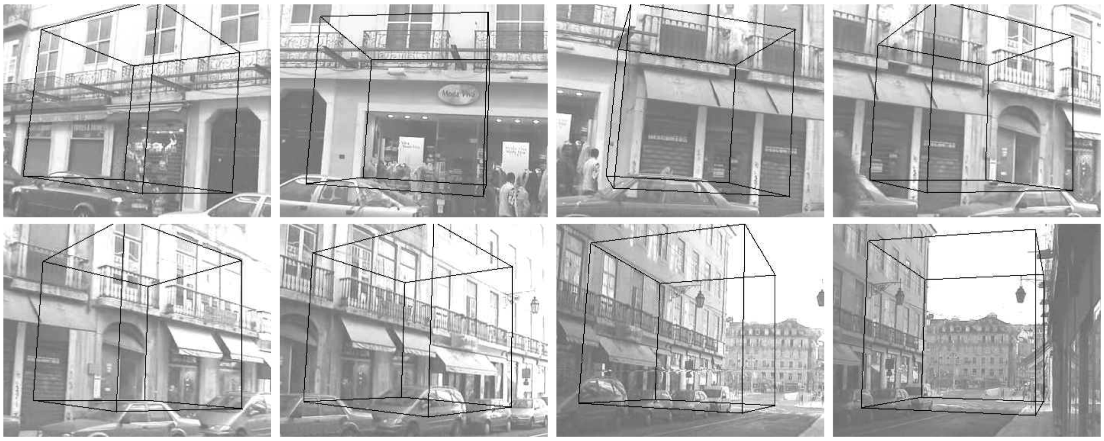

Fig. 5. Orientation estimates (superimposed cubes represent the estimated MW axes) for the first and several other frames of a video sequence.

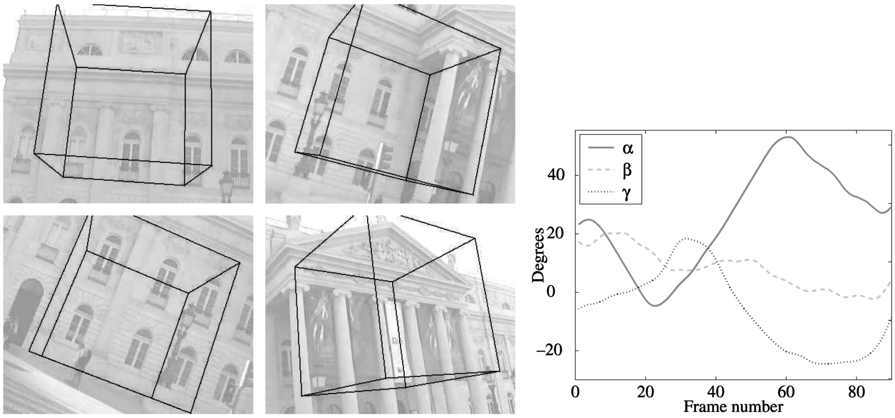

Fig. 6. Left: Frames 20, 30, 40, and 50 of another video sequence. Right: Camera angle estimates.

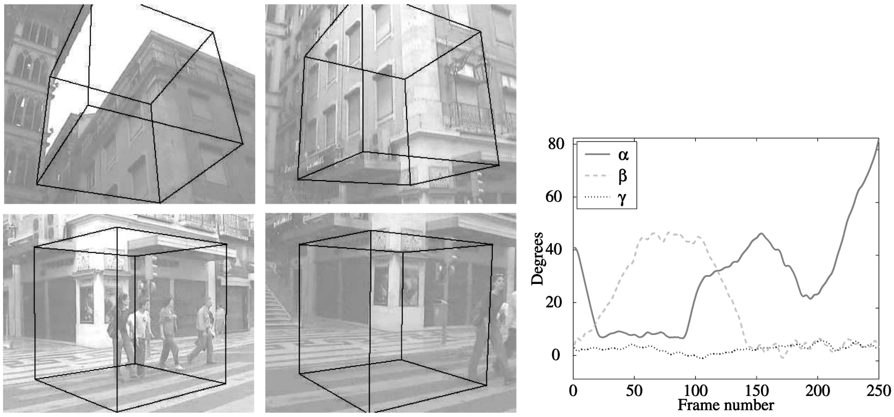

Fig. 7. Left: Frames 110, 130, 150, and 170 of a third video sequence. Right: Camera angle estimates.

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} f^2\left(\frac{f^2}{P_i^2} - 1\right) & \Leftarrow i = j \\ -f^2 & \Leftarrow i \neq j, \end{cases} \tag{19}$$

which gives us both the Euclidean distance $d_i = (\mathbf{v}_i^T \mathbf{v}_i)^{1/2}$ between points $\mathbf{v}_i$ and $\mathbf{p} = (0,0)^T$ and the angle $\theta_{ij} = \mathrm{acos}\frac{\mathbf{v}_i^T \mathbf{v}_j}{d_i d_j}$ formed by the two lines $[\mathbf{p}\mathbf{v}_i]$ and $[\mathbf{p}\,\mathbf{v}_j]$, with $i \neq j$.

Consider now the disk $\mathcal{D}$ with radius f centered at $\mathbf{p}$, i.e., $\mathcal{D} = \{(u,v) \in \mathbb{R}^2 : u^2 + v^2 \leq f^2\}$. We have $\mathbf{v}_i \in \mathcal{D}$ iff $d_i \leq f$, which, by (19), is equivalent to $P_i^2 \geq f^2/2$. Since $P_x^2 + P_y^2 + P_z^2 = f^2$, the condition $P_i^2 \geq f^2/2$ implies that $P_j^2 \leq f^2/2$ for any $j \neq i$, which means that there cannot exist more than one VP in the interior of disk $\mathcal{D}$. Furthermore, the three VPs are all in the exterior or at the boundary of $\mathcal{D}$ iff $P_i^2 \leq f^2/2$, for $i \in \{x, y, z\}$.

To complete our proof, we need the following intermediate result:

Proposition 6. *Any two VPs $\mathbf{v}_i$ and $\mathbf{v}_j$, with $i \neq j$, verify $\cos\theta_{ij} \leq 0$. Furthermore, if $\mathbf{v}_k \in \mathcal{D}$, with $k \neq i$ and $k \neq j$, then $\cos\theta_{ij} \geq -\frac{1}{3}$.*

Proof. The first statement comes directly from (19). To prove the second statement, we obtain $\min\cos\theta_{ij} = -\frac{f^2}{\min d_i d_j}$, w.r.t. the principal point coordinates P_i and P_j, over the domain defined by $P_i^2 + P_j^2 \leq f^2/2$. The minimum occurs for $|P_i| = |P_j| = \frac{f}{2}$ with value $-1/3$.

Since $\frac{1}{2}\mathrm{acos}(-\frac{1}{3}) = \mathrm{atan}\sqrt{2} \approx 54.7^\circ$,the shaded area in Fig. 3 is a simple consequence of Proposition 6. To show (4), consider an orientation O and let $\mathbf{v}_i$ be a VP in this shaded area. Proposition 2 then guarantees the existence of an equiprojective orientation O^*, satisfying: 1) $\mathbf{v}^*_\mathbf{z} = \mathbf{v}_i$ and 2) $d^*_x \leq d^*_y$. From (2)-(3), we have, due to 1), that $\beta^* \in]-\pi/2, \pi/2]$ and $\gamma^* \in]-\mathrm{atan}\sqrt{2}, \mathrm{atan}\sqrt{2}]$ and, due to 2), that $\alpha^* \in]-\pi/2, \pi/2]$.

Acknowledgments

An early version of this work appeared in [1]. This work was partially supported by the (Portuguese) Foundation for Science and Technology, grants POSI/SRI/41561/2001 and POSI/33143/SRI/2000.

References

[1] A. Martins, P. Aguiar, and M. Figueiredo, "Navigating in Manhattan: 3D Orientation from Video without Correspondences," *Proc. IEEE Int'l Conf. Image Processing,* 2003.

[2] O. Faugeras, *Three-Dimensional Computer Vision.* Cambridge, Mass.: MIT Press, 1993.

[3] R. Hartley and A. Zisserman, *Multiple View Geometry.* Cambridge Univ. Press, 2000.

[4] *Geometric Invariants in Computer Vision,* J. Mundy and A. Zisserman, eds. Cambridge, Mass.: MIT Press, 1992.

[5] E. Lutton, H. Maitre, and J. Lopez-Krahe, "Contribution to the Determination of Vanishing Points Using Hough Transform," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 4, pp. 430-438, Apr. 1994.

[6] S. Utcke, "Grouping Based on Projective Geometry Constraints and Uncertainty," *Proc. IEEE Int'l Conf. Computer Vision,* 1998.

[7] J. Kosecka and W. Zhang, "Video Compass," *Proc. European Conf. Computer Vision,* 2002.

[8] B. Horn and E. Weldon Jr., "Direct Methods for Recovering Motion," *Int'l J. Computer Vision,* vol. 2, no. 1, pp. 51-76, 1988.

[9] G.P. Stein and A. Shashua, "Model-Based Brightness Constraints: On Direct Estimation of Structure and Motion," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 9, pp. 992-1015, Sept. 2000.

[10] J. Coughlan and A. Yuille, "Manhattan World: Compass Direction from a Single Image by Bayesian Inference," *Proc. IEEE Int'l Conf. Computer Vision,* 1999.

[11] J. Coughlan and A. Yuille, "The Manhattan World Assumption: Regularities in Scene Statistics which Enable Bayesian Inference," *Proc. Neural Information Processing Systems,* 2000.

[12] J. Deutscher, M. Isard, and J. MacCormick, "Automatic Camera Calibration from a Single Manhattan Image," *Proc. European Conf. Computer Vision,* 2002.

[13] G. Schindler and F. Dellaert, "Atlanta World: An Expectation-Maximization Framework for Simultaneous Low-Level Edge Grouping and Camera Calibration in Complex Man-Made Environments," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2004.

[14] N. Gordon, A. Doucet, and N. Freitas, *Sequential Monte Carlo Methods in Practice.* New York: Springer-Verlag, 2001.

[15] M. Isard and A. Blake, "CONDENSATION–Conditional Density Propagation for Visual Tracking," *Int'l J. Computer Vision,* vol. 29, no. 3, pp. 5-28, 1998.