A DIFFERENTIAL GEOMETRIC APPROACH FOR VISUAL NAVIGATION IN INDOOR SCENES

L. Fuentes and M. Gonzalo-Tasis
MOBIVA Group
University of Valladolid

G. Bermudez and J. Finat
MOBIVA Group
University of Valladolid

Keywords: Visually Navigation, Motion Analysis, Matching Correspondence and Flow, Kalman Filtering.

Abstract: Visual perception of the environment provides a detailed scene representation which contributes to improve motion planning and obstacle avoidance navigation for wheelchairs in non-structured indoor scenes. In this work we develop a mobile representation of the scene based on perspective maps for the automatic navigation in absence of previous information about the scene. Images are captured with a passive low-cost video camera. The main feature for visual navigation in this work is a map of quadrilaterals with apparent motion. From this mobile map, perspective maps are updated following hierarchical grouping in quadrilaterals maps given by pencils of perspective lines through vanishing points. Egomotion is interpreted in terms of maps of mobile quadrilaterals. The main contributions of this paper are the introduction of Lie expansion/contraction operators for quadrilateral/cuboid and the adaptation of Kalman filtering for moving quadrilaterals to estimate and predict the egomotion of a mobile platform. Our approach is enough modular and flexible for adapting to indoor and outdoor scenes provided at least four homologue cuboids be present in the scene between each pair of sampled views of a video sequence.

1 INTRODUCTION

Visually based automatic navigation for platforms is a common subject in Motion Planning and Navigation along the nineties, with a lot of applications, including interactive wheelchairs navigation. This problem has been also developed in several related works (P. Trahanias and Orphanoudakis, 1997), including landmarks (M. Rous and Kraiss, 2001), stereo vision (Mazo and et al, 2002) for navigation robotics. Localization of beacons, simultaneous correspondence between homologue points increase the computational cost and troubles for decision making.

The design of smart wheelchairs with sensor fusion and hybrid control is an active research subject(Mazo and et al, 2002) (Levine et al., 1999) (R. Simpson and Nourbakhsh, 2004) (Yanco, 1998). Two main problems concern to safety tasks based in reactive behavior (Cortés et al., 2003) and navigation-oriented tasks focused towards the generation of environment maps and motion planning (Yuki, 2000). Changing or non-structured environments, and complex interactions with the environment are benefited from visual information. Active and passive sensors can be used for semi-automatic or automatic navigation with wheelchairs (A. Pruski and Morre, 2002). Active sensors are crucial for obstacle avoidance, but a reactive behavior is not enough for complex tasks in unknown environments. An accurate and robust motion planning, requires detailed information about the scene for mapping and localizations tasks (P. Trahanias and Orphanoudakis, 1997). Modeling intelligent behaviors for wheelchairs have received an important attention along last years and it is the main motivation for this work. Nevertheless, the present approach is focused towards the exploitation of information provided by a low-cost video camera. Range sensors require information fusion and often provide incomplete information about changing environment, with some limitations linked to reactive behavior. A low-cost non-calibrated camera mounted in a mobile platform provides global information about the scene, which can be corrected, updated and completed with range sensors for metric information (A. Pruski and Morre, 2002)

Furthermore, to increase the usability in non-structured environments, it is convenient to avoid beacons or previous information about the scene. To decrease the computational effort and to achieve a better adaptation to these applications for disabled per-

Fuentes L., Gonzalo-Tasis M., Bermudez G. and Finat J. (2006).
A DIFFERENTIAL GEOMETRIC APPROACH FOR VISUAL NAVIGATION IN INDOOR SCENES.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 468-476
DOI: 10.5220/0001378104680476
Copyright © SciTePress

sons, we shall suppose that environments are given by indoor or outdoor architectural scenes. In ordinary architectural scenes it is easy to identify objects whose boundaries support elements for generating "perspective maps" given by vanishing points, perspective lines and perspective planes. A methodology for generating perspective maps is developed in the §3. Perspective maps provide a $2\frac{1}{2}$ reconstruction. Most of cameras can be considered as perspective devices which project some part of the visible scene on the camera plane.

Visual navigation is planned depending on the selection of visual targets. Visual targets must be real-time located and updated in a faithful representation of the scene. To simplify, in this work we suppose that the identification of visual targets or tasks to be achieved is externally performed by the user; otherwise, a recognition and a motion planning module must be added for decision making. A difficult problem concerns to the representation updating for the workspace. In this work, the chosen representation is given by the lifting of quadrilaterals maps $\mathcal{Q}$. A quadrilateral map is a $2d$ quadrangular mesh adapted to a perspective representation of the scene (M. Gonzalo and Aguilar, 2002) generated by the intersection of pencils of perspective lines through vanishing points.

Edges of $\mathcal{Q}$ lie on visible support whose boundaries provide a support for perspective lines. The computer management of the image information contained in each view is performed in terms *visible* elements. The information management is performed on a symbolic representation given by a small subgraph of a quadtree corresponding to visible elements in each view. A characterization of simple events linked to quadrilaterals of $\mathcal{Q}$ simplifies the search of homologue elements. From the motion viewpoint, evolving homologue quadrilaterals give information about the magnitude and direction of motion changes, involving the mobile platform itself, and other external agents.

Usual accurate volumetric representations have a high computational cost, and it is difficult to obtain a faithful representation of the scene which are updated on-line (twice per second). Corridor scenes are used in the experimental set-up. In this case, by taking a mobile reference centered in the mobile object, at most two vanishing points (v_z, v_x, e.g.) are at finite distance, and at least a third one v_y is at infinite distance (vertical lines must be parallel). The simple nature of analyzed scenes allows to generate maps of cuboids by intersecting pencils of planes through the three vanishing lines connecting each pair of vanishing points.

The information management in terms of octrees has in general a complexity $O(N^2\ log\ N)$ in the number N of planes of the scene. However, the simple nature of the corridor scene, allows to generate a mobile perspective representation of the volumetric scene. Each 3d perspective representation of the whole scene is obtained by lifting the quadrilateral map $\mathcal{Q}$ to a $3D$ model. The ordered lifting of the quadrilateral map has a complexity at most $O(N \log\ N)$ linked to ordering planes contained in perspective maps which are meaningful for visibility issues. The incorporation of a graphics card would allow to avoid this increasing thanks to the use of a typical z-buffer algorithm. However, for lowering costs, a simplified perspective representation is introduced, which reduces the computation to visible cuboids. The resulting maps of cuboids $\mathcal{C}$ play a very similar role to the maps of quadrilaterals. To give a representation of contraction/expansion of cuboid/quadrilateral maps we introduce Lie contraction/expansion operators along motion directions. The construction of contraction/expansion operators between maps of cuboids $\mathcal{C}$ and maps of quadrilaterals $\mathcal{Q}$ requires a robust estimation of vanishing points and the ego-motion description in terms of maps of quadrilaterals, which is the main contribution of this work. To achieve it, we use a variant of Kalman Filtering (Marion, 2002)

Kalman filtering is a tool for control of mobile systems, including motion estimation, tracking, and prediction from estimation. They have been used in Motion Analysis by Computer Vision, in particular to provide an assistance for visually guided automatic navigation. A Kalman filter (Faugeras, 1993) is a recurrent technique for solving a dynamic problem by the least squares method. Measures can be corrupted by white noise, and must be corrected. In this work, an adaptation of Kalman filtering is developed for maps of quadrilaterals, including an implementation in C++ for motion estimation and tracking in architectural indoor scenes (M. Gonzalo and Aguilar, 2002) by developing some aspects appearing in (Marion, 2002)

The paper is organized as follows: We start with a very short review of related approaches. Next, we develop some elements for modeling the scene. The fourth section is devoted to the motion analysis and to sketch the adaptation of Kalman filtering to maps of quadrilaterals.

2 RELATED APPROACHES

The design of smart wheelchairs with sensor fusion and hybrid control is an active research subject along last ten years ((Levine et al., 1999), (Mazo and et al, 2002), (R. Simpson and Nourbakhsh, 2004), (Yanco, 1998)). Two main problems concern to safety tasks based in reactive behavior following agent-based technologies (Cortés et al., 2003) and navigation-oriented tasks focused towards the generation of en-

vironment maps and motion planning ((A. Pruski and Morre, 2002), (Yuki, 2000)). The integration of embedded systems for real-time sensor-based navigation is a challenge specially in presence of human interaction (J. Minguez and Montano, 2004).

Changing or non-structured environments, and complex interactions with the environment are benefited from visual information. Visually based automatic navigation for platforms is a common subject in Robotics Navigation along the nineties. Its application to wheelchairs has been also developed in several related works (P. Trahanias and Orphanoudakis, 1997), including landmarks (M. Rous and Kraiss, 2001), stereo vision (Mazo and et al, 2002), between others. Localization of beacons, simultaneous correspondence between homologue points increase the computational cost and decision taking. The nearest approach to ours is (Toedemé and Gool, 2004). In our work the accent is put in explaining the generation and updating of the geometric model, by sketching algorithms for grouping around geometric primitives of positive dimension. Often, beacons or "salient" corners can be partially occluded for relative localizations of the mobile platform. Thus, it is advisable to work with geometric primitives supported on lines or quadrilaterals, instead of points. Robustness of perspective models suggests the choice of $2d$ perspective maps as the initial support for relative localization and tracking of apparent motion of the platform.

3 MODELLING THE SCENE

On-line generation and updating of a perspective map provide us to get a model of the scene. The experimental set-up is as follows: To imitate arbitrary motions of a disabled person mounted in a wheelchair, a mobile platform is manually displaced along a corridor with diffuse light and non-conditioned material in an arbitrary (sometimes erratic) way . Kinematic characteristics of the trajectories are unknown. The mobile platform has a low-cost uncalibrated video camera providing low-quality images, which are sampled twice each second.

3.1 Perspective Models

Lines of the ordinary space provide the main features for perspective models. The set of space lines through an eventually mobile point (focus of the camera) is a homogeneous compact manifold called a projective plane $\mathbb{P}^2$. Often, in architectural scenes, some "meaningful lines" can be grouped in pencils through each "vanishing point". Structural elements for a perspective map of the scene are given by vanishing points $\mathbf{V}_i$, perspective lines PL_j and perspective planes PP_k. Vanishing points $\mathbf{v}_i$ are the intersection locus of at least four perspective lines PL_j parallel in the scene (three lines can converge in a corner which is not a vanishing point in general). In complex architectural scenes there can be a lot of vanishing points, due to the misalignment of walls or faades, in indoor or outdoor scenes, respectively. For clarity purposes, we restrict ourselves to simpler cases of indoor scenes with at least three independent vanishing points.

In frontal views of a large number of architectural simple scenes, there are three main vanishing points. Typical frontal views have two fixed vanishing points at infinite distance, labelled as $\mathbf{v}_x$ and $\mathbf{v}_y$, corresponding to the intersection of pencils of horizontal and vertical lines, respectively. Usually, the vanishing point $\mathbf{v}_y$ is fixed. However, rotations of the camera give displacements of $\mathbf{v}_x$, providing the main input for the motion estimation. Furthermore, even for rectified views, turning a corridor generates abrupt changes in the identification of perspective elements involving the localization of $\mathbf{v}_x$.

Figure 1: Perspective Model and Quadrilateral Map.

Collinear segments are grouped to obtain "large" lines ℓ_j with a length ℓ (number of collinear mini-segments) larger than a threshold selected by the user. So, a collection of allowed lines ℓ_{j_i} is obtained as candidates for perspective lines through the vanishing point $\mathbf{v}_i$. Candidate perspective lines are obtained by an automatic grouping of segments, and a selection of pencils Λ of lines which converge in a vanishing point $\mathbf{v}_i$. Validation of $\mathbf{v}_i$ is performed by minimizing $\sum(\ell_{j_i}^T \mathbf{v}_i)^2$. After estimating each vanishing point $\mathbf{v}_i$, candidate perspective lines are replaced by new perspective lines to provide a "coarse rectification" linked to a globally coherent perspective representation of the scene; this step is necessary for correcting views provided by the low cost video camera.

Each pair of vanishing points $\mathbf{V}_i, \mathbf{V}_j \in \mathbb{P}^3$ determines an *horizon line* $L_{\infty,ij} = \overline{\mathbf{V}_i \mathbf{V}_j} \subset \mathbb{P}^3$ which is computed by the cross-product $\mathbf{V}_i \wedge \mathbf{V}_j$; any point on $L_{\infty,ij}$ is also a vanishing point. Often, visual target

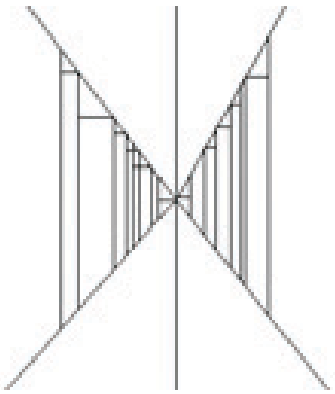

Figure 2: Perspective Model and Quadrilateral Map.

is located in $L_{\infty,ij}$, and the visual perception of egomotion in the image is represented by an apparent displacement along the horizon line $L_{\infty,ij}$; in this case, it suffices to track this displacement for egomotion estimation. To achieve a $2\frac{1}{2}$ reconstruction as general framework for visual navigation, a projective representation of the plane at infinity is required.

Three non-collinear vanishing points $\mathbf{V}_i$, $\mathbf{V}_j$, $\mathbf{V}_k$ generate a *plane at infinity* in $\mathbb{P}^3$ which is denoted as Π_∞ without specifying the choice of vanishing points. Thus, π_∞ can be taken as an invariant reference plane for motion estimation and tracking. Similarly, pencils $\mathcal{L}_i$, $\mathcal{L}_j$, $\mathcal{L}_k$ or perspective lines through the vanishing points $\mathbf{V}_i$, $\mathbf{V}_j$, $\mathbf{V}_k$ give a dual reference.

Let us denote by $\mathcal{L}_i$ a pencil of perspective lines through the vanishing point $\mathbf{v}_i$. Two pencils , $\mathcal{L}_i$, $\mathcal{L}_j$ of coplanar perspective lines determine a perspective plane which is denoted by $\mathcal{L}_i + \mathcal{L}_j$. The perspective representation of an ordinary architectural scene is visualized by means of three pencils $\mathcal{L}_h$, $\mathcal{L}_v$, $\mathcal{L}_d$ of horizontal, vertical and depth perspective lines. This representation is a local version of the projective representation, and it can be interpreted as a dual reference system for the projective plane which extends the plane of each view. A perspective representation of the scene in terms of perspective lines is more robust than representations based on points.

There are *three basic perspective models* which are labeled as frontal, angular and skew depending on the availability of one, two or three vanishing points at finite distance. If there is no vanishing point at finite distance we have a parallel projection, which is not a perspective model, properly said. A corridor scene (resp. room scene) is represented by a frontal (resp. angular) perspective model, whereas in turning a corner of a corridor an angular perspective model provides the transition between two frontal perspective model. In outdoor architectural scenes, the localization of the camera gives skew perspective models for the transition between angular perspective models.

3.2 Quadrilaterals Map

The low quality of conventional video cameras and the bad illumination conditions, impose serious restrictions for achieving robust results in real-time, including an on-line localization for a mobile platform. Visual perception of rectangular regions is more robust than corners surrounding them. In frontal perspective models one has at least a vanishing point $\mathbf{v}_z$ at finite distance. In this model, rectangular regions are ideally perceived as trapezoids, i.e., as quadrilaterals with two parallel vertical edges. To simplify the tracking of quadrilaterals, we restrict ourselves to horizontal and vertical segments. Hence, the first task is the construction and tracking of two trapezoidal maps linked to $\mathbf{v}_z$. There are two *extremal perspective lines* $PL_{M,z}$, $PL_{m,z}$ through $\mathbf{v}_z$ with maximal M and minimal m slope and affine equations $\ell_{M,z} = 0$, $\ell_{m,z} = 0$, respectively. Both perspective lines decompose the plane in four signed regions, depending on the sign evaluation for affine equations at each point. Vertical (respectively, horizontal) segments contained in regions $(+,-)$ and $(-,+)$ (respectively, in regions $(+,+)$ and $(-,-)$) are extended till to arrive to extremal perspective lines. So, we generate two *trapezoidal maps* $\mathcal{T}_{z,ver}$ and $\mathcal{T}_{z,hor}$ linked to the vanishing point $\mathbf{v}_z$. Both trapezoidal maps are bounded by extremal perspective lines with vertical and horizontal parallel segments, respectively. After retracing extremal perspective lines, each trapezoid of a trapezoidal map linked to a vanishing point is characterized by three points. For each oriented trapezoid $T_{i,j} \in \mathcal{T}_{z,j}$ with $j = ver$ or $j = hor$, we associate in a canonical way a pair of bivectors representing the trapezoid uniquely; a bivector is the cross-product of two ordinary vectors.

The same argument is applied for angular and skew perspective models, but by replacing the trapezoid by a general quadrilateral, i.e is the image of a rectangle by an affine transformation, whose opposite edges are supported on perspective lines through the same vanishing point. Two coplanar pencils Λ_i and Λ_j of projective lines through the vanishing points $\mathbf{V}_i$, $\mathbf{V}_j$ give a *quadrilateral map*. The quadrilateral map $\mathcal{Q}$ is supported by a plane through the vanishing line $L_{\infty,ij} := \overline{\mathbf{V}_i\mathbf{V}_j}$. Data structures for a quadrilateral map $\mathcal{Q}$ are supported on visible segments lying in two coplanar pencils $\mathcal{L}_i$, $\mathcal{L}_j$ of perspective lines. Thus, for a typical architectural scene, we have three families of maps of quadrilaterals $\mathcal{Q}_{12}$, $\mathcal{Q}_{13}$, $\mathcal{Q}_{23}$ linked to pencils of perspective planes through the vanishing lines $L_{\infty,12}$, $L_{\infty,13}$, $L_{\infty,23}$. A sheaf depending on two (resp. three) parameters is called a net (resp. web). Hence, by packing quadrilateral maps a web of quadrilaterals is obtained.

The basic cell of a a quadrilateral map is given by an ordered quadrilateral $\mathbf{Q}_{ij}$ with edges

$\mathbf{a}_{i1}, \mathbf{b}_{j1}, \mathbf{a}_{i2}, \mathbf{b}_{j2}$. Edges are supported on two pairs of lines lying on pencils of perspective lines; so we have $\mathbf{a}_{i1} \subset \ell_{i1} \in \mathcal{L}_i$, $\mathbf{b}_{j1} \subset \ell_{j1} \in \mathcal{L}_j$ and so on. The *oriented area* for the quadrilateral $\mathbf{Q}_{ij}$ is given by the bivector

$$\mathcal{A}(\mathbf{Q}_{ij}) = \frac{1}{2}[\mathbf{a}_{i1} \times \mathbf{b}_{j1} + \mathbf{a}_{i2} \times \mathbf{b}_{j2}] \qquad (1)$$

In particular, if $\mathcal{Q}$ is a map of quadrilaterals for a frontal perspective model, then each quadrilateral is a trapezoid. Quadrilateral maps are referred in projective representation with respect to the common vanishing point (corresponding to the preservation of vertical direction), and a rotation of the horizon line opposite to the vanishing point. In the general case, rotation is computed following the methodology described in (Jelinek and Taylor, 2001).

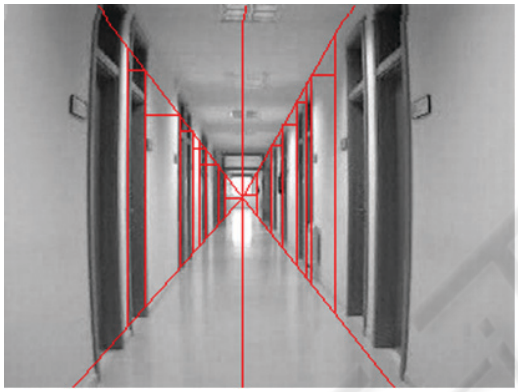

Figure 3: Perspective Model and Quadrilateral Map.

Turning a corridor implies severe alterations in maps of quadrilaterals, and it is necessary to have a $3d$ model for integrating turning quadrilateral maps in a common framework. The proposed common framework for the on-line management of $2\frac{1}{2}d$ information is given by a map of cuboids. A cuboid is given by the cross product of three *linearly independent vectors*; a parallelepiped is a cuboid given by the triple product of three *orthogonal* vectors. In architectural scenes, edges (respectively, faces) of cuboids are supported on perspective lines (respectively, planes). The intersection of two adjacent cuboids is a face of one of them. Modeling of cuboid maps $\mathcal{C}$ is similar to modeling of trapezoidal maps, but increasing the dimension in a unity. So, instead of considering pencils Λ_i of lines through vanishing points $\mathbf{V}_i$, it is necessary to consider pencils $\mathcal{L}_{k\ell}$ of planes through vanishing lines $\overline{\mathbf{V}_k\mathbf{V}_\ell}$. Computer management is also similar, but by replacing quadtrees by octrees. To avoid excessive subdivisions, only non-empty cellules are subdivided depending on the identification of visible perspective elements in the 3d scene. Furthermore, only local information is preserved in short memory whereas some part of the scene is visible. In this way, a verification of global coherence is avoided in updating 3d models.

4 LIE OPERATORS FOR EXPANDING/CONTRACTING MAPS

Motion estimation from optical motion is not well-determined. It is necessary to add constraints, predictions and measurements about the scene or the motion, which must be estimated, validated and updated. In this section, a differential model is provided for a simultaneous evaluation of the motion and the scene structure from the updating of perspective representation of the scene. The key tool is the introduction of simple transformations between perspective maps which can be read in terms of maps of quadrilaterals and cuboids. Some additional concepts are needed for showing the main steps for both operations

4.1 Vector Fields for Motion Estimation

A *smooth vector field* X is a section of the tangent bundle $T\mathbb{R}^3 \simeq \mathbb{R}^3 \times \mathbb{R}^3$, i.e., a representation of the velocities field at each point. Hence, it is given by $a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y} + c\frac{\partial}{\partial z}$, where a, b, c are smooth functions or more synthetically as $\sum a_i \frac{\partial}{\partial x_i}$.

The *optical flow* is a $2d$ vector field θ which is the projection of the true $3d$ motion field. The *egomotion* is a piecewise smooth vector field ξ providing a estimation of the optical flow along a temporal sequence of views with integral curve parametrized by λ.

Our differential geometric approach to the estimation of the $3d$ motion vector field requires 1) the estimation of the $2d$ egomotion vector field ξ, the expression of $2d$ optical flow θ from the $2d$ egomotion vector field ξ and the lifting of θ to the $3d$ motion field by using the updating of the perspective model of the indoor scene.

4.2 Differential Forms for the Perspective Representation of the Scene

A differential form ω is a section of the cotangent bundle $\Omega^1\mathbb{R}^3 := HOM(T\mathbb{R}^3, \varepsilon^1_{\mathbb{R}^3}) \simeq \mathbb{R}^3 \times \mathbb{R}^3$ where $\varepsilon^1_{\mathbb{R}^3} := \mathbb{R}^3 \times \mathbb{R}$ is the trivial vector bundle of rank 1 on $\mathbb{R}^3$. A differential form $\omega \in \Omega^1$ is given by $\sum b_i dx_i$. The evaluation $\omega(X)$ of a differential form on a field

gives a real number. Similarly, Ω^1_c denotes the first degree differential forms with compact support.

The canonical bases dx_1, dx_2, dx_3 and $\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}$ are dual between them, i.e., $dx_i(\frac{\partial}{\partial x_j}) = \delta_{ij}$ (delta de Kronecker) for $1 \leq i, j \leq 3$. In particular, dz vanishes $\frac{\partial}{\partial x}$, and $\frac{\partial}{\partial y}$. Thus, in an ideal euclidian model where z represents the depth, it suffices to make adjustments with respect to the component c of $\frac{\partial}{\partial z}$. Unfortunately, the perception model of any video camera is not euclidian. Thus, some additional considerations are required.

A map of Segments (resp. Quadrilaterals, resp. Cuboids), is given as the support of a pencil of 1-degree (resp. a net of 2-degree, resp. a web of 3-degree) differential form with compact support $\omega \in \Omega^k_c(\mathbb{R}^3)$ of degree k for $k = 1$ (resp $k = 2$, resp. $k = 3$. The choice of compact support for differential forms is justified by practical and theoretical reasons: the objects to be tracked (quadrilaterals) are compact and the need of representing volumetric elements by "non-trivial" models from the cohomological viewpoint. To fix ideas, a static simple corridor scene is modeled in a frontal perspective view with vertical and horizontal trapezoidal maps corresponding to lateral walls and floor-roof, respectively. Pencils of lines belonging to each trapezoidal map verify incidence conditions with respect to vanishing points. For example, the vertical left trapezoidal map is determined by two pencils of lines through the vanishing points $\mathbf{v_y}$ for parallel vertical lines and $\mathbf{v_z}$ for depth map. Similarly, the horizontal lower trapezoidal map is determined by two pencils of lines through the vanishing points $\mathbf{v_x}$ for parallel horizontal lines and $\mathbf{v_z}$ for depth map.

The real motion of the mobile platform generates an apparent displacement of the vanishing points $\mathbf{v_x}$ and $\mathbf{v_z}$ (vertical direction is preserved). The apparent displacement of vanishing points generates displacements of vertical and horizontal quadrilateral maps in each view. The displacement is controlled by 1) a transversal direction to the vertical trapezoid given a perspective horizontal line passing through $\mathbf{v}_x$ for vertical trapezoidal map; and 2)a transversal direction to the horizontal trapezoid given a perspective depth line passing through $\mathbf{v}_z$ for horizontal trapezoidal map.

The exterior differential of a k-degree differential form gives a $(k+1)$-degree differential form. In particular, if we would have the simplest representation for the left vertical trapezoidal map given by the 2nd degree differential form $\omega_{vert} = a_{vert}(x, y, z)dz \wedge dx$, then its differential $d\omega_{vert} = \frac{\partial a_{vert}}{\partial y} dx \wedge dy \wedge dz$ would give a cuboid map relative to the lacking orthogonal direction. Similarly, if we would have the simplest representation for the lower horizontal trapezoidal map given by the 2nd degree differential form $\alpha_{hor} = b_{hor}(x, y, z)dy \wedge dz$, then its differential $d\alpha_{hor} = \frac{\partial b_{hor}}{\partial x} dx \wedge dy \wedge dz$ would give a cuboid map relative to the lacking orthogonal direction. As both cuboid maps must be the same, one has a structural constraint $d\omega_{vert} = d\alpha_{hor}$. Hence, it suffices to consider only a trapezoidal map. By visibility reasons it is more practical to restrict to a left vertical trapezoidal map $\mathcal{T}$, e.g., generated by the intersection of pencils of projective lines through $\mathbf{v}_z$ and $\mathbf{v}_y$.

4.3 Lie Contraction/Expansion Operators

Even when prior information about the scene is available, optical flow is not easy to compute due to noise, partial occlusions, etc. Quadrilateral are easier than points for tracking, and they provide a support for differential forms with compact support. Each mobile vector, bivector or three-vector (with support containing a segment, a quadrilateral or a cuboid) is the dual of a 1, 2 or 3-degree differential form in a point. Homologue elements are linked along the motion by the directional or Lie derivative.

The Lie (directional) derivative of a k-th degree differential form ω is the k-th degree differential form defined by $\mathcal{L}_\xi \omega(\underline{x}) = \frac{d}{d\lambda}(f^*_\lambda \omega)(\underline{x}) \mid_{\lambda=0}$; let us remark that $\frac{d}{d\lambda}$ is a linear operator.

The motion field θ corresponding to the apparent motion of homologue elements along the sequence of views is topologically modeled by a local diffeomorphism $f_\lambda(\xi)$ arising from the integration of the vector field ξ. The linearization of the diffeomorphism $f_\lambda(\xi)$ (differential evaluated at origin) gives an element of the general linear group $GL(3;\mathbb{R})$. Its semi-direct product with the translation gives the affine linear map between homologue elements along the sequence of temporal views. The estimation of θ is performed from the comparison between homologue trapezoids or, more generally, quadrilaterals in terms of Kalman filters (see below).

The flow conservation' equation $f^*_\lambda(d\omega) = d(f^*_\lambda \omega)$ implies the conservation of homologue $2d$ rigid elements (trapezoidal or quadrilateral maps) along the temporal sequence of views. The lifting of the flow conservation to the $3d$ ambient space implies the preservation of homologue $3d$ rigid elements along the motion. The homologue rigid elements are visually perceived as cuboids corresponding to different views of the same rectangular parallelepiped. Hence, they are related between them through affine transformations which must be estimated on line.

The exterior differentiation commutes with the Lie derivative which is a linear operator $\frac{d}{d\lambda}$. Thus, $d(\mathcal{L}_X \omega) = \mathcal{L}_X(d\omega)$ for every $\omega \in \Omega^k_M$, which can be expressed as $d\mathcal{L}_X = \mathcal{L}_X d$ in a more synthetic way.

The piecewise-linear expansion of the $2d$ quadrilateral map $\mathcal{Q}$ (given in particular by a the trapezoidal map $\mathcal{T}$) to a cuboid map $\mathcal{C}$ is locally given by the support of the exterior product $\omega \wedge \eta$ where η is a differential form supported on lines through the lacking vanishing point $\mathbf{v}_x$.

The apparent displacement of rigid elements identified on the quadrilateral map $\mathcal{Q}$ (respectively, cuboid map $\mathcal{C}$) is modeled by the Lie derivative of the 2-degree differential form ω (respectively, the 3-degree differential form $d\omega$) along the motion field ξ.

Given a differentiable manifold M, for any vector field $X \in \mathcal{X}(M)$ on M and for any $k+1$-degree differential form $\omega \in \Omega_M^{k+1}$, the *contraction of ω along the vector field X*, denoted as $i_X\omega$ is defined as the k-th differential form given by $i_X\omega(X_1,\ldots,X_k) := \omega(X,X_1,\ldots,X_k)$ for any vector fields $X_1,\ldots,X_k \in \mathcal{X}(M)$

The piecewise contraction of the cuboid map $\mathcal{C}$ to the three quadrilateral maps $\mathcal{Q}_{ij}$, for $1 \leq i < j \leq 3$ is locally given by the Lie contraction of Ω along the Lie vector field ξ of the egomotion.

The simplest *local representation* of mobile maps of cuboids (respectively, quadrilaterals) is given by orthographic three-dimensional reconstructions, where basic pieces are rectangular parallelepipeds, which are distorted by perspective effects. By means a projective transformation, it is possible to send the three vanishing points to the infinity, to construct a parallepiped model for the scene with euclidian information (for scaled orthographic resulting representation), solve the contraction/expansion, and to perform the inverse transformations. However, this would have a high computational cost, and it would be some difficult to obtain a real-time implementation. Thus, it is important to develop a forward estimation of egomotion directly supported on maps of homologue quadrilaterals. To fix ideas, let us restrict to the trapezoidal map $\mathcal{T}$ corresponding to $\mathbf{v}_y$ and $\mathbf{v}_z$

5 MOTION ANALYSIS

The general problem is which kind of information about the motion vector field Θ can be extracted from the egomotion vector field ξ in an indoor scene. Homologue quadrilaterals allow to evaluate magnitude and direction of motion field. Usual differential models for ego-motion estimation are based on optical flow. Some troubles are related with aperture problem, noise for geometric features, indeterminacy or ambiguity about homologue elements, between others. In our case, the differential formalism allows to reduce the estimation/tracking of a trapezoid to the estimation/tracking of two segments supported on perspective lines through $\mathbf{v}_z$. An important issue concerns to the estimation of $1D$ (perspective lines) or $2D$ data (trapezoids, quadrilaterals), and the adaptability to changing conditions. Thus, we develop an approach able of supporting estimation and tracking of positive dimensional elements.

After modeling, main problems concern to 1) the prediction from the current localization, 2) the identification of homologue quadrilaterals along sampled views of a video sequence, 3) model's validation and 4) prediction updating.

5.1 Dynamic Modeling

A discrete dynamical system can be represented by a *state equation* $s_i = \psi_{i,i-1}s_{i-1}w_i$ where s_i is the *state vector* at instant i, $\psi_{i,i-1}$ is a linear matrix, and w_i is the random noise representing a perturbation on the system. To simplify, one supposes a white noise, i.e. $E[w_i] = 0$, $E[w_i w_i^T] = Q_i$ where the covariance matrix Q_i is usually determined in a heuristic way. If we suppose that data are not correlated between them, then a diagonal matrix can be chosen for Q_i. To avoid heuristics, a choice based on triangular matrices allows to maintain dependence relations between nested elements.

5.2 Dynamical Estimation of the Trapezoidal Map

In this subsection, wand adaptation of (Faugeras, 1993) is developed for estimating and measuring the trapezoidal map as a discrete dynamical system. We have followed an *incremental estimation* approach:

- Static states for each trapezoid of the trapezoidal map $\mathcal{T}$ are represented by a $2d$ vector $\mathbf{t}_i$ corresponding to a coordinate of a corner, and the width of trapezoid.
- Kinematic states $\mathbf{s}_i$ are given by a $4d$ vector $(\mathbf{t}_i, \dot{\mathbf{t}}_i)$ where $\dot{\mathbf{t}}_i$ is the temporal variation of $\mathbf{t}_i$.
- The plant's or *state's equation* is $\mathbf{s}_i = \psi_{i|i-1}\mathbf{s}_{i-1} + \underline{w}_i$ where $\psi_{i|i-1}s_{i-1}$ is a triangular matrix responsible for the kinematic coupling and $\underline{w}_i$ is a white noise with $E[\underline{w}_i.\underline{w}_i^T] = Q_i$. To begin with, $\psi_{1|0} = \begin{pmatrix} I_2 & I_2 \\ 0 & I_2 \end{pmatrix}$ where I_2 is the 2×2 identity matrix.
- The *measurement equation* is given by $\mathbf{m}_i = H_i\mathbf{s}_i + \underline{\nu}_i$ with noise $\underline{\nu}_i$. It gives as output the 4×4 linear transformation H_i for each trapezoid.
- *Covariance matrix:* Due to the lack of information about estimated values, we take an initial covariance matrix P_i given by a 4×4 diagonal matrix with high positive eigenvalues.

- The *prediction of the covariance matrix* before comparing the 0-th (before motion) and the first image is equal to $P_{1|0} = \psi_0 P_1 \psi_0^T + Q_0$ where ψ_0 has been estimated before, and $Q_0 = E[\underline{w}.\underline{w}^T]$ for the i-th trapezoid.

5.3 Kalman Filtering

A Kalman filter (Faugeras, 1993) is a recurrent linear technique for solving a dynamic problem by the least squares method, allowing to integrate noised measures. Measures can be corrupted by white noise, and must be corrected without having all the measures. In our case, we have adapted Kalman filtering to maps of quadrilaterals, (Marion, 2002). The estimation based on KF is performed without waiting to have all the measures.

Typical steps of standard KF are given by initialization, measurement, validation and updating of prediction. The initialization has been already described below.

5.3.1 Mahalanobis Distance for Updating Measurements

Under uncertainty conditions, the covariance matrix Λ_i provides information about the dispersion of data, and allows to compute the *Mahalanobis distance* $d_{iM}(\underline{x}, \underline{\mu}) := (\underline{x} - \underline{\mu})\Lambda_i(\underline{x} - \underline{\mu})^T$ of a vector $\underline{x}$ with respect to the current mean value $\underline{\mu}$. In our case, the comparison between 0-th and first images gives an initial measurement matrix H. From H, the covariance matrix $\Lambda_0 = HR_0H^T + R_1$.

Along the iteration,$d_{iM} = (\mathbf{m}_{i+1} - H\mathbf{s}_i)\Lambda_i^{-1}(\mathbf{m}_{i+1} - H\mathbf{s}_i)^T$ where $\mathbf{m}_{i+1}$ is the measure at time $i+1$ with covariance matrix R_i, and $\Lambda_i = HP_iH^T + R_{i+1}$.

5.3.2 Validation

If the computed Mahalanobis distance for the segment representing the trapezoid, is lesser than a selected threshold, then the estimation is accepted. Otherwise, another trapezoid must be selected, and the process re-starts.

5.3.3 Updating of the Kalman Filter

The information is updated in the standard form as follows: 1) Gain matrix: $K_0 = R_0H^T(HR_0H^T + R_1)^{-1}$; 2) Updating of the state's estimation: $\hat{s}_1 0 s_0 + K_0(\mathbf{m}_1 - H\mathbf{s}_0)$; 3) Updating of P_0 in $\hat{P}_0 = (R_0 + H^T R_1^{-1} H)^{-1}$

The process is iterated for all trapezoids contained in the central part of the image with homologues easily identifiable in the precedent and consecutive view of the sequence.

5.3.4 Extended Kalman Filter EKF

Most processes in real world are not linear. Hence, we must adapt or extend KF (EKF) to the non-linear case: In each step a linearization of the precedent estimated state is performed. If the initialization is not enough good or if the noise is high, then the convergence is not guaranteed. In this case, it is necessary to modify initial constraints or to perform a new processing of the acquired information.

5.4 Estimation and Tracking of Quadrilaterals

The pipeline for estimating and tracking quadrilaterals is as follows

1. Grouping of mini-segments in long segments
2. Computation of intersections of lines supported by long segments
3. Identification of putative perspective lines
4. Estimation of vanishing points
5. Retracing of "true" perspective lines
6. Grouping of perspective lines in three pencils $\mathcal{L}_i$ through vanishing points $\mathbf{V}_i$
7. Maps of Quadrilaterals $\mathcal{Q}_{ij}$ linked to each pair $\mathcal{L}_i$, $\mathcal{L}_j$ of perspective lines.
8. Identification of allowed simple events involving changes in quadrilaterals (grouping or decompositions).
9. Representation of transformations involving indecomposable quadrilaterals by triangular matrices.
10. Solving systems by successive reduction in triangular matrices.

6 CONCLUSION

A geometric framework based on mobile perspective models is superimposed to structured scenes. An online updating of changing perspective models is performed from quadrilateral maps by sampling two images each second. Expansions of maps of quadrilaterals are constructed as maps of cuboids by using differential constraints. A new representation of the egomotion is provided in terms of Lie derivative of the quadrilateral maps and the Lie contraction of the cuboid map along the egomotion field ξ. The estimation of the egomotion is directly computed on the quadrilateral map by using an adaptation of Kalman filtering. Some challenges for the next future are related with a) the implementation of a *real-time* module for generating and tracking mobile perspective

models will be developed, b) improve metric localization from uncalibrated video sequences, c) identification of mobile human beings in the scene compatible with the motion mounted on the platform, d) evaluation of behaviors to improve the interaction.

ACKNOWLEDGEMENTS

The authors acknowledge to the MAPA Project and specially thanks to J.J. Fernandez-Martin and J.I.SanJose Alonso for their partial finantial support.

REFERENCES

A. Pruski, M. E. and Morre, Y. (2002). Vahm: A user adapted intelligent wheelchair. In *IEEE Intl Conf on Control Application (Glasgow, Scotland, UK)*. IEEE.

Cortés, U., Annicchiarico, R., Vázquez-Salceda, J., Urdiales, C., namero, L. C., López-Sánchez, M., Sànchez-Marrè, M., and Caltagirone, C. (2003). e-tools: The use of assistive technologies to enhance disabled and senior citizens' autonomy. In *EU-LAT Workshop on e-Health*, pages 117–133.

Faugeras, O. (1993). *Three-dimensional Computer Vision. A Geometric Viewpoint*. The MIT Press, Cambridge, Mass.

J. Minguez, L. M. and Montano, L. (2004). An architecture for snsor-based navigation in realistic dynamic and troublesome scenarios. In *Proc. IROS)*. IEEE.

Jelinek, D. and Taylor, C. (2001). Reconstruction of linearly parametrized models from single images with camera of unknown focal length. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(7):767–773.

Levine, S. P., Bell, D. A., Jaros, L. A., Simpson, R. C., and Koren, Y. (1999). The navchair assistive wheelchair navigation system. *IEEE Trans on Rehabilitation Engineering*, 7 (4):443–451.

M. Gonzalo, J. Finat, M. A. and Aguilar, S. (2002). Dynamical trapezoidal maps for coarse perspective model in indoor scenes. In *ISPRS'02, 1st International Conference on Enterprise Information Systems*. ISPRS.

M. Rous, H. L. and Kraiss, K. F. (2001). Vision-based landmark extraction using indoor scene analysis for mobile robot navigation. In *IEEE/RSJ International Conference*. IEEE.

Marion, G. (2002). *Estimacin y Prediccin del Movimiento Propio en Escenas de Interior mediante Filtros Kalman*. Master's thesis, Valladolid.

Mazo, M. and et al (2002). Experiences in assisted mobility: The siamo project. In *Proc. IEEE Intl Conf on Control Applications (Anchorage, Alaska)*, pages 766–771. IEEE.

P. Trahanias, M. Lourakis, A. A. and Orphanoudakis, S. C. (1997). Navigational support for robotic wheelchair platforms. In *Proc. ICRA-IEEE*. IEEE.

R. Simpson, E. LoPresti, S. H. and Nourbakhsh, I. (2004). The smart wheelchair component system. *J. of Rehabilitation Research and Development*, 41:429–442.

Toedemé, M. Nuttin, N. T. and Gool, K. V. (2004). Vision based intelligent wheelchair control: the role of vision and inertial sensing in topological navigation. *J. of Robotic Systems*, 21(2):85–94.

Yanco, H. A. (1998). Integrating robotic research: a survey of robotic wheelchair development. In *AAAI Spring Symp. on Integrating Robotic Research (Stanford, CA)*. IEEE.

Yuki, O. (2000). *Corridor Navigation of a Mobile Robot using a camera and sensors-multiagent approach*. Thesis, UCLA University, Los Angeles, CA.