University of Florence
Faculty of Engineering
Dipartimento di Sistemi e Informatica

Dottorato di ricerca in informatica, multimedialità e telecomunicazioni

Two Results in Computer Vision using Projective Geometry

by

Federico Pernici

Doctoral committee:

Prof. Alberto Del Bimbo

Prof. Giacomo Bucci

Year 2005

To my family and myself

Abstract

In the first part of this thesis the problem of metric reconstruction and texture acquisition from a single uncalibrated view of a surface of revolution (SOR) is addressed. Geometric constraints induced in the image by the symmetry properties of the SOR structure are exploited to perform self-calibration of a natural camera, 3D metric reconstruction, texture acquisition & mosaicing and automatic reconstruction. By exploiting the analogy with the geometry of single axis motion, i demonstrate that the imaged apparent contour and the visible segments of two imaged cross sections in a single SOR view provide enough information for these tasks.

In the second part of this thesis the problem of scheduling an active observer to visit as many[1] targets in an area of surveillance as possible is proposed. I show how it is possible to plan a sequence of decisions regarding what target to look at through such a foveal-sensing action. I propose a framework in which a pan/tilt/zoom camera executes saccades in order to visit, and acquire high resolution images (at least one) of, as many moving targets as possible before they leave the scene. The whole problem is casted into a dynamic discrete optimization framework. In particular, we will show that the problem can be solved by modeling the attentional gaze control as a kinetic traveling salesperson problem whose solution is approximated by iteratively solving time dependent orienteering problems

[1] An intelligent choice of the order of sensing the targets can significantly reduce the total dead-time wasted by the active camera and, consequently, its cycle time.

Acknowledgement

I would like to acknowledge the efforts and input of my supervisor, Professor Alberto Del Bimbo, and my colleagues of the VipLab, who were invaluable in shaping the course of this research.

Preface

This thesis address two problems in Computer Vision in which the proposed methods are related by the use of *Projective Geometry*. Projective Geometry is the important tool which forms the basis for the formulation of the two main frameworks here proposed. The geometry of single views is explored and monocular vision is shown to be sufficient to obtain a partial/complete three-dimensional reconstruction or measurements of a scene. To achieve this the properties of planar homographies, planar homologies and conics are extensively exploited.

The first part of this thesis regards about 3d reconstruction through camera (auto) calibration. A camera is a remarkably useful measuring device - it not only produces a realistic picture of a scene, but also provides information from which geometric properties of the scene can be measured and so reconstructed. Reconstructing scene geometry from images is one of the most active areas in computer vision. This task has proved to be very challenging. Still, with the almost exponential increase in computation resources that the last decade has seen, and with the application of principled mathematical methods, much progress has been made in the last decade especially in the geometry of multiple images. In this part the author propose a new method for the 3D reconstruction of surfaces of revolution from only a single uncalibrated view. Uncalibrated means that nothing is known about the geometry

of the camera taking the picture. In particular no knowledge of the camera internal parameters (such as focal length and aspect ratio) or its pose (position and orientation) is required at any time. In this minimal setting i'll show mathematically how to recover the global 3D structure (up to an overall scaling factor) of a SOR and how to calibrate the camera from the solely image content. The novel contributions of the thesis will be addressed more precisely in the conclusion. In broad terms, however, the methods described integrate constraints on calibration and reconstruction of surfaces of revolution into a single framework extending previous approaches dealing with piecewise planar scenes.

In the second part of the thesis a very novel problem is proposed in which an active observer (i.e. an active zooming camera) plans a sequence of decisions regarding what target to look at, through a foveal sensing action aimed to obtain high resolution images of such targets. The motivation is to collect images at some minimal ground resolution. This task might be defined to support gait recognition, or the acquisition of an appearance model that could be used to subsequently identify the person. I propose a optimization framework in which Projective Geometry of monocular vision is used for evaluating the cost in the optimization. This work is motivated by the goal of reproducing the ability of humans to recognize a person in a crowd of moving people for surveillance purposes.

The overall layout of subsequent chapters is as follows. Chapter 1 discuss some background and the single view geometry of surfaces of revolution. Chapter 2 propose a surface mosaicing algorithm for registering into a global images single views of SOR. Chapter 3 address the problem of giving an automatization to the whole process of SOR reconstruction from a single image. In chapter 4 (the second part of this thesis) present an approach

for planning an active zooming camera for the task of high resolution image sensing. Chapter 5 draws the thesis to a close, summing up the contributions made in the research and outlining areas of further investigation.

Contents

Part I

Single view geometry and Surfaces of Revolution

Chapter 1

Single view Reconstruction

Image analysis and computer vision can be effectively employed to recover the three-dimensional structure of imaged objects, together with their surface properties. In this chapter, we address the problem of metric reconstruction and texture acquisition from a single uncalibrated view of a surface of revolution (SOR). Geometric constraints induced in the image by the symmetry properties of the SOR structure are exploited to perform self-calibration of a natural camera, 3D metric reconstruction, and texture acquisition. By exploiting the analogy with the geometry of single axis motion, we demonstrate that the imaged apparent contour and the visible segments of two imaged cross sections in a single SOR view provide enough information for these tasks. Original contributions of this part of the thesis are: single view self-calibration and reconstruction based on planar rectification, previously developed for planar surfaces, has been extended to deal also with the SOR class of curved surfaces; self-calibration is obtained by estimating both camera focal length (one parameter) and principal point (two parameters) from three independent linear constraints for the SOR fixed entities; the invariant-based description of the SOR scaling function has been extended from affine

to perspective projection. The solution proposed exploits both the geometric and topological properties of the transformation that relates the apparent contour to the SOR scaling function. Therefore, with this method, a metric localization of the SOR occluded parts can be made, so as to cope with them correctly. For the reconstruction of textured SORs, texture acquisition is performed without requiring the estimation of external camera calibration parameters, but only using internal camera parameters obtained from self-calibration.

1.1 Introduction

In the last few years, the growing demand of realistic 3D object models for graphic rendering, creation of non-conventional digital libraries, and population of virtual environments has renewed the interest in the reconstruction of 3D objects and the acquisition of their surface properties from one or more camera images. In fact, solutions based on image-analysis can be effectively employed in all those cases in which the original object is not available and only its photographic reproduction can be used (for example it no longer exists), or its material does not work with structured light methods, or its size is too large for the other automatic acquisition methods. The knowledge of internal camera parameters is essential for the purpose of metric reconstruction, of the 3D object shape and structure and its texture, from image data. Self-calibration [1] is particularly important in that, although less accurate than off-line calibration [2], [3], it is the only possible solution when no direct measurements can be made in the scene, as for example in applications dealing with archive photographs and recorded video sequences. In this thesis we address the task of metric reconstruction and texture acquisition from

a single uncalibrated image of a SOR. We follow a method which exploits geometric constraints of the imaged object without requiring any knowledge of both internal and external camera calibration parameters. We derive the geometric constraints induced in the image by the symmetry properties of SOR structure. The key idea is that, since a SOR is a non trivial repeated structure generated by the rotation of a planar curve around the axis, in principle it can be recovered by properly extending and combining together single image planar scene reconstruction and single axis motion constraints. A preliminary version of the approach described has been presented in [4] and this is the first published paper dealing with uncalibrated single view SOR metric reconstruction. In the following sections, we present the algorithms and the principles of projective geometry exploited, for the three stages of: camera calibration, 3d shape and structure metric reconstruction and texture acquisition. Metric reconstruction of the 3D shape and structure of the SOR is reformulated as the problem of determining the shape of a meridian curve. The inputs to the algorithms are two elliptical imaged SOR cross sections and the silhouette of the object apparent contour. Working with image curves has two main advantages. On the one hand, curves are dominant features that can be easily and reliably extracted from images; on the other hand, using curves allows avoiding point correspondences and making reconstruction possible also for textureless, translucent and transparent objects. Constraints for camera calibration are obtained from the computation of the projective geometry of the SOR from a single view. Actually, this removes the 1D projective reconstruction ambiguity due to underconstrained calibration that has been discussed in the literature of turntable sequences (see e.g. [5]). It also represents an improvement w.r.t. the approach presented in [6], that requires the presence of two different SORs in the same view. The contri-

bution of our reconstruction approach to the recent literature is three-fold. First, it extends to the SORs the planar rectification framework originally introduced in [7] for planar surfaces. Second, it develops the invariant-based description of SORs discussed in [8] from affine to perspective projection, by proving that also in this case, the reconstructed scaling function of the imaged SOR forms a "canonical frame" in which invariants can be computed. Finally, in the case of self occlusions, since the approach operates locally on the apparent contour, occluded parts can be precisely identified and eventually recovered from multiple views. In this regard, it improves the SOR reconstruction approach described in [9].

1.2 Related Work

3D shape and structure reconstruction approaches that have been reported in the literature (see [10] for a recent survey) include classic triangulation [11], [12], visual hulls [13]—typically enhanced by voxel coloring [14], dense stereo [15] and level sets methods [16]. However, more effective reconstruction can be obtained by exploiting prior knowledge about the scene, encoded in the form of constraints on either scene geometry or motion. Most of the recent research contributions that used prior knowledge employ geometric scene constraints. The presence of a "repeated structure" [17] is a classical example of geometric constraint frequently used. This happens because the image of a repeated structure is equivalent to multiple views of a single structure. In real applications this can regard planes, lines, etc. occurring in particular (e.g., parallel, orthogonal) spatial arrangements. In a repeated structure, the epipolar geometry induced in the image by multiple instances of the same object can be expressed through projective homologies, which

require less parameters and therefore are more robust to estimate [18]. A further advantage of geometrically constrained reconstruction is that fewer (and, in special cases, just one) images are required. In [19], an interactive model-based stereo approach was proposed by Taylor et al., where the scene is represented as a constrained hierarchical model of parametric polyhedral primitives—such as boxes, prisms—called blocks. The user can constrain the sizes and positions of any of the blocks in order to simplify the reconstruction problem. All these constraints are set in the 3D space, thus requiring a complex non-linear optimization to estimate camera positions and model parameters. Liebowitz et al. perform calibration from scene constraints in order to reconstruct textured piecewise planar architectural scenes [7]. Moreover, they rectify the projective distortion in the imaged planar portions so as to represent the original texture as a rectangular image. Single view piecewise planar reconstruction is also addressed by Sturm and Maybank [20], [21]. The main difference w.r.t. [7] is that reconstruction does not include the rectification step. This saves computation time but requires a larger memory space to store textures. In [22], Cross and Zisserman use quadrics for 3D modeling and image registration from two or more perspective views. They also show that surfaces can be approximated piecewise by quadrics that are obtained from a piecewise conic approximation of their outlines in the image plane.

Useful constraints for calibration and/or reconstruction can also be derived if the scene undergoes planar motion [23]. Recent works exploit single axis motion to reconstruct objects of any shape that rotate on a turntable [24], [25], [5] [26]. Apart from algorithmic differences in the reconstruction phase, motion fixed entities like the imaged axis of rotation and the vanishing line of the plane of rotation are first estimated from the image sequence, and

then used to calibrate the camera. However, the turntable approaches above do not succeed to perform a complete camera calibration. As a consequence of this, reconstruction is affected by a 1D projective ambiguity along the rotation axis.

On the other hand, for the creation of realistic 3D object models, texture acquisition must be combined with 3D shape and structure acquisition. Furthermore image geometry/registration is only the first step for texture reconstruction infact a correct blending is necessary to remove lighting from the albedo and remove the redundant information by compacting the whole pictorial surface. Standard approaches of texture acquisition generally require the knowledge of camera position and orientation (i.e., the external calibration parameters) in order to backproject correctly image data onto surface shape. There are basically two methods for estimating external calibration from image data and a known 3D structure [27], [28], [29], [30]. The first method exploits the correspondence between selected points on the 3D object and their images. The second method works directly in the image plane, and minimizes the mismatch between the original object silhouette and the synthetic silhouette obtained by projecting the 3D object onto the image. However, in general, if there are internal calibration or surface reconstruction errors, texture acquisition introduces visual artifacts which do not reflect the real object appearance.

Surfaces of Revolution (SORs) represent a class of surfaces very common in man-made objects and thus of great relevance for a large number of applications. SORs are a subclass of Straight Homogeneous Generalized Cylinders (SHGCs). SHGCs have been extensively studied under different aspects: description, grouping, recognition, recovery, and qualitative surface recostruction. Most of the existing techniques follow [31], for an extensive

review, see [32]. Abdallah and Zisserman, in [8] have discussed the invariance properties of SORs scaling function under affine viewing conditions — finding the analogous invariants in the perspective view case is left to future work — for the purpose of object recognition from a single view. Reconstruction of a generic SHGC from a single view, either orthographic or perspective, is known to be an underconstrained problem, except for the case of SORs [33]. However, with the noticeable exception of the work by Wong [9], virtually no contributions exist addressing the problem of metric reconstruction of SORs from a single perspective view. In that paper, the 3D structure of a SOR is reconstructed from its silhouette from an uncalibrated image. The calibration is obtained with the method described in [34], [6]. However with this method only the focal length can be estimated from a single view with the assumptions of zero skew and principal point being at the image center.

The reconstruction is affected by a 1-parameter ambiguity: although this can be fixed by localizing an imaged cross section of the surface, however, a major problem is that in this approach the silhouette is related directly to its generating contour on the surface, which is an incorrect assumption. In fact, this makes it impossible to capture the correct object geometry in the presence of self-occlusions, which is a very common situation. Very recently two new papers were published: [35] [36] in which two imaged cross-section as described in [4] were used to perform respectively a projective reconstruction and a multi-view metric reconstruction. However in [36] the two imaged cross-section were not fully exploited since the circular point constraint on the IAC is not used.

Texture acquisition of straight uniform generalized cylinders (SUGCs), which are a special subclass of SORs [37], was addressed by Puech et al.. In this approach, texture is obtained as a mosaic image gathering visual

information from several images. Since texture is not metrically sampled, the quality of the global visual appearance of the object is somewhat affected.

Another important aspect is that several paper exists for the automatic segmentation (mainly devoted for recognition purposes) of SHCG and in particular for SOR [38] [39] [40] [41] [32]. This gives a relevant interest to our work since the possibilities to perform the reconstruction from images complete automatically (albeit grouping for recognition is rather different than for metric reconstruction since high accuracy in the estimated entity is needed). Texture acquisition is also obtained by exploiting the special properties of SOR structure. In our approach, texture is sampled in the image plane, and represented as a flat rectangular image through a cartographic parametrization. The surface texture is acquired by a special 2D transformation applied to the imaged SOR region. This act as a simplification of the geometry to image registration phase. This is different to the general method in which the texture images acquired by a color camera need to be registered with the reconstructed 3D points by means of camera external calibration which establish the projective mapping of the texture image onto the 3d points [42]. In our method external camera pose wasn't not explicitly computed: the image to geometry registration its computed entire in the image space.

Our methods gives also a benefit in the acquisition of the whole texture which involves the combination (blending) of all the texture maps acquired for an SOR object into a single non-redundant map over the entire object with adjusted color. The texture acquired with our methods can be combined with a translation motion model by mean of image registration. Our methods in general reduces the sensitivity to artifacts since a lower number of parameters are involved in the estimation.

Calibration information is exploited in the resampling phase. Fully metric textures are obtained for developable surfaces. In that the method is not limited to the class of SUGCs, it also extends the solution presented in [37]. The content of this chapter is organized as follows. Section 1.3 provides background material on basic geometry and states the analogy between single axis motion and surfaces of revolution. Section 1.4 describes the various parts of the approach, and specifically computation of the fixed entities, camera calibration, reconstruction of 3D structure and texture acquisition. In section 1.5 experimental results on both synthetic and real data are presented and discussed. Finally, in section 1.6 conclusions are drawn and future work is outlined. Mathematical proofs are reported in the Appendices.

1.3 Background

In this section we review the basic terminology and geometric properties of SORs under perspective projection. We also discuss an important analogy between properties as derived from a single SOR image and those of a sequence of images obtained from single axis motion: this analogy will be exploited in the calibration, reconstruction and texture acquisition algorithms, discussed in section 1.4.

1.3.1 Basic terminology

Mathematically, a *surface of revolution* can be thought of as obtained by revolving a planar curve $\rho(z)$, referred to as *scaling function*, around a straight axis z. Therefore, SORs can be parametrized by $\mathbf{P}(\theta, z) = (\rho(z)\cos(\theta), \rho(z)\sin(\theta), z)$, with $\theta \in [0, 2\pi]$, $z \in [0, 1]$. In the 3D space, all *parallels* (i.e., cross sections with planes $z =$ constant) are cir-

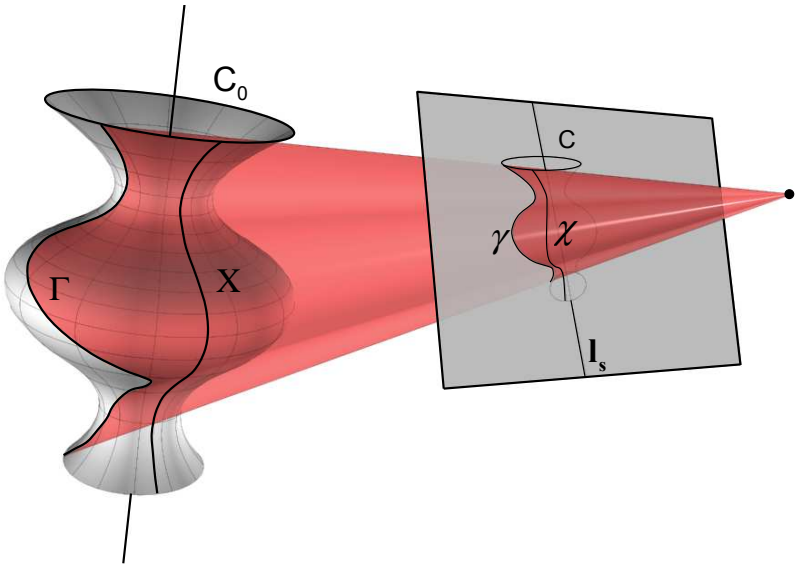


Figure 1.1: Imaged SOR geometry. Γ and γ are respectively part of the contour generator and of the apparent contour. The translucent surface is the visual hull for the apparent contour. X and χ are respectively a meridian and its projection. The ellipse $\mathtt{C}$ is the edge corresponding to the parallel $\mathtt{C}_o$.

cles. *Meridians* (i.e., the curves obtained by cutting the SOR with planes θ = constant) all have the same shape, coinciding with the SOR scaling function. Parallels and meridians are locally mutually orthogonal in space, but not in a 2D view. Two kinds of curves can arise in the projection of a SOR: *limbs* and *edges* [43]. A limb, also referred to as *apparent contour*, is the image of the points at which the surface is smooth and projection rays are tangent to the surface. The corresponding 3D curve is referred to as *contour generator*. An edge is the image of the points at which the surface is not smooth and has discontinuities in the surface normal. Fig. 1.1 depicts a SOR and its projection. Under general viewing conditions, the contour generator is not a planar curve, and is therefore different from a meridian [44]. Depending on this, the apparent contour also differs from the imaged meridian. Parallels always project onto the image as ellipses. Edges are el-

liptical segments which are the projection of partially or completely visible surface parallels.

1.3.2 Basic imaged SOR properties

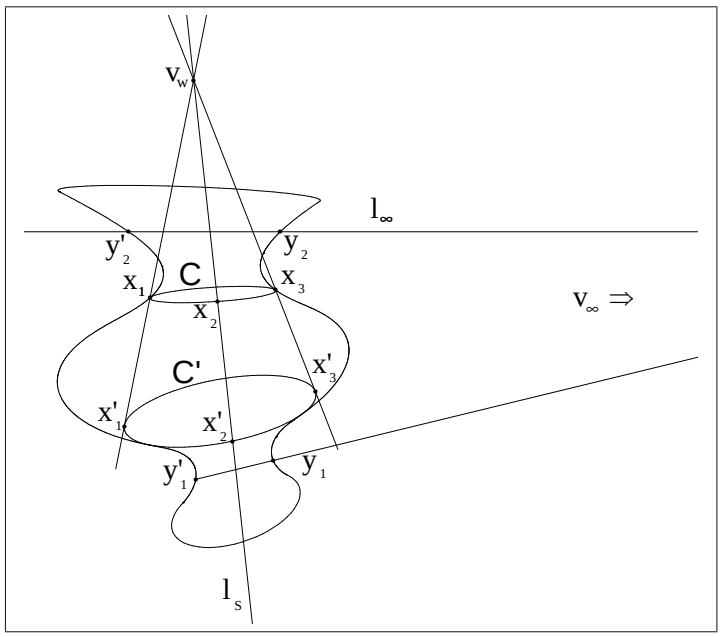


Figure 1.2: Basic projective properties for an imaged SOR. Property 1.3.1: Points $\mathbf{x}_i$ and $\mathbf{x}'_i$ correspond under $\mathtt{W}$; all lines $\mathbf{x}'_i \times \mathbf{x}_i$ meet at $\mathbf{v}_\mathtt{W} \in \mathbf{l}_s$. Property 1.3.2: Points $\mathbf{y}_i$ and $\mathbf{y}'_i$ correspond under $\mathtt{H}$; all lines $\mathbf{y}'_i \times \mathbf{y}_i$ meet at $\mathbf{v}_\infty \in \mathbf{l}_\infty$ (not shown in the figure).

Most of the properties of imaged SORs can be expressed in terms of projective transformations called *homologies*. These are special planar transformations that have a line of fixed points (the homology axis) and a fixed point (the vertex) that does not belong to the axis [45]. In homogeneous coordinates, a planar homology is represented by a 3×3 matrix $\mathtt{W}$ transforming points as $\mathbf{x}' = \mathtt{W}\mathbf{x}$. This matrix has two equal and one distinct real eigenvalues, with eigenspaces respectively of dimension two and one. It can

be parametrized as

$$\mathtt{W} = \mathtt{I} + (\mu - 1)\frac{\mathbf{v}\,\mathbf{l}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\,\mathbf{l}} \ , \tag{1.1}$$

where $\mathtt{I}$ is the 3×3 identity matrix, $\mathbf{l}$ is the axis, $\mathbf{v}$ is the vertex and μ is the ratio of the distinct eigenvalue to the repeated one. A planar homology has five degrees of freedom (dof); hence, it can be obtained from three point correspondences. In the special case $\mu = -1$, the dofs are reduced to four, and the corresponding homology $\mathtt{H}$ is said to be *harmonic*.

An imaged SOR satisfies the following two fundamental properties, whose geometric meaning is illustrated in Fig. 1.2.

Property 1.3.1. Any two imaged SOR cross sections are related to each other by a planar homology $\mathtt{W}$. The axis of this homology is the vanishing line $\mathbf{l}_\infty$ of the planes orthogonal to the SOR axis. The image of the revolution axis $\mathbf{l}_s$ contains the vertex $\mathbf{v}_\mathtt{W}$ of the homology [8][32].

Property 1.3.2. The apparent contour of an imaged SOR is transformed onto itself by an harmonic homology $\mathtt{H}$, whose axis $\mathbf{l}_s$ is the imaged axis of symmetry of the SOR. The vertex $\mathbf{v}_\infty$ of the homology lies on the aforementioned vanishing line $\mathbf{l}_\infty$ [46].

Denoting with $\mathtt{C}$ and $\mathtt{C}'$ the 3×3 symmetric conic coefficient matrices associated with two generic cross sections that correspond pointwise under the planar homology $\mathtt{W}$, it holds $\mathtt{C}' = \mathtt{W}^{-\mathrm{T}}\mathtt{C}\mathtt{W}^{-1}$. The harmonic homology generalizes the usual concept of bilateral symmetry under perspective projection. In fact, the imaged axis of symmetry splits the imaged SOR into two parts, which correspond pointwise through $\mathtt{H}$. This is true, in particular, for imaged cross sections, that are fixed as a set under the harmonic homology: $\mathtt{C} = \mathtt{H}^{-\mathrm{T}}\mathtt{C}\mathtt{H}^{-1}$ (or $\mathtt{C} = \mathtt{H}^{\mathrm{T}}\mathtt{C}\mathtt{H}$, being $\mathtt{H}^{-1} = \mathtt{H}$). The two elliptical cross-section $\mathtt{C}$ and $\mathtt{C}'$ of Fig. 1.2 are therefore related by a planar homology $\mathtt{W}$ with axis

$\mathbf{l}_\infty$ and vertex $\mathbf{v_w}$. The vertex $\mathbf{v_w}$ is always on the imaged axis of revolution $\mathbf{l_s}$. Cross-section points $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ correspond to $\mathbf{x}'_1$, $\mathbf{x}'_2$, $\mathbf{x}'_3$ under $\mathtt{W}$. Cross-section points $\mathbf{x}_1$, $\mathbf{x}'_1$, $\mathbf{x}_2$, $\mathbf{x}'_2$ also correspond to $\mathbf{x}_3$, $\mathbf{x}'_3$, $\mathbf{x}_2$, $\mathbf{x}'_2$ under $\mathtt{H}$. The points on the apparent contour $\mathbf{y}'_1$, $\mathbf{y}'_2$ also correspond to $\mathbf{y}_1$, $\mathbf{y}_2$ under $\mathtt{H}$. Lines through points $\mathbf{y}'_1$, $\mathbf{y}_1$ and $\mathbf{y}'_2$, $\mathbf{y}_2$ meet at $\mathbf{v}_\infty$.

1.3.3 The analogy between SOR geometry and single axis motion

Given a static camera, and a generic object rotating on a turntable, single axis motion (SAM) provides a sequence of different images of the object. This sequence can be imagined as produced by a camera that undergoes a *virtual* motion of pure rotation about the turntable axis while viewing a fixed object. Single axis motion can be described in terms of its fixed entities—i.e., those geometric objects in space or in the image that remain invariant throughout the sequence [23]. In particular, the imaged fixed entities can be used to express orthogonality relations of geometric objects in the scene by means of the image of the absolute conic ω—an imaginary point conic directly related to the camera matrix $\mathtt{K}$ as $\omega = \mathtt{K}^{-\mathrm{T}}\mathtt{K}^{-1}$ [11].

Important fixed entities for the SAM are the imaged circular points $\mathbf{i}_\pi$ and $\mathbf{j}_\pi$ of the pencil of planes π orthogonal to the axis of rotation, and the horizon $\mathbf{l}_\pi = \mathbf{i}_\pi \times \mathbf{j}_\pi$ of this pencil. The imaged circular points form a pair of complex conjugate points which lie on ω:

$$\mathbf{i}_\pi^{\mathrm{T}}\,\omega\,\mathbf{i}_\pi = 0, \;\; \mathbf{j}_\pi^{\mathrm{T}}\,\omega\,\mathbf{j}_\pi = 0 \;\; . \tag{1.2}$$

In practice, as $\mathbf{i}_\pi$ and $\mathbf{j}_\pi$ contain the same information, the two equations above can be written in terms of the real and imaginary parts of either points. Other relevant fixed entities are the imaged axis of rotation $\mathbf{l}_a$, and

the vanishing point $\mathbf{v}_n$ of the normal direction to the plane passing through $\mathbf{l}_a$ and the camera center. These are in pole-polar relationship with respect to ω:

$$\mathbf{l}_a = \omega \mathbf{v}_n \quad . \tag{1.3}$$

Eqs. 1.2 and 1.3 were used separately in the context of approaches to 3D reconstruction from turntable sequences. In particular, Eq. 1.2 was used in [24] and in [5] to recover metric properties for the pencil of parallel planes π given an uncalibrated turntable sequence. In both cases, reconstruction was obtained up to a 1D projective ambiguity, since the two linear constraints on ω provided by Eq. 1.2 were not enough to calibrate the camera. On the other hand, Eq. 1.3 was used in [26] to characterize the epipolar geometry of SAM in terms of $\mathbf{l}_a$ and $\mathbf{v}_n$ given a calibrated turntable sequence. Clearly, in this case, the a priori knowledge of intrinsic camera parameters allows to obtain an unambiguous reconstruction. In the case of a SOR object, assuming that its axis of revolution is coincident with the turntable axis, the apparent contour remains unchanged in every frame of the sequence. Therefore, for a SOR object, the fixed entities of the motion can be computed from any single frame of the sequence. According to this consideration, *a SOR image and a single axis motion sequence share the same projective geometry*: the fixed entities of SOR geometry correspond to the fixed entities of single axis motion. In particular: $\mathbf{l}_a$ corresponds to $\mathbf{l}_s$; $\mathbf{v}_n$ corresponds to $\mathbf{v}_\infty$; $(\mathbf{i}_\pi, \mathbf{j}_\pi)$ correspond to $(\mathbf{i}, \mathbf{j})$; $\mathbf{l}_\pi$ corresponds to $\mathbf{l}_\infty = \mathbf{i} \times \mathbf{j}$ where $\mathbf{i}$ and $\mathbf{j}$ denote the imaged circular points of the SOR cross sections.

The analogy between SOR and SAM imaged geometry was implicitly exploited in [6] to calibrate the camera from two SOR views under the assumption of zero camera skew. In that paper, the pole-polar relationship of $\mathbf{l}_s$ and $\mathbf{v}_\infty$ w.r.t. the image of the absolute conic was used to derive two

constraints on ω. Also in [26] a similar exploitation is used to locate the rotation axis and the vanishing point. In section 1.4.2 we will exploit the analogy one step forward, and show that it is possible to apply both Eqs. 1.2 and 1.3 to SORs for camera calibration and 3D reconstruction given a single SOR view.

1.4 The Approach

In this section we demonstrate that, given a single SOR view and assuming a zero skew known aspect ratio camera, the problems of camera calibration, metric 3D reconstruction and texture acquisition are solved if the apparent contour γ and two distinct imaged cross sections $\mathtt{C}_1$ and $\mathtt{C}_2$ are extracted from the original image. Preliminary to this, we demonstrate that the fixed entities $\mathbf{l}_s$, $\mathbf{v}_\infty$, $\mathbf{l}_\infty$, $\mathbf{i}$ and $\mathbf{j}$—that are required for all the later processing—can be derived from the two imaged cross sections.

1.4.1 Derivation of the fixed entities

The non linear system

$$\begin{cases} \mathbf{x}^{\mathrm{T}}\mathtt{C}_1\mathbf{x} = 0 \\ \mathbf{x}^{\mathrm{T}}\mathtt{C}_2\mathbf{x} = 0 \end{cases} \tag{1.4}$$

that algebraically expresses the intersection between $\mathtt{C}_1$ and $\mathtt{C}_2$ has four solutions $\mathbf{x}_k$, $k = 1 \ldots 4$—of which no three are collinear [45]—that can be computed as the roots of a quartic polynomial [47]. In the case that $\mathtt{C}_1$ and $\mathtt{C}_2$ are two imaged SOR cross sections, at least two solutions (say $\mathbf{x}_1$ and $\mathbf{x}_2$) are complex conjugate. In fact, since any imaged cross section intersects the vanishing line $\mathbf{l}_\infty$ at the imaged circular points $\mathbf{i}$ and $\mathbf{j}$, these complex conjugate points must belong to the solution set of Eq. 1.4. According to

this, the remaining two solutions (say $\mathbf{x}_3$ and $\mathbf{x}_4$) are either real or complex conjugate.

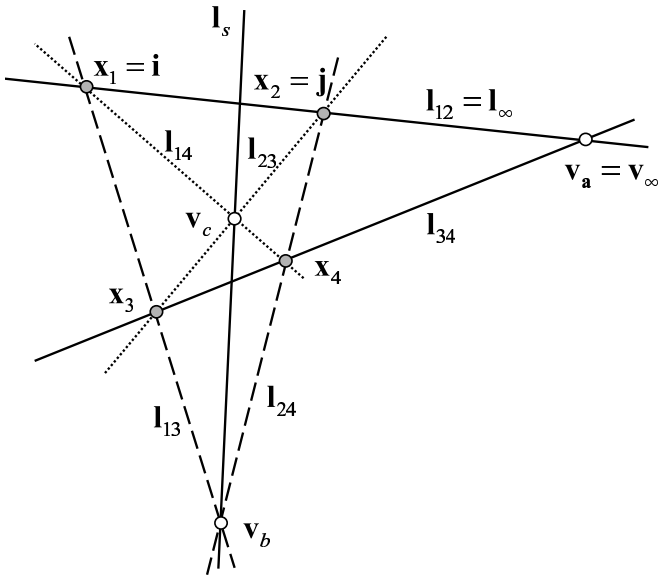
Fig. 1.3(a) shows the geometric construction for the derivation of the fixed entities $\mathbf{v}_\infty$ and $\mathbf{l}_s$. The four solutions $\mathbf{x}_k$'s form a so called "complete quadrangle" and are represented in the figure as filled circles. In the figure it is assumed, with no loss of generality, that $\mathbf{x}_1$ and $\mathbf{x}_2$ are the two imaged circular points $\mathbf{i}$ and $\mathbf{j}$. The $\mathbf{x}_k$'s may be joined in pairs in three ways through the six lines $\mathbf{l}_{ij} = \mathbf{x}_i \times \mathbf{x}_j$, $i = 1, \ldots 3$, $j = i+1, \ldots 4$. Each pair of lines has a point of intersection, and the three new points (hollow circles in the figure) form the vertices of the so called "diagonal triangle" associated with the complete quadrangle. The vertex of the harmonic homology $\mathbf{v}_\infty$ is the vertex of the diagonal triangle which lies on the line $\mathbf{l}_{12}$ connecting the two complex conjugate points $\mathbf{x}_1$ and $\mathbf{x}_2$. The imaged axis of symmetry $\mathbf{l}_s$ is the line connecting the remaining two vertices of the diagonal triangle. In particular, the vertex of the harmonic homology and the imaged axis of symmetry can be computed respectively as

$$\mathbf{v}_\infty = \mathbf{l}_{12} \times \mathbf{l}_{34} \tag{1.5}$$

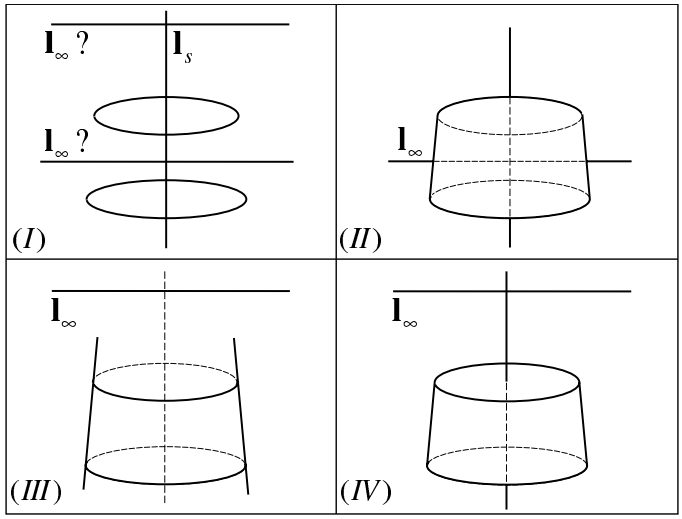
and

$$\mathbf{l}_s = (\mathbf{l}_{13} \times \mathbf{l}_{24}) \times (\mathbf{l}_{14} \times \mathbf{l}_{23}) \ . \tag{1.6}$$

The proof of this result is given in Appendix A.1. The computation of the vanishing line $\mathbf{l}_\infty$ is straighforward when $\mathbf{x}_3$ and $\mathbf{x}_4$ are real. In this case, $\mathbf{x}_1$ and $\mathbf{x}_2$ are the imaged circular points and, by consequence, $\mathbf{l}_\infty = \mathbf{l}_{12}$. On the other hand, when $\mathbf{x}_3$ and $\mathbf{x}_4$ are complex conjugate, an ambiguity arises in the computation of $\mathbf{l}_\infty$, since both $\mathbf{l}_{12}$ and $\mathbf{l}_{34}$ are physically plausible vanishing lines. In fact, a pair of imaged cross sections $\mathtt{C}_1$ and $\mathtt{C}_2$ with no real points of intersection are visually compatible with two distinct views of the



(a)



(b)

Figure 1.3: (a): Geometric properties of the four intersection points of $\mathtt{C}_1$ and $\mathtt{C}_2$ with the hypotesis $\mathbf{l}_\infty = \mathbf{l}_{12}$. (b): Ambiguity and visibility issues in determining the vanishing line. This symbolic view can be interpreted as two different view of two coaxial parallel 3d world circle represented with the two shown vanishing line. If these circles comes from a SOR three possible interpretation can be exist with the same appearence.

planar cross sections, where each view corresponds to a different vanishing line. Fig. 1.3(b)(I) shows an example of two imaged cross sections and the two possible solutions for the vanishing line; Fig. 1.3(b)(II) shows the correct solution for the vanishing line when the camera center is at any location in between the two planes of the cross sections; Figs. 1.3(b)(III) and (IV) show the correct solution for the vanishing line when the camera center is at any location above the two planes of the cross sections. The example shows that, unless the two imaged cross sections are one inside the other—which is indeed not relevant for the purpose of our research, since in this case no apparent contour could be extracted—, at least one of them is not completely visible. This suggests a simple heuristics to resolve the ambiguity based on visibility considerations. When both $\mathtt{C}_1$ and $\mathtt{C}_2$ are both not completely visible, the correct vanishing line $\mathbf{l}_\infty$ is the one whose intersection with $\mathbf{l}_s$ belongs to $h_1 \cap h_2$, where h_i is the half-plane generated by the major axis of $\mathtt{C}_i$ that contains the majority of the hidden points. In the case in which one of the two ellipses, say $\mathtt{C}_1$, is completely visible, then the correct $\mathbf{l}_\infty$ leaves both $\mathtt{C}_1$ and $\mathtt{C}_2$ on the same side. Once $\mathbf{l}_\infty$ is associated to the correct $\mathbf{l}_{ij} = \mathbf{x}_i \times \mathbf{x}_j$, the imaged circular points are simply obtained as the points $\mathbf{i} = \mathbf{x}_i$ and $\mathbf{j} = \mathbf{x}_j$. In other words the correct $\mathbf{i}$ and $\mathbf{j}$ are simply two out of the four intersection of the two conic. The above result demonstrates that the visible segments of two ellipses are in any case sufficient to extract unambiguously the vanishing line and the imaged circular points. This relaxes the conditions claimed by Jiang et al. in [5], where three ellipses are requested to compute the imaged circular points.

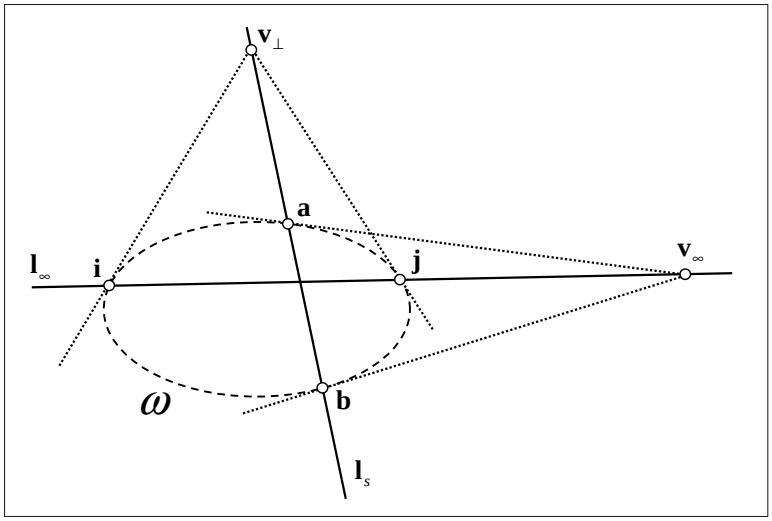


Figure 1.4: A symbolic representation of the geometry of camera calibration. The IAC is shown dashed to remind that it is a pure imaginary conic.

1.4.2 Camera calibration

In order to perform camera calibration from a single image of a SOR, we exploit the analogy between a single SOR image and single axis motion discussed in section 1.3.3. According to this, we can rewrite Eqs. 1.2 and 1.3 in terms of the SOR fixed entities $\mathbf{i}$, $\mathbf{j}$, $\mathbf{l}_s$ and $\mathbf{v}_\infty$. The resulting system

$$\begin{cases} \mathbf{i}^{\mathrm{T}}\,\omega\,\mathbf{i} = 0 \\ \mathbf{j}^{\mathrm{T}}\,\omega\,\mathbf{j} = 0 \\ \mathbf{l}_s = \omega\mathbf{v}_\infty \end{cases} \tag{1.7}$$

provides four linear constraints on the image of the absolute conic ω. However, it can be demonstrated (see AppendixA.2) that the system has only three independent linear constraints. Therefore, the available constraints are sufficient to calibrate a natural camera (zero skew and known aspect ratio: 3 dofs) from a single image. By rewriting the third of Eqs. 1.7 as $\mathbf{l}_s \times \omega\mathbf{v}_\infty = \mathbf{0}$, the system can be transformed into a homogeneous system and solved by singular value decomposition (see appendix A.4). Once ω is computed, the

camera matrix $\mathtt{K}$ can be obtained by the Cholesky factorization.

Different conditions can also be considered: (i) a single image with n SOR objects provides—except than in special configurations—$3n$ constraints that can be used to perform a full pinhole camera calibration (5 dofs); (ii) m distinct images of a SOR—obtained without varying the internal camera parameters—provide $3m$ linear constraints for full camera calibration.

Fig. 1.4 shows the geometrical relationships between the fixed entities and the image of the absolute conic, as mathematically expressed by the system of Eq. 1.7. The three points $\mathbf{v}_\infty$, $\mathbf{v}_s = \mathbf{l}_\infty \times \mathbf{l}_s$ and $\mathbf{v}_\perp \in \mathbf{l}_s$ are the vanishing points of three mutually orthogonal directions in the 3D space. In particular, $\mathbf{v}_\perp$ is the vanishing point of the directions parallel to the SOR symmetry axis; since this point cannot be measured from a single SOR view, its associated constraint $\mathbf{l}_\infty = \omega\mathbf{v}_\perp$ cannot be used for calibration purposes. If $\mathbf{v}_\infty$ is a finite point, there exists only one IAC such that $\mathbf{l}_\infty$ intersects ω at the fixed points $\mathbf{i}$ and $\mathbf{j}$, and the tangent lines to ω from $\mathbf{v}_\infty$ have the tangent points on $\mathbf{l}_s$. In the case in which the optical axis of the camera pierces the revolution axis of the SOR, the principal point is on $\mathbf{l}_s$ and as a consequence $\mathbf{v}_\infty$ becomes an ideal point. The effect of this is a 1-dof ambiguity in the position of the principal point, which can be anywhere on the imaged axis of symmetry. A practical solution to this problem is to choose as the principal point the point on $\mathbf{l}_s$ nearest to the image center [48].

When the principal point is close to $\mathbf{l}_s$, although not exactly on it, a near degenerate condition occurs. In this case, the accuracy of calibration strongly depends on the accuracy of the estimation of the fixed entities, and particularly of $\mathbf{v}_\infty$.

1.4.3 3D Metric Reconstruction

Given the IAC, it is possible to remove the projective distortion of any imaged plane for which the vanishing line is known—a technique known as planar rectification [11]. According to this, if the image χ of any SOR meridian, the corresponding vanishing line $\mathbf{m}_\infty$ and the imaged axis of symmetry $\mathbf{l}_s$ are available, it is possible to guarantee a solution for the problem of 3D metric SOR reconstruction.

As a first step, we compute χ and $\mathbf{m}_\infty$ from one imaged cross section $\mathtt{C}$ and the apparent contour γ under full perspective conditions.

The imaged meridian χ and $\mathbf{l}_s$—the latter obtained as shown in the previous section—will then be rectified in order to compute the SOR scaling function $\rho(z)$.

Computation of the imaged meridian

The following properties for the apparent contour and the imaged cross sections of a SOR extend the basic imaged SOR properties discussed in section 1.3.2, and provide the theoretical foundation for computing the imaged meridian χ.

Property 1.4.1. The apparent contour is tangent to an imaged cross section at any point of contact [32].

Property 1.4.2. The lines tangent to two distinct imaged cross sections $\mathtt{C}$ and $\mathtt{C}'$ at any two points $\mathbf{x}$ and $\mathbf{x}'$ related by the planar homology $\mathtt{W}$ as $\mathbf{x}' = \mathtt{W}\mathbf{x}$ have the same vanishing point $\mathbf{u}_\infty$, which lies on $\mathbf{l}_\infty$.

Property 1.4.3. The 3D points whose images $\mathbf{x} \in \mathtt{C}$ and $\mathbf{x}' \in \mathtt{C}'$ are related as $\mathbf{x}' = \mathtt{W}\mathbf{x}$ belong to the same SOR meridian.

With reference to fig.1.5 the properties are illustrated. The cross-section $\mathtt{C}'$ is tangent to the apparent contour γ and the point $\mathbf{x}'_\gamma$ is the contact point showing the property 1.4.1. The line $\mathbf{l}'$ tangent at $\mathbf{x}'_\gamma$ meet $\mathbf{l}_\infty$ at $\mathbf{u}_\infty$. The line $\mathbf{l}$ passing by $\mathbf{u}_\infty$ and tangent to $\mathtt{C}$ gives the point $\mathbf{x}$. The point $\mathbf{x}$ and $\mathbf{x}'_\gamma$ are the two corresponding point under $\mathtt{W}$ of property 1.4.3 while $\mathbf{l}'$ and $\mathbf{l}$ are the tangent lines of property 1.4.2.

Always with reference to Fig. 1.5, given the apparent contour γ, there exists a unique imaged cross section $\mathtt{C}'$ that includes the generic point $\mathbf{x}'_\gamma \in \gamma$. Correspondingly, once the vanishing line $\mathbf{l}_\infty$ is given, there exists a unique planar homology $\mathtt{W}$ that maps a reference imaged cross section $\mathtt{C}$ onto $\mathtt{C}'$. As $\mathbf{x}'_\gamma$ varies on γ, the vertex $\mathbf{v}_\mathtt{W}$ and the characteristic invariant $\mu_\mathtt{W}$ of $\mathtt{W}$ also vary, while $\mathbf{l}_\infty$ remains fixed. Therefore, as $\mathbf{x}'_\gamma \in \mathtt{C}'$ is moved along γ, it gives rise to a family of planar homologies $\mathtt{W} : \mathtt{C} \longrightarrow \mathtt{C}'$.

Practically the above properties can be translated in a algorithm for computing an imaged meridian whose rectification will furnish the scaling function of the SOR. We now show how to compute the planar homology $\mathtt{W}$ at a given $\mathbf{x}'_\gamma$. According to property 1.4.1, there exists an imaged cross section $\mathtt{C}'$ such that γ and $\mathtt{C}'$ share the same tangent line $\mathbf{l}'$ at $\mathbf{x}'_\gamma$. The tangent line $\mathbf{l}'$ intersects the vanishing line $\mathbf{l}_\infty$ at the point $\mathbf{u}_\infty$: according to property 1.4.2, this is the vanishing point of all the lines which are tangent to the SOR along the same meridian. Therefore, the tangent line $\mathbf{l}$ to $\mathtt{C}$ from $\mathbf{u}_\infty$ meets $\mathtt{C}$ at the point $\mathbf{x}$ such that $\mathbf{x}'_\gamma = \mathtt{W}\mathbf{x}$, and the planar homology vertex $\mathbf{v}_\mathtt{W}$ is the point where the line through $\mathbf{x}$ and $\mathbf{x}'_\gamma$ intercepts the imaged axis of symmetry $\mathbf{l}_s$:

$$\mathbf{v}_\mathtt{W} = (\mathbf{x} \times \mathbf{x}'_\gamma) \times \mathbf{l}_s \ . \tag{1.8}$$

This fixes two of the three degrees of freedom left for $\mathtt{W}$. The remaining degree

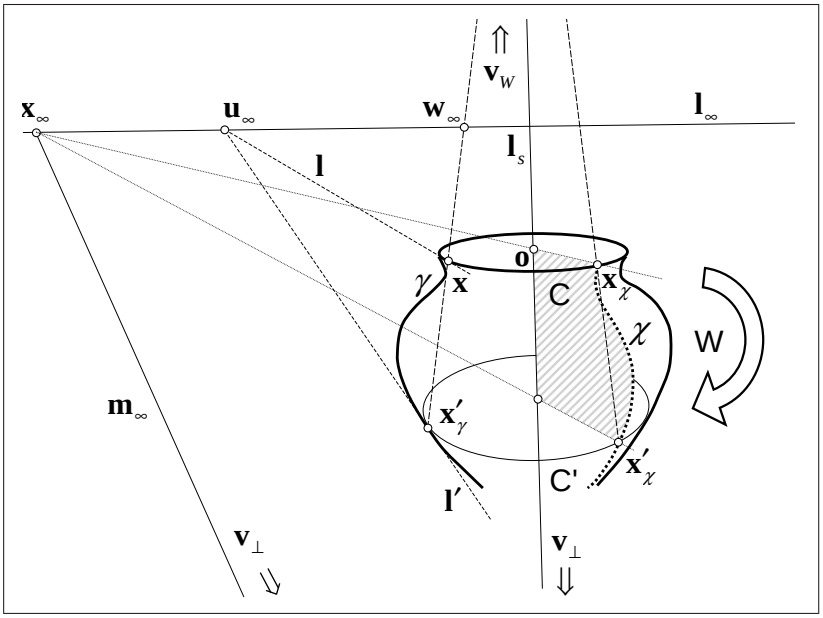


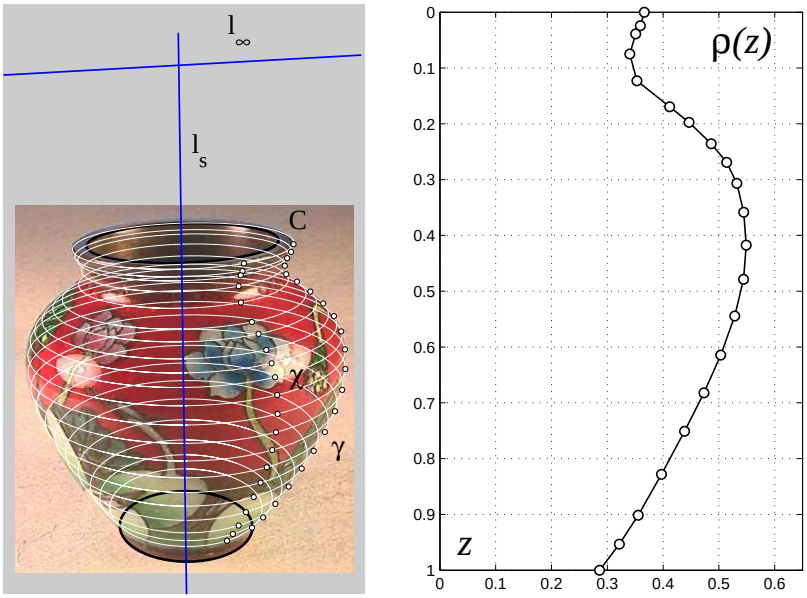
Figure 1.5: The geometry of imaged meridian reconstruction and rectification.

of freedom is fixed by computing the characteristic invariant $\mu_{\mathtt{W}}$ as

$$\mu_{\mathtt{W}} = \{\mathbf{v}_{\mathtt{W}}, \mathbf{w}_\infty, \mathbf{x}, \mathbf{x}'_\gamma\} \ , \tag{1.9}$$

where $\mathbf{w}_\infty = (\mathbf{x} \times \mathbf{x}'_\gamma) \times \mathbf{l}_\infty$ is the point where the line through $\mathbf{x}$ and $\mathbf{x}'_\gamma$ intercepts the vanishing line $\mathbf{l}_\infty$, and $\{\}$ denotes the usual cross ratio of four points [45].

For each $\mathtt{W}$ that is obtained from the steps above, by exploiting the property 1.4.3, a point $\mathbf{x}'_\chi$ on the imaged meridian χ that passes through the point $\mathbf{x}_\chi \in \mathtt{C}$ is computed as $\mathbf{x}'_\chi = \mathtt{W}\mathbf{x}_\chi$. The imaged meridian χ is then recovered as the set of all the points $\mathbf{x}'_\chi$ obtained for different points $\mathbf{x}'_\gamma$ sampled on the apparent contour (see Fig. 1.6, (a)).



(a) (b)

Figure 1.6: Recovery (a) and rectification (b) of an imaged meridian.

Rectification of the imaged meridian

The rectification of χ requires the availability of both the image of the absolute conic ω and the vanishing line $\mathbf{m}_\infty$ of the plane π_χ through the meridian and the SOR symmetry axis. As the SOR symmetry axis lies by construction on π_χ, once the rectifying homography $\mathtt{M}_r$ for this plane is known, we are able to rectify both the imaged meridian χ and the imaged axis of symmetry $\mathbf{l}_s$ according to

$$\begin{cases} \mathbf{x}_\rho = \mathtt{M}_r \mathbf{x}'_\chi \\ \mathbf{l}_z = \mathtt{M}_r^{-\mathrm{T}} \mathbf{l}_s \end{cases} . \tag{1.10}$$

By computing the distance between any point $\mathbf{x}_\rho$ and the line $\mathbf{l}_z$, it is then possible to obtain the values of z and $\rho(z)$ for each $\mathbf{x}'_\chi$ given the reference SOR axis $\mathbf{l}_s$ (see Fig. 1.6(b)).

The vanishing line $\mathbf{m}_\infty$ can be obtained as $\mathbf{m}_\infty = \mathbf{x}_\infty \times \mathbf{v}_\perp$, where $\mathbf{x}_\infty$ and

$\mathbf{v}_\perp$ are respectively the vanishing point of the direction of all lines in π_χ that are orthogonal to the symmetry axis, and the vanishing point of the direction of the symmetry axis (see Fig. 1.5). The vanishing point $\mathbf{x}_\infty$ is computed as

$$\mathbf{x}_\infty = (\mathbf{x}_\chi \times \mathbf{o}) \times \mathbf{l}_\infty = \left(\mathbf{x}_\chi \times \mathtt{C}^{-1}\mathbf{l}_\infty\right) \times \mathbf{l}_\infty \ , \tag{1.11}$$

where $\mathbf{o} = \mathtt{C}^{-1}\mathbf{l}_\infty$ is the image of the center of the cross section that projects onto $\mathtt{C}$; this point is in pole-polar relationship with $\mathbf{l}_\infty$ with respect to $\mathtt{C}$. Since ω is known, the vanishing point $\mathbf{v}_\perp$ is computed as $\mathbf{v}_\perp = \omega^{-1}\mathbf{l}_\infty$. The vanishing line $\mathbf{m}_\infty$ can now be intersected with ω in order to obtain the imaged circular points $\mathbf{i}_\chi$ and $\mathbf{j}_\chi$. This intersection can be algebraically computed by solving for λ the quadratic equation $(\mathbf{x}_\infty + \lambda\mathbf{v}_\perp)^{\mathrm{T}}\,\omega\,(\mathbf{x}_\infty + \lambda\mathbf{v}_\perp) = 0$, where $\mathbf{x}_\infty + \lambda\mathbf{v}_\perp$ denotes the generic point on $\mathbf{m}_\infty$. The imaged circular points are obtained from the two complex conjugate solutions λ_1 and λ_2 respectively as $\mathbf{i}_\chi = (\mathbf{x}_\infty + \lambda_1\mathbf{v}_\perp)$ and $\mathbf{j}_\chi = (\mathbf{x}_\infty + \lambda_2\mathbf{v}_\perp)$. According to [49], the rectifying homography for the plane π_χ is

$$\mathtt{M}_r = \begin{pmatrix} \beta^{-1} & -\alpha\,\beta^{-1} & 0 \\ 0 & 1 & 0 \\ m_1 & m_2 & 1 \end{pmatrix} \ , \tag{1.12}$$

where $\mathbf{m}_\infty = (m_1, m_2, 1)$ and $\mathbf{i}_\chi = \mathrm{conj}(\mathbf{j}_\chi)$ is expressed as $\mathtt{M}_r^{-1}(1, i, 0) = (\alpha - i\beta, 1, -m_2 - m_1\alpha + im_1\beta)$.

Discussion

The above two-step method for 3D metric reconstruction is equivalent to the computation of the set of pairs $\{(z, \rho(z))\}$, where z is the point of the SOR symmetry axis that corresponds to a point $\mathbf{x}'_\gamma$ sampled on the apparent contour γ. This correspondence can be expressed in terms of a function

$\zeta : \gamma \rightarrow [0,1]$ such that

$$z = \zeta(\mathbf{x}'_\gamma) \ . \tag{1.13}$$

The function ζ is defined only at points $\mathbf{x}'_\gamma$ at which γ is smooth and has a unique tangent line. These points belong to a unique imaged cross section $\mathtt{C}'$, whose corresponding pair $(z, \rho(z))$ can then be correctly recovered with the method above. In the presence of self-occlusions, the apparent contour can have singular points at which γ is not smooth and has two distinct tangent lines. The values z_- and z_+ corresponding to the two tangent lines at a singular point delimit the portion of the z axis at which no $\rho(z)$ can be computed with the method above. In this case, the method still guarantees that the scaling function be correctly recovered piecewise as a non connected curve.

If a uniform sampling strategy for γ is used, a non uniform sampling of z is obtained. If, instead, a uniform sampling on the z axis is required, then the inverse of ζ should be used.

However, according to the definition of γ used so far, the function ζ is not invertible. In fact, the apparent contour is split by the imaged axis of symmetry into two halves, whose points correspond in pairs under the harmonic homology. The two points of a pair carry the same reconstruction information, since both of them are mapped by ζ onto the same z. Without loss of generality we can restrict the domain of ζ to one of two halves of γ, say γ', so as to ensure that the function $\zeta^{-1} : [0,1] \rightarrow \gamma'$ exists. This maps any point z at which the value $\rho(z)$ can be recovered with the method above onto a single point $\mathbf{x}'_{\gamma'}$ of the apparent contour.

We adopt an algorithm for the evaluation of $\mathbf{x}'_{\gamma'} = \zeta^{-1}(z)$ at the generic z similar to those described in [50]. The algorithm is reported in detail in Tab.2.2.

1.4.4 Texture acquisition

As shown in Fig. 2.3(a), the SOR texture is the rectangular image $T(\theta, z) = I(\mathbf{x}(\theta, z))$, where I is the image function and $\mathbf{x}(\theta, z)$ is the image projection of the 3D point $\mathbf{P}(\theta, z)$ parametrized as in section 1.3.1.

Texture acquisition following the canonical parameterization θ, z can be solved using the well known cartographic method of *normal cylindrical projection* [51]. However, if parallels and meridians of the imaged object are sampled at regular (θ, z) in the euclidean space, a non-uniform sampling of the texture is created. In order to avoid this, we follow the inverse method (from a regular grid of (θ, z) on the texture plane to points on the image plane) that assures that a uniformly sampled texture is created.
To obtain a metric texture, θ and z are therefore sampled at regular intervals. The resulting texture image has M rows and N columns. The unknown image point $\mathbf{x}(\theta, z)$ is the intersection between the imaged meridian $\chi(\theta)$ corresponding to the SOR meridian at θ and the visible portion of the imaged parallel $\mathtt{C}(z)$ corresponding to the SOR parallel at z. Therefore, the rows of the texture image are composed of image pixels sampled from $\mathtt{C}(z)$ at regular intervals of θ.

A method to sample the visible portion of an imaged parallel $\mathtt{C}(z)$ at a given value of the Euclidean angle θ is described in Appendix A.3. The method permits the Laguerre's formula [12]

$$\theta = \frac{1}{2i} \log(\{\mathbf{v}_\theta, \mathbf{v}_s, \mathbf{i}, \mathbf{j}\}) \tag{1.14}$$

to be inverted so as to compute the vanishing point $\mathbf{v}_\theta$ and to obtain, from this, the sampled point $\mathbf{x}(\theta, z)$—see Fig. 2.3(b). This inversion is necessary to obtain the inverse image trasformation which avoid holes or overlapping while resampling the original image [52].

The algorithm for the computation of a generic texture row $\{T(\theta, z),\ \ \theta = \theta_1, \ldots \theta_N\}$ is:

0. Choose a reference imaged parallel $\mathtt{C}$.
1. Compute $\mathbf{x}'_{\gamma'} = \zeta^{-1}(z)$ as shown in section 1.4.3.
2. Use the planar homology $\mathtt{W}$ associated to $\mathbf{x}'_{\gamma'}$ (see section 1.4.3) to compute the imaged parallel $\mathtt{C}' = \mathtt{W}^{-\mathrm{T}}\mathtt{C}\mathtt{W}^{-1}$.
3. Sample $\mathtt{C}'$ at all values $\theta = \theta_1, \ldots \theta_N$ as described in Appendix A.3.
4. For each of the N points $\mathbf{x}'_{\chi(\theta)} = \mathbf{x}(\theta, z)$ thus obtained, set $T(\theta, z) = I(\mathbf{x}'_{\chi(\theta)})$.

Texture acquisition is achieved by repeating the steps 1 through 4 for all the M rows of the texture image, sampled at regular intervals of z.

It is worth noting that not all the texture image pixels can be computed by the algorithm above. In particular, singular points on the apparent contour γ' due to self-occlusions give rise to row intervals $[z_-, z_+]$ for which the inverse function $\zeta^{-1}(z)$ cannot be computed (see section 1.4.3). A similar situation occurs for the range of θ values for which the surface is not visible. In this case, for each imaged parallel $\mathtt{C}(z)$, the Laguerre's formula—with the value $\mathbf{u}_\infty$ of section 1.4.3 used in Eq. 1.14 in the place of $\mathbf{v}_\theta$—can be used to determine the interval $[\theta_{\min}(z), \theta_{\max}(z)]$ for which the parallel is visible.

The method for texture acquisition described above has some advantages with respect to other solutions presented in the literature. It compute the geometry to image registration avoiding the external calibration, thus requiring only the knowledge of internal camera parameters. This mean that no explicit external calibration and projection of 3D point are computed so

keeping all the computations in image space. It computes a image to image transformation which permits to use the standard parameterization of the SOR to index in the transformed image texture the surface texture. To transform and correctly resample the original image *inverse texture mapping* is used; with this method it is possible to avoids "holes" in the resulting transformed image [52]. To perform this, the inverse image transformation is needed and as shown is computed by an analytical inversion of the angle with the Laguerre formula and by a recursive subdivision scheme similar to [50] for the z.

Moreover the surface texture in this new coordinate system is defined invariantly once that two imaged cross-section are identified in more views of the same SOR. This permits to obtain the full texture by simply register the transformed overlapping texture with a simple translation motion model (see chapter 2).

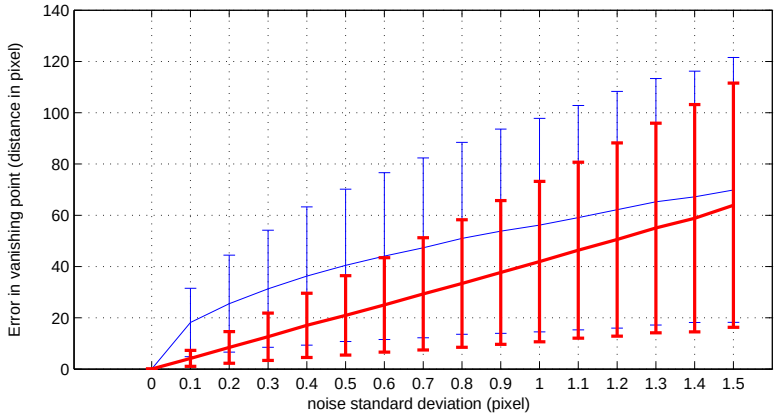
1.5 Experimental Results

1.5.1 Accuracy evaluation

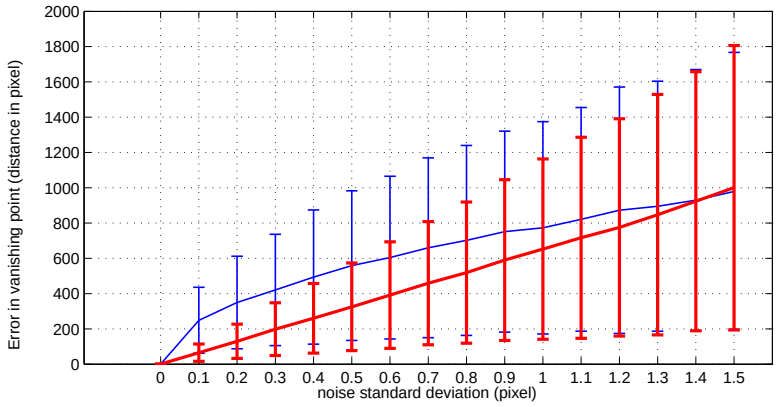
Several experiments were performed in order to test the accuracy of the approach. In particular we assessed the accuracy of the vanishing point estimation, of camera self-calibration and the error in the reconstruction of the SOR scaling function. Two view are chosen as the more representative to evaluate our presented method. These two views are synthetically created by initially positioning a SOR (with scaling function: $\rho(z) = \frac{1}{10}(\cos(\frac{\pi}{2}(\frac{19}{3}z + 1)) + 2)$ with $z \in [0, 1]$) such that the imaged axis of revolution coincides with the y-axis passing by the principal point (this is a degenerate view). From this position the camera it is panned respectively by 13.932° and 3.548°. In

this sense these two view can be termed respectively: *non degeneracy* view and *near degeneracy*. Internal parameters of the camera are set as: $f = 750$, $u_o = 400$, $v_o = 300$. The top and bottom cross sections of the 3D SOR were projected onto the image so as to generate the two corresponding ellipses (this can be done analitically). The ellipses were then resampled with a resolution similar to the image size (800x600) and the resulting points were perturbed with zero mean isotropic Gaussian noise with standard deviation between 0 and 1.5 pixel. These points were successively used to estimate the ellipses with the algorithm described in [53]. Ellipse intersection problem used for the computation of the fixed entities is solved analitically by means of Gröbner basis [47] by finding the roots (numerically) of a quartic polynomial. The system of eq.1.7 is solved using SVD. The influence of this noise was tested by running a Monte Carlo simulation with 10000 trials using different seeds to generate the noise.

In particular we study two scenarios in the two views already mentioned: in the first the imaged elliptical cross-section are partially visible (i.e. only the visible points are resampled from the imaged cross-sections) and in the second the ellipses are fully visible (i.e. all the points from the imaged cross-section are resampled). The former scenario reflects a SOR scene viewed with the camera center in between the two 3d planes containing the cross-section while the latter could reflects two non coplanar 3D points rotating on a turntable. The SOR view with two partially visible cross-section can be regarded as the worst case in terms of camera calibration, since both ellipses are partially visible giving lower point measurements for the ellipse estimation algorithm. A better viewing condition would be obtained by viewing one of the ends of the object; in this condition an entire cross section gives a far better estimated ellipse. We start by showing in fig. 1.7 the accuracy



(a)



(b)

Figure 1.7: Vanishing point estimation accuracy. This figure shows the means and the relative rms of the estimation error in the computation of the vanishing point for different noise level in: (a) non degenerate view condition (ground truth: $\mathbf{v}_\infty = [3421.978 \;\; 209.049 \;\; 1]^{\mathrm{T}}$), (b) degeneracy view condition (ground truth: $\mathbf{v}_\infty = [12493.024 \;\; 206.432 \;\; 1]^{\mathrm{T}}$). The error bars shows the relative upper and lower values of rms error. This gives the reliability of the estimation. The two curves on each graph shows the effect of using only the visible points of the imaged cross-section (light curve) or the fully cross-section (bold curve).

of vanishing point estimation—the most critical among the fixed entities with respect to increased level of noise. In particular Fig. 1.7(a) shows the analysis conducted on the imaged SOR of 1.8(b) and Fig. 1.7(b) shows the

analysis conducted on the imaged SOR of 1.9(b). These are respectively synthetic images of the *non degeneracy* view and *near degeneracy* view. The two curves on each graph of fig.1.7 shows the means and the relative RMS of the estimation error (Euclidean distance w.r.t. the ground truth vector in near degenerate $\mathbf{v}_\infty = [3421.978 \;\; 209.049 \;\; 1]^{\mathrm{T}}$ and non degenerate $\mathbf{v}_\infty = [12493.024 \;\; 206.432 \;\; 1]^{\mathrm{T}}$) in the computation of the vanishing point by using the visible cross-section point measurements (light curve) and by using all the cross-section point measurements (bold curve). The vertical bars represent the upper and lower values of the RMS error which gives the reliability of the estimation. Albeit the graphs looks similar the error scale differ of one order of magnitude. The error appear to be larger when using only the visible cross-section point measurements. While the responses to increasing noise are almost linear if all the points can be measured. The noticeable bias is mainly due to the known bias present in using the algebraic distance in estimating the ellipses which in turn affect the computed fixed entities. Also the estimation of the absolute conic from the fixed entities is estimated using the algebraic distance. This affect substantially the estimation of the internal camera parameter as shown if comparing the two graphs of fig.1.8(a) and of fig.1.9(a). In particular Fig.1.8(a) shows the means and the relative RMS error in the computation of the focal length and principal point for different noise level for the non degenerate view shown in Fig.1.8(b). The two curves on each graph shows the computed estimation error using the visible cross-section point measurements (light curve) and by using all the cross-section point measurements (bold curve). Once more results are represented as error bars showing the mean and the relative RMS error. The lower graphs is the focal length mean error while the upper one is principal point mean distance error. In Fig.1.8(b) it is further shown the view under testing with

the three standard deviation ellipses of the principal point computed from Monte Carlo simulation at increasing noise values by using only the visible ellipse segment. The ground truth principal point position $[400\ 300]^{\mathrm{T}}$ is indicated as a cross. To appreciate better the uncertainty in the principal point estimation of fig.1.8(b), a zoomed figure is shown in fig.1.10(a) and fig.1.10(c) respectively using the visible cross-section and the entire one. Here are reported the three standard deviation ellipses at increasing values of 0.2 for the noise level starting from $\sigma = 0.1$ and ending at $\sigma = 1.5$. It is evident how using all the feature points of the imaged cross-section gives a substantial accuracy in location of the principal point. This is useful for example in calibrating turntable video sequences from two always visible rotating points. The same analysis were conducted for the near degenerate case and results are represented in fig.1.9. If compared with those shown in fig.1.8 regarding the non degenerated view the error is more than doubled. The effect of the increased estimation error is due to the fact that the pole polar constraint becomes irrelevant since the vanishing point is approaching the infinity in the image. This is also verified by the uncertainty analysis shown in fig. 1.10(b) and (d). Fig. 1.10 gives a good description in what happen to the principal point as the SOR position in the image gets closer to the image center, the uncertainty ellipses become increasingly larger, and with the major axis more and more parallel to the imaged symmetry axis. This is motivated by the fact that, in pure degeneracy conditions, an infinite uncertainty affects the principal point coordinate along the imaged symmetry axis.

The average and standard deviation of the error in the reconstruction of the scaling function can be defined respectively as $\int_0^1 |\rho_e(x) - \rho_{gt}(x)|dx$ and $\sqrt{\int_0^1 [\rho_e(x) - \rho_{gt}(x)]^2 dx}$ where $\rho_e(x)$ and $\rho_{gt}(x)$ are respectively the es-

timated and the ground truth scaling functions.

In order to obtain a corrupted smooth curve and to keep tractable the Montecarlo simulation the points of the generating contour were projected in the image plane and a interpolating cubic spline curve passing to 50 equispaced of these points is used as input to our reconstruction algorithm. These points are perturbed with increasing isotropic gaussian noise with zero mean and standard deviation σ measured in pixels. The mean and relative rms error as define before were obtained after running the experiment 10000 times using different seeds to generate the noise. Fig. 1.11 shows the extent to which the presence of noise affects the reconstruction error for the apparent contour and the imaged cross-sections. We choose as test case the near degenerate synthetic SOR views of fig.1.9(b). This can be regarded as a worst case in terms of camera calibration furthermore from this position it can be avoided (at least for non high values of error) the self-occlusion phenomenon which by its nature gives non measurable reconstruction. In particular in Fig. 1.11(a) the light curve shows the error for noisy apparent contour and noisy ellipses (the latter are used to calibrate the camera), while the bold curve show the error when only both cross-sections are noisy. It is evident that the apparent contour is the major responsible for the reconstruction error. Fig. 1.11(b) show the reconstruction error when the light curve of the previous figure (here dashed) is compared with the curve of the reconstruction error with noise only on the apparent contour (continuous curve). It can be observed that the noise on the apparent contour is a predominant source of error for reconstruction. This can be explained by the fact that reconstruction involves the computation of lines tangent to the apparent contour, that are very sensitive to noise. The camera calibration errors are in practical absorbed in the overall error of the reconstruction. This reflect the good

perceptual result obtained in the reconstruction examples giving a chance of existence to the minimalist calibration such as that presented in this chapter.

To further investigate the error in the camera parameters obtained from the imaged cross-section in Fig. 1.11(c) the light curve show the error with visible ellipse segment while keeping a perfect apparent contour. The the bold curve is the same of that in Fig. 1.11(a) here represented at different scales to better appreciate the differences. The jitter-linear behavior is due to the ellipse fitting algorithm which is quite sensible to error in measurement especially if the are not uniform distributed over the curve. We conclude this section showing Fig. 1.11(d) shows the comparison of the reconstruction error when apparent contour and ellipses are corrupted with noise. I particular the the light curve represent the error when only visible segment are used, while the bold curve shows the error when full ellipses are used. The more jagged behaviour of the curves is due mainly to the isotropic noise present at the high curvature segment of ellipses which due the known bias present at those points.

1.5.2 3D SOR reconstruction

Fig. 1.12 shows examples of 3D shape and structure reconstruction from a single uncalibrated view of SOR objects. For each object the original image and the 3D solid obtained after the application of the algorithm are shown. All the images have been obtained with moderate perspective distortion. The apparent contour and cross sections that are needed as an input of the algorithm have been manually extracted following precisely the imaged object boundary; the results shown can therefore be regarded as close to those obtainable in the absence of noise. Fig. 1.12(a) and Fig. 1.12(c)(d) show the reconstruction of a can and of two archaeological vase with differently shaped

profiles. It can be observed that both in the case of curvilinear and linear profiles the SOR structures are correctly reconstructed. Fig. 1.12(b) show the reconstruction of a transparent glass in this case reconstruction with a laser scanner would be difficult to obtain due to the physical properties of the media. Instead, the apparent contour and cross-sections can be easily extracted and measured. Fig. 1.12(e) shows the original (*First row*) and the reconstructed 3D model (*Second Row*) of the first "wireframe" drawing in history, made by the Renaissance florentine artist Paolo Uccello (1397–1475). Also in this case the apparent contour and cross-sections were extracted manually. The wireframe drawing has provided information also for parts of the apparent contour that would have been occluded in a photograph. This has permitted a more complete 3D reconstruction of the object model as can be appreciated in the view of the 3D SOR in Fig. 1.12(e) (*Second Row*). Fig. 1.14(a) present the case of reconstruction of a 3D SOR from a single view with large perspective distortion. Looking at Fig. 1.14(a) (*middle*) it can be noticed that the scaling function is recovered correctly also in this case (the part of the scaling function that corresponds to the cylindrical body of the bottle, is reconstructed as almost perpendicular to the reference plane). Fig. 1.14(a) (*Right*) show the 3D rendered surface obtained from the scaling function. Fig. 1.14(b) *Left* show an example of SOR scene in which severe self-occlusion is present. Self-occlusion can be described in terms aspect graph; this representation was extensively studied in [43] for the class of solids of revolution under perspective view condition. In practical what it can be stated in general is that the relation between the apparent and the scaling function is not only geometric. Self-occlusion can produce visual-events i.e. a particular alignment of features that changes the topology of the view. So in general the apparent contour cannot be 'globally' related to its scaling

function. Fig. 1.14(b) *Right* show that a 3D metric reconstruction is hovewer still possible (albeit partially) if a piece-wise apparent contour is specified. For each apparent contour in the figure γ_1, γ_2, γ_3, γ_4 the corresponding piece-wise scaling functions ρ_1, ρ_2, ρ_3, ρ_4 can be obtained and maintain the global metric structure of the reconstructed SOR scene.

1.5.3 Texture acquisition

Acquisition of the flattened texture permits both the complete three-dimensional reconstruction of the visible part of the SOR object as well as a separate analysis of the true texture properties, regardless of the perspective distortion. Texture flattening makes image details more evident than in the original picture, and also gives the same importance to central and peripheral figures. Fig. 1.15(a) shows the flattened texture from the image of the Greek vase of fig.1.12(c). In this case the original texture is applied to a quasi-spherical surface. While areas are locally preserved, in making the texture a flattened surface, distortions are introduced in those parts of the object surface that either are farthest from the cylindrical surface built around the SOR object or have high curvature. Fig. 1.15(b) shows the case of texture of a cylindrical surface (the can in Fig.1.12(a)). In this case the flattened texture preserves the global geometry of the original surface. This allows to recover the 'AL' mark in the texture, removing the deformation that is present in the original photo. The texture portions closest to the apparent contour have not been considered, in that their re-sampling is typically affected by aliasing due to the strong foreshortening (see Fig. 1.15(c) for the vase of 1.12(d)).

Flattened textures can be easily superimposed on the reconstructed 3D model, so as to obtain photorealistic three-dimensional reconstructions from image data. Fig.1.16 shows the flattened texture of Fig. 1.15(b) and Fig. 1.15(c)

superimposed to the reconstructed 3D can and vase model of Fig.1.12(a) and Fig.1.12(d) respectively.

1.6 Summary

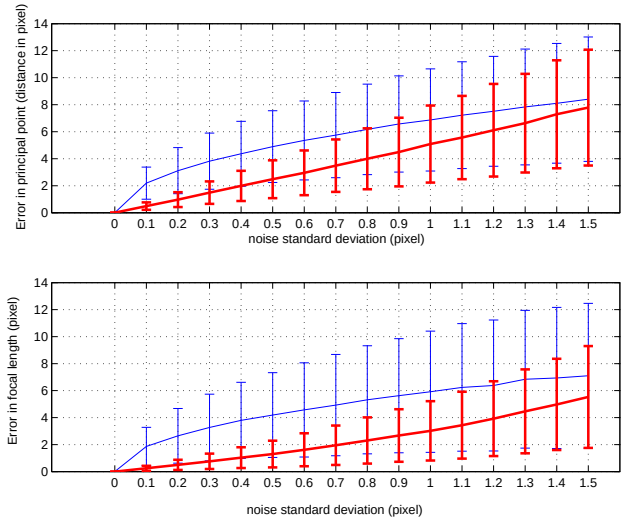
In this chapter we discussed a novel method to recover the original 3D structure of a generic SOR object and its (eventually) superimposed texture from a single uncalibrated view. The solution proposed exploits geometric properties of perspective projection, namely the planar and harmonic homologies and their relationships with internal camera parameters, and the basic properties of SOR objects. Camera calibration is directly obtained from the analysis of imaged SOR features. It has proved to be sufficiently accurate, although sensitive to noise. The 3D structure of the SOR object is derived from the application of the homology constraints to image data - the SOR apparent contour and the elliptic segments of two SOR cross sections - and therefore from calculations in the 2D domain.

Since the homology constraints are of general applicability, the solution can be applied under full perspective conditions to any type of surface of revolution with at least two partly visible cross-sections. In that, the method provides a significant advancement with respect to close research contributions that used homology constraints for 3D recongition/reconstruction, but restricted to the affine projection case [32] or to full perspective of planar surfaces [54] [20].

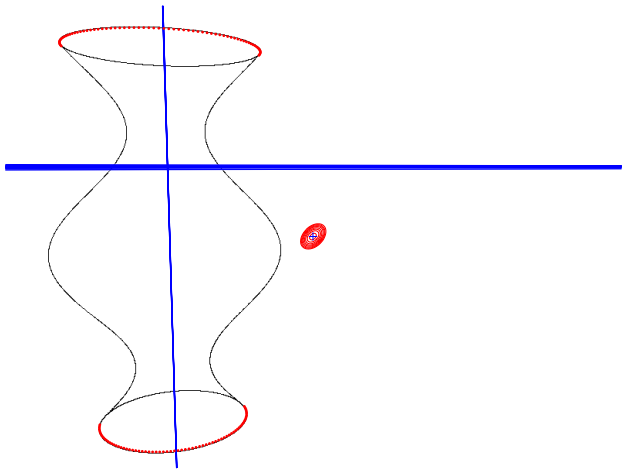
The method can be used reliably, in all those cases in which only a photograph or a drawing of the SOR object is available and structured light acquisition methods cannot be employed for the acquisition of the solid structure. Particularly, in the case of cultural heritage objects that are no more available

as original, or that cannot be moved from their site. It can also be applied in those cases where the nature of the object material makes it impossible or cumbersome acquisition with laser-based techniques. The possibility of recovering the texture superimposed on the SOR as a flattened image allows a complete reconstruction (albeit limited to the imaged part of the object) of the SOR 3D structure and appearance. Descriptions of the 3D shape and textural properties can be used in image database applications to retrieve from large collections, objects that are similar to each other according to their true structural and superficial aspect.

Extraction of the apparent contour and cross section segments, although done manually in the experiments reported in the paper, can be also performed automatically (see for example [55]), with relatively low complexity and good reliability, provided that the object has a dominant size with respect to other objects and non-symmetric textures, and that the background is almost uniform . A research activity on this problem is presently ongoing, aimed at the development of a fully-automated system that creates a 3D model of a SOR from a single uncalibrated view. Finally, as a natural extension of the present work, a complete reconstruction of the 3D textured SOR object can be easily obtained by mosaicing multiple views of the same object, taken from the same camera under the same illumination conditions.

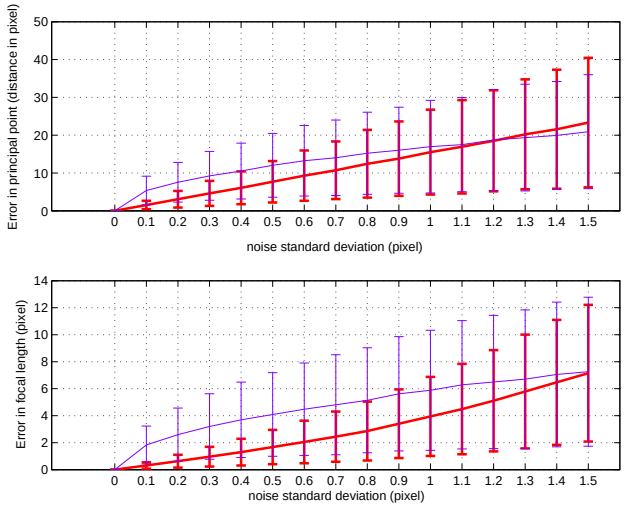


(a)

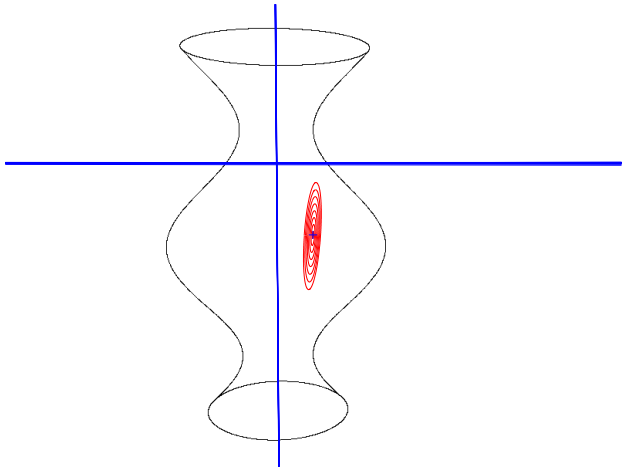


(b)

Figure 1.8: The figure (a) shows the means and the relative RMS error in the computation of the principal point (upper) and focal length (lower) for different noise level of the synthetic non degenerate view shown in (b). The two curves on each graph shows the effect of using only the visible points of the imaged cross-section (light curve) or the fully cross-section (bold curve). In particular in (b) are shown the three standard deviation ellipses of the principal point computed from Monte Carlo simulation at increasing noise values. The ground truth $[400\ 300]^{\mathrm{T}}$ is plotted as a a cross.



(a)



(b)

Figure 1.9: The figure (a) shows the means and the relative RMS error in the computation of the principal point (upper) and focal length (lower) for different noise level of the synthetic near degeneracy view shown in (b). The two curves on each graph shows the effect of using only the visible points of the imaged cross-section (light curve) or the fully cross-section (bold curve). In particular in (b) are shown the three standard deviation ellipses of the principal point computed from Monte Carlo simulation at increasing noise values. The ground truth $[400\ 300]^{\mathrm{T}}$ is plotted as a a cross. If compared with 1.8 the error is more than doubled.

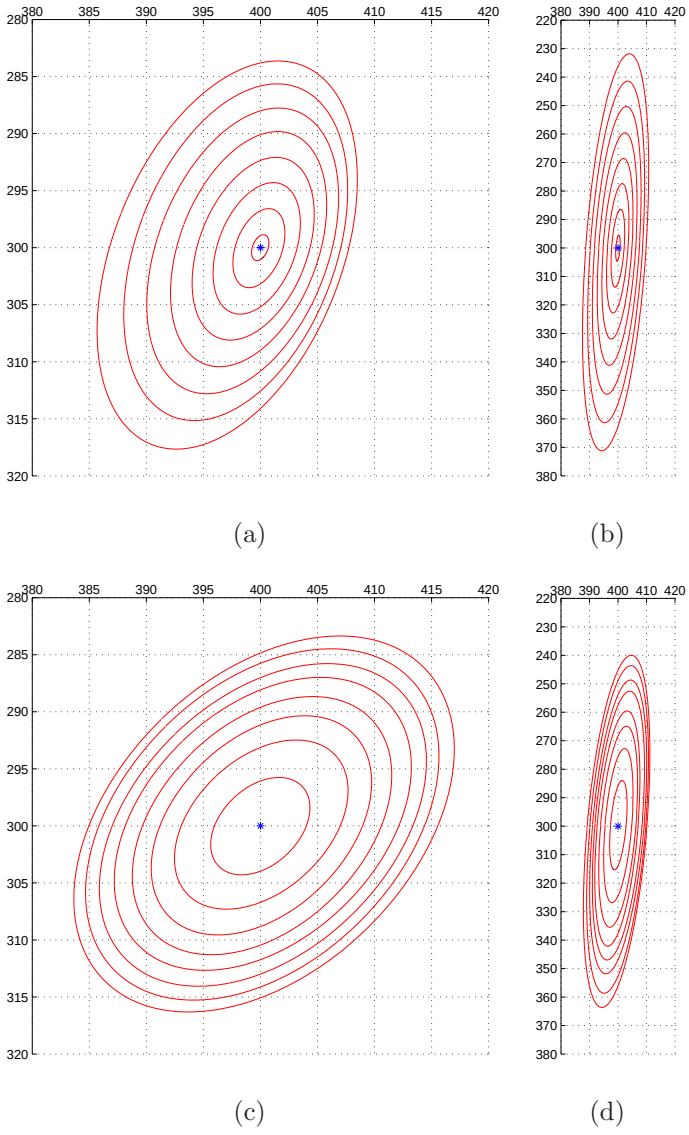


Figure 1.10: The figures shows three standard deviation ellipses at increasing values of increment 0.2 for the noise level starting from $\sigma = 0.1$ and ending at 1.5 for the estimation of the principal point. In particular in fig.(a) and fig.(b) are shown the results respectively for the non degenerate and near degenerate views using all point measurements. While in fig.(c) and fig.(d) are shown the results respectively for the non degenerate and near degenerate views using instead only the visible point measurements.

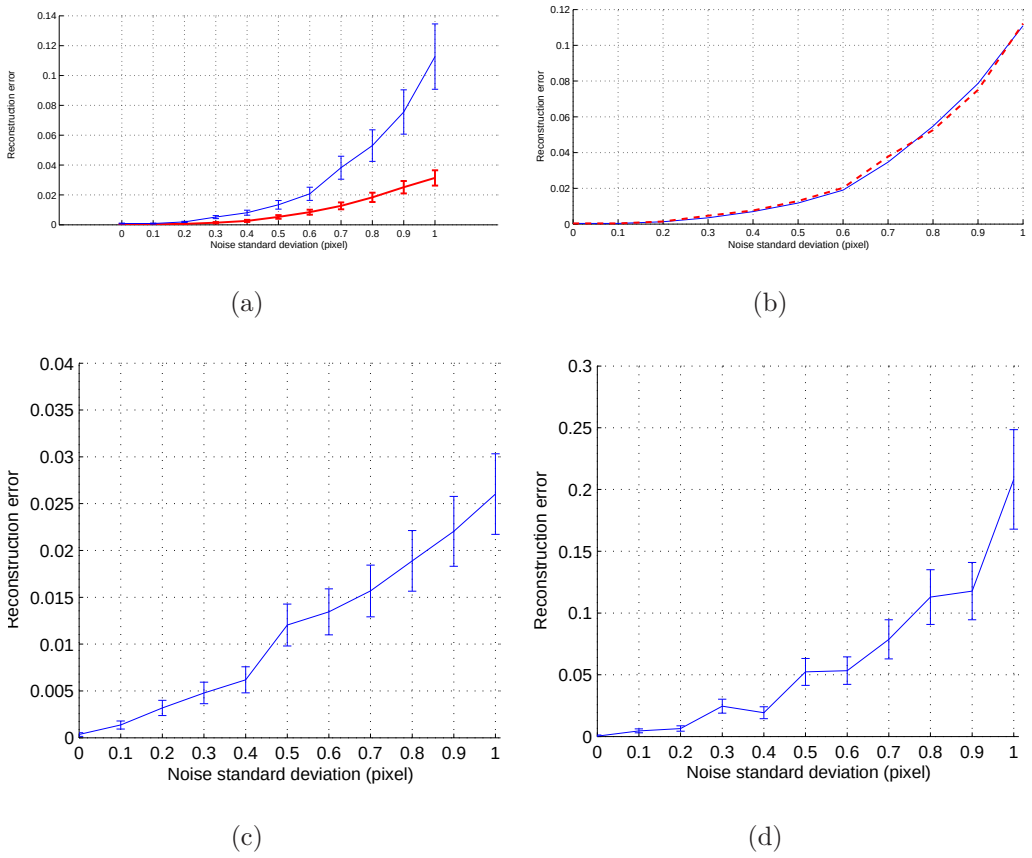


Figure 1.11: Residual reconstruction error. (a): The light curve shows the mean and the RMS for noisy apparent contour and noisy ellipses (the latter are used to calibrate the camera), the bold curve show the error when only both cross-sections are noisy. (b): comparison between the reconstruction with noisy ellipses (dashed curve) and with noise only in the apparent contour (continuous curve). (c): The same bold curve in (a) is shown here with the reconstruction error when noisy visible segment of ellipses are tested (light curve). (d): comparison between noisy visible ellipses and noisy apparent contour (light curve) and noisy ellipses and apparent contour (bold curve).

(a) (b) (c) (d)

Figure 1.12: Typical SOR objects and their reconstructed models. (*Top*) Single uncalibrated views of three SORs. *Bottom*: Corresponding 3D models reconstructed.

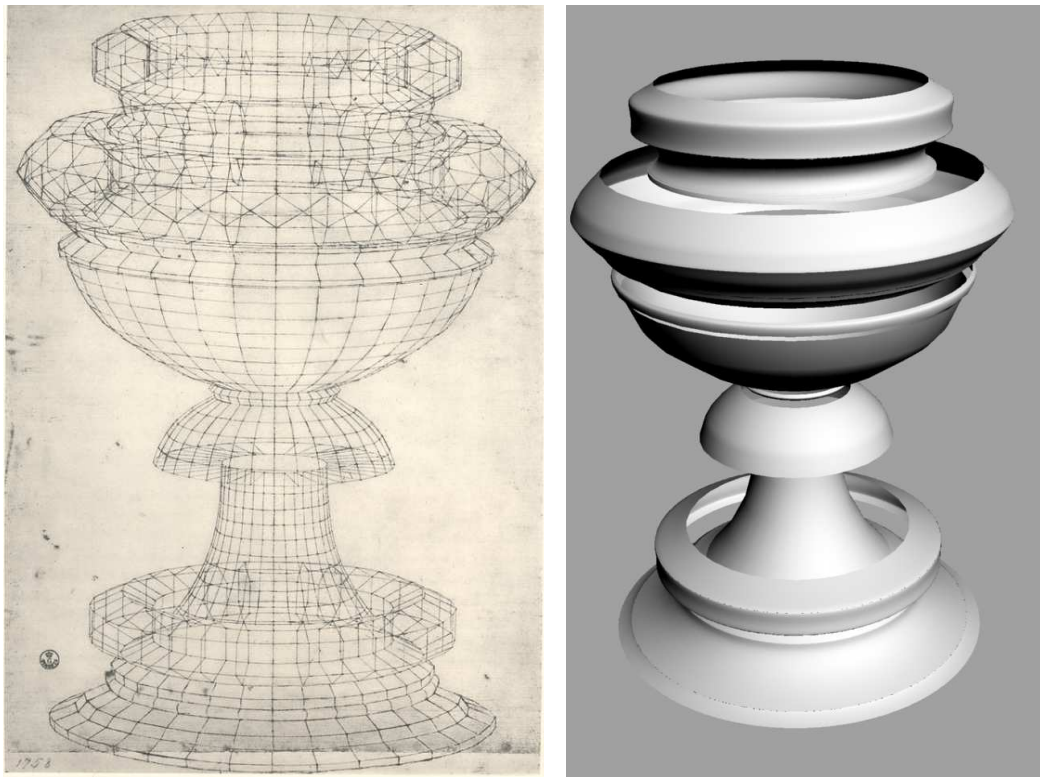
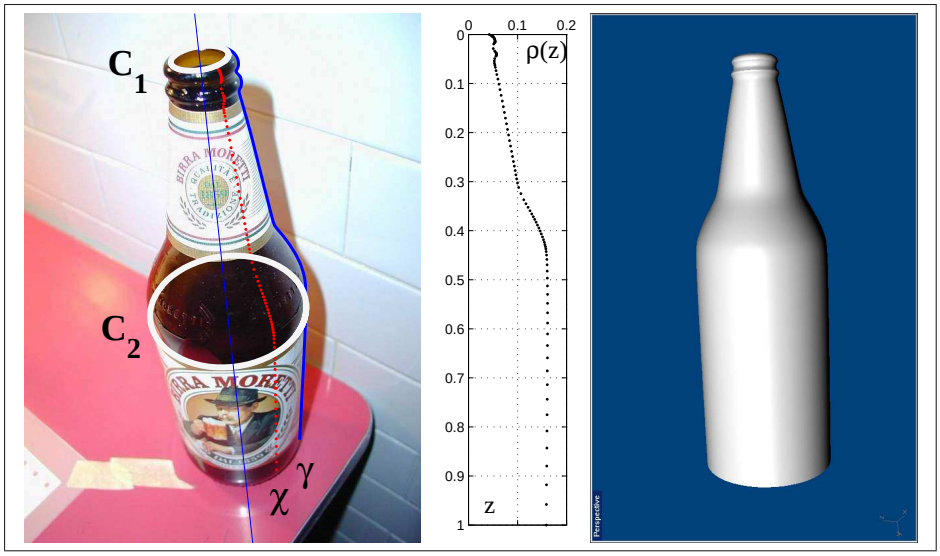
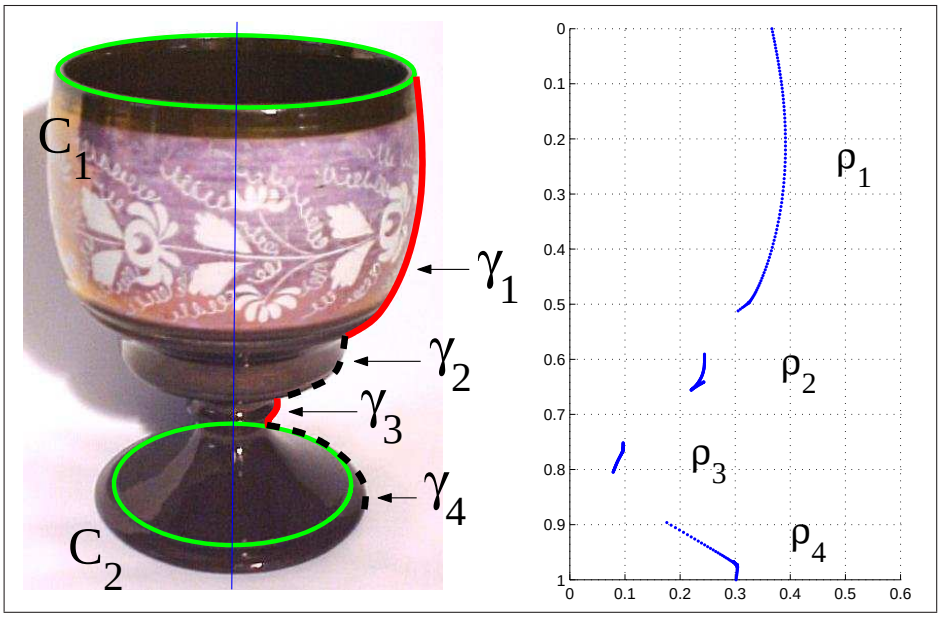


Figure 1.13: *Left*:Paolo Uccello's wireframe drawing of a chalice. *Middle and Right*: two views of the reconstructed model with evidence of self-occluded parts.



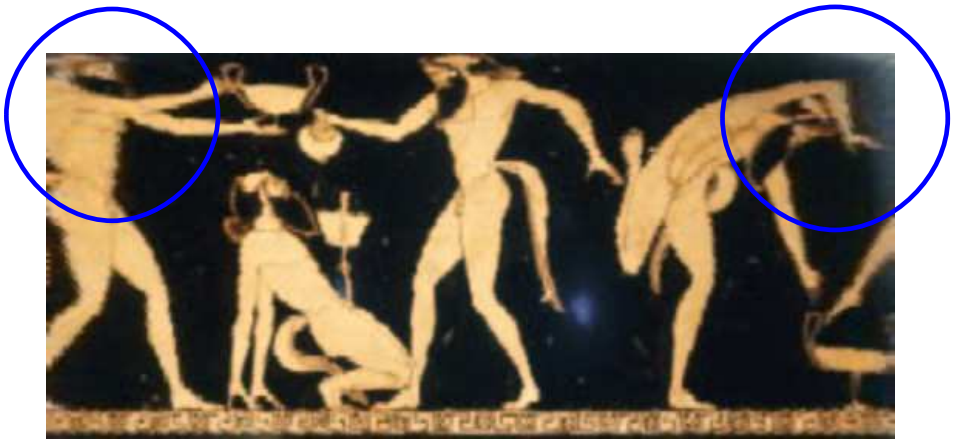
(a)



(b)

Figure 1.14: (a) *Left*: A SOR view taken under strong perspective conditions with indications of the two cross section C_1 and C_2, the apparent contour γ and the projection of the SOR meridian χ. (a) *Middle*: The SOR scaling function rectified. (a) *Right*: The 3D reconstructed model. (b) *Left*: A SOR view with severe self-occlusion with specified piece-wise unoccluded apparent contour. (b) *Right*: The partial reconstructed metric scaling function.

.



(a)

(b) (c)

Figure 1.15: (a): Flattened texture from the image of the archaeological vase from Fig.1.12(c). Surface region with the largest distortion are indicated with circles. (a) The flattened textures for the can and the Chinese vase of Fig.1.12(a) and Fig.1.12(d) respectively.

Figure 1.16: The textured 3D reconstruction for the can and the Chinese vase of Fig.1.12(a) and Fig.1.12(d) respectively.

Chapter 2

Mosaicing

We present a novel approach to obtain a mosaic image for the surface texture content of a surface of revolution (SOR) from a collection of uncalibrated views. The SOR scene constraint is used to calibrate each view and align the corresponding pictorial content into a global representation. Metric surface properties are extracted from each view by exploiting special properties of the imaged SOR geometry expressed in terms of homologies. Image alignment is achieved by projecting imaged surface elements onto a reference plane, and then registering them according to a translational motion model. This work extends previous research on calibrated scenes of right circular cylinders to the more general case of uncalibrated SOR scenes. Experimental results with images taken from the web demonstrate the effectiveness and the general applicability of the approach.

2.1 Introduction and Related Work

Image mosaicing consists in merging collections of images having a partially overlapping content. The process can be decomposed into three main steps. First, the transformations relating each image coordinate system with a ref-

erence one are computed; the images are then warped according to the associated transformations, and finally aligned to each other to compose a single mosaic image.

Several approaches were proposed in the literature, mainly differing in the alignment methods adopted and in the class of warping transformations considered [56], [57]. The most common warping transformations fall into the class of projective and affine planar parametric models, related to special camera motions (e.g., rotations only) and/or scene geometry (e.g., planar scenes) [58], [11]. A typical example is that of panoramic mosaics, which are obtained either with a pan-only or with a pan/tilt camera, and are further warped onto a cylindrical or a spherical manifold, respectively, to obtain a 360° panorama [59], [60], [61]. A general technique for projecting an image mosaic onto a curved manifold according to camera motion is developed in [62].

If camera motion is unconstrained but the scene is planar, standard parametric models (i.e., planar homographies) can still be used for mosaicing purposes. This is not the case for general curved scenes, that require image transformation models which are more complex than homographies and are typically—even if not always [63]—non parametric. Due to its intrinsic difficulties, the problem of curved scene mosaicing under general motion has been largely neglected by the research community. As an exception, a method for mosaicing the pictorial content painted on right circular cylinders was presented in [37]. That method requires that internal camera parameters are known in advance. The external orientation parameters and the imaged symmetry axis are obtained from two imaged circular cross sections. As the method does not fully exploit prior knowledge about scene geometry, pictorial surface elements cannot be metrically sampled in the warping step,

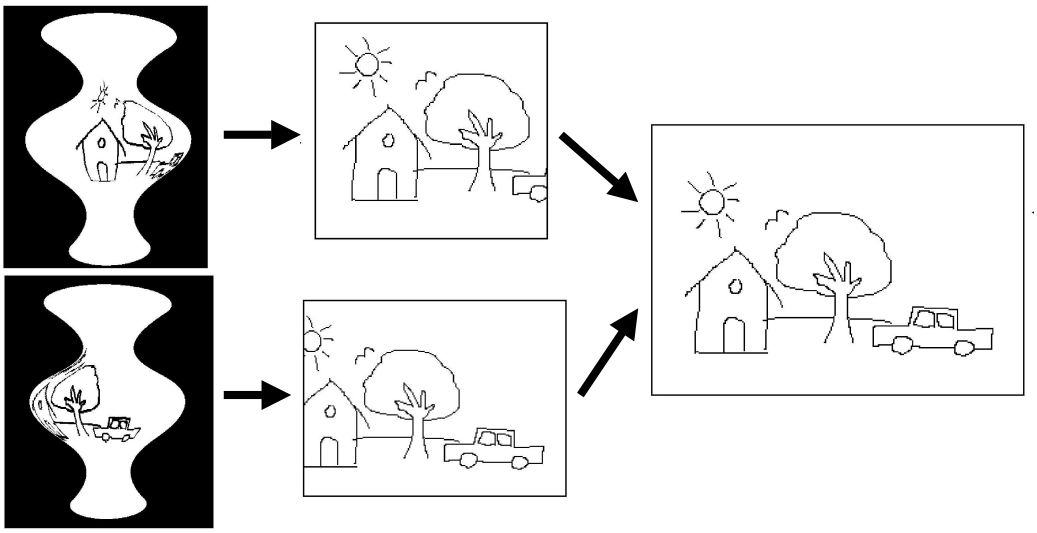


Figure 2.1: Mosaicing from two SOR images.

and geometric distortions are introduced. This may affect significantly the subsequent alignment step and the visual quality of the obtained mosaic.

In this chapter, a novel approach is presented for the creation of mosaics from collections of uncalibrated perspective views of a Surface of Revolution (SOR). A projective model of SOR scene geometry based on homologies is used both to calibrate each single view and to align metrically the corresponding surface pictorial content from each view according to a translational motion model. The original contribution of this work is two-fold. First, the approach extends previous literature on curved surfaces to the broad class of SOR objects. Second, calibration information need not be known in advance, but can be obtained directly from SOR scene geometry. Experimental results with images taken from the web demonstrate the effectiveness and the general applicability of the approach. In particular, the visual quality of the results is comparable with the one obtained with expensive 3D laser scanning technologies (see e.g. [64]) typically used for cultural heritage applications.

2.2 The Approach

Fig. 2.1 shows an example of mosaic creation from two SOR views. The imaged SOR surface regions visible from each view (Fig. 2.1, left) are first individually mapped onto a reference plane (Fig. 2.1, middle), whereon they can be aligned together and merged into a single mosaic image (Fig. 2.1, right).

The mosaicing approach consists of three main steps. In the first step, the imaged SOR geometry of each view is estimated from the visible segments of two imaged cross sections, and used to compute the internal camera parameters. The imaged SOR geometry is then exploited together with the calibration information and the imaged SOR silhouette to obtain, for each view, a SOR parameterization common to all the views up to a translation (second step). This removes the projective distortion due to the image formation process, and allows the imaged SOR regions visible from each view to be warped onto a common reference plane. In the final step (alignment and compositing), the unknown translation for each warped image is estimated by region-based image registration, and used to create the mosaic image. In the following sections, each of the steps above will be described in detail.

2.2.1 Imaged SOR Geometry and Camera Calibration

A SOR can be parameterized as $\mathbf{P}(\theta, z) = (\rho(z)\cos(\theta), \rho(z)\sin(\theta), z)$, where $\theta \in [0, 2\pi]$ and $z \in [0, 1]$. The scaling function $\rho(z)$ controls the 3D shape of the SOR. (In the special case of constant $\rho(z)$, the SOR reduces to the right circular cylinder addressed in [37].) The perspective projection of a SOR gives rise to two different kinds of image curves, namely the *apparent contour* and the *imaged cross sections*. The former is the image of the points

at which the surface is smooth and the projection rays are tangent to the surface; the shape of this curve is view dependent. On the other hand, imaged cross sections are view independent elliptical curves, that correspond to parallel coaxial circles in 3D and arise from surface normal discontinuities or surface texture content. Both the apparent contour and the imaged cross sections of a SOR are transformed onto themselves by a 4-dof harmonic homology $\mathtt{H} = \mathtt{I} - 2\frac{\mathbf{v}_\infty \mathbf{l}_s^{\mathrm{T}}}{\mathbf{v}_\infty^{\mathrm{T}} \mathbf{l}_s}$, where $\mathbf{l}_s$ and $\mathbf{v}_\infty$ are respectively the imaged axis of revolution and the vanishing point of the normal direction of the plane passing through $\mathbf{l}_s$ and the camera center [17]. These geometric entities are strictly related to the calibration matrix $\mathtt{K}$, which embeds information about the internal camera parameters. In particular it holds $\mathbf{l}_s = \omega \mathbf{v}_\infty$, where $\omega = \mathtt{K}^{-\mathrm{T}}\mathtt{K}^{-1}$ is the image of the absolute conic [11]. Moreover, since cross sections are parallel circles in 3D, they intersect at the circular points of the families of planes orthogonal to the SOR symmetry axis. Their projection in the image, $\mathbf{i}$ and $\mathbf{j}$, are also related to the image of the absolute conic as $\mathbf{i}^{\mathrm{T}}\,\omega\,\mathbf{i} = 0$ and $\mathbf{j}^{\mathrm{T}}\,\omega\,\mathbf{j} = 0$. The resulting system of eq.1.7 here reported for convenience

$$\begin{cases} \mathbf{i}^{\mathrm{T}}\,\omega\,\mathbf{i} = 0 \\ \mathbf{j}^{\mathrm{T}}\,\omega\,\mathbf{j} = 0 \\ \mathbf{l}_s = \omega \mathbf{v}_\infty \end{cases} \tag{2.1}$$

provides four linear constraints on ω, whose coefficients can be computed from two imaged ellipses as shown in [4] [65]. A symbolic representation of the geometrical relationships involved in Eq. 1.7 is shown in Fig. 1.4. It can be demonstrated that only three out of the four constraints above are actually independent. Therefore, the system of Eq. 1.7 can be used to calibrate a natural camera (zero skew and known aspect ratio: 3 dofs) from a single image.

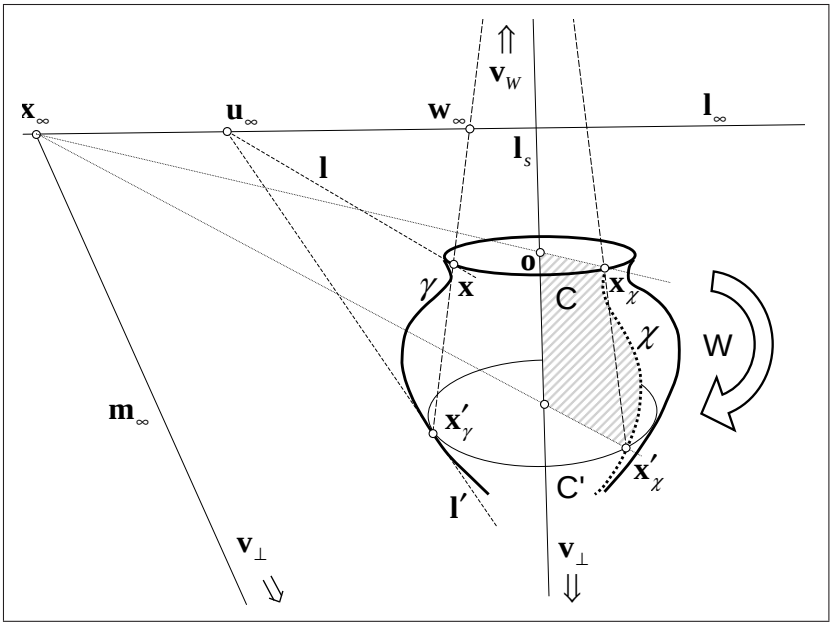


Figure 2.2: The geometry of imaged SOR parametrization.

2.2.2 Imaged SOR Parameterization

The image of any SOR point can be backprojected uniquely onto the coaxial right cylinder $(\cos(\theta), \sin(\theta), z)$; this, in turn, can be unrolled onto the plane (θ, z) used as reference plane for mosaicing. The imaged SOR parametrization just described is nothing but the (unknown) warping transformation connecting the image and the mosaicing planes. This section shows how to evaluate the imaged SOR parametrization from the imaged SOR geometry introduced in the previous section and the apparent contour.

The three properties 1.4.1 1.4.2 1.4.3 hold; with reference to Fig.2.2, the cross section $\mathtt{C}'$ is tangent to the apparent contour γ and the point $\mathbf{x}'_\gamma$ is the contact point of property 1.4.1. The line $\mathbf{l}'$ tangent at $\mathbf{x}'_\gamma$ meets $\mathbf{l}_\infty$ at $\mathbf{u}_\infty$. The line $\mathbf{l}$ passing by $\mathbf{u}_\infty$ and tangent to $\mathtt{C}$ gives the point $\mathbf{x}$. The points $\mathbf{x}$ and $\mathbf{x}'_\gamma$ correspond under the homology $\mathtt{W}$ of property 1.4.3, while $\mathbf{l}'$ and $\mathbf{l}$ are the tangent lines of property 1.4.2. The properties above are

Objective:

- Given $\mathtt{C}$, $\mathbf{l}_\infty$, $\mathbf{l}_s$, $\mathbf{v}_\infty$ and $\mathbf{x}'_\gamma \in \gamma$. Compute the imaged cross section $\mathtt{C}'$ tangent at $\mathbf{x}'_\gamma$.

Algorithm:

1. Compute the tangent line $\mathbf{l}'$ at $\mathbf{x}'_\gamma \in \gamma$.
2. $\mathbf{u}_\infty = \mathbf{l}' \times \mathbf{l}_\infty$
3. Compute $\mathbf{x} \in \mathtt{C}$ at which the tangent line $\mathbf{l}$ is incident with $\mathbf{u}_\infty$
4. $\mathbf{v}_\mathtt{W} = (\mathbf{x} \times \mathbf{x}'_\gamma) \times \mathbf{l}_s$
5. $\mathbf{w}_\infty = (\mathbf{x} \times \mathbf{x}'_\gamma) \times \mathbf{l}_\infty$
6. $\mu_\mathtt{W} = \{\mathbf{v}_\mathtt{W}, \mathbf{w}_\infty, \mathbf{x}, \mathbf{x}'_\gamma\}$
7. $\mathtt{W} = \mathtt{I} + (\mu_\mathtt{W} - 1)\frac{\mathbf{v}_\mathtt{W}\mathbf{l}_\infty^\mathrm{T}}{\mathbf{v}_\mathtt{W}\mathbf{l}_\infty}$
8. $\mathtt{C}' = \mathtt{W}^{-\mathrm{T}}\mathtt{C}\mathtt{W}^{-1}$.

Table 2.1: Computation of the imaged cross section $\mathtt{C}'$ tangent to an assigned point $\mathbf{x}'_\gamma$ on the apparent contour γ. (see Fig.2.2)

used in the algorithms shown in Tables 2.1, 2.2(a), 2.2(b) and 2.3 to solve the imaged SOR parametrization problem. The algorithms in Tab. 2.3 is the top-most algorithm performing image transformation by using the algorithms in Tabs. 2.1, 2.2(a) , 2.2(b). The algorithm in Tab. 2.1 computes the imaged cross section $\mathtt{C}'$ tangent at γ at its generic point $\mathbf{x}'_\gamma \in \gamma$ by transforming a visible cross section $\mathtt{C}$. The inputs $\mathbf{l}_\infty$, $\mathbf{l}_s$, $\mathbf{v}_\infty$ are computed from two visible cross section as described in [4]. This algorithm allows one to "move" projectively along imaged cross sections. In particular Tab. 2.2(a) and 2.2(b) show how to index a generic imaged surface element with a unique value of θ and z respectively.

Solving for the Euclidean θ. The angle between two lines in a world plane π can be computed in the image in terms of the vanishing points of the lines and the imaged circular points of the plane as shown in Fig. 2.3(a).

In the figure, the Euclidean (world angle) θ between the two imaged lines $\mathbf{l}_\theta$ and $\mathbf{l}_s$ can be calculated by the Laguerre's formula [12]

$$\theta = \frac{1}{2i}\log(\{\mathbf{v}_\theta, \mathbf{v}_s, \mathbf{i}, \mathbf{j}\}) \ , \tag{2.2}$$

where $\{\}$ denotes the usual cross ratio of four points [45]. In order to obtain an imaged point at a given θ in a generic imaged cross section $\mathtt{C}$, the Eq. A.10 is inverted. By expressing the generic point on the vanishing line $\mathbf{l}_\infty$ of π as $\mathbf{v}(\lambda) = \mathbf{i} + \lambda(\mathbf{i} - \mathbf{j})$, Eq. A.10 can be rewritten as

$$e^{i2\theta} = \{\lambda_\theta, \lambda_s, \lambda_i, \lambda_j\} \ , \tag{2.3}$$

where λ_θ, λ_s, $\lambda_i = 0$ and $\lambda_j = -1$ are the values of the complex parameter λ respectively for the points $\mathbf{v}_\theta$, $\mathbf{v}_s$, $\mathbf{i}$ and $\mathbf{j}$. Given the imaged axis of revolution $\mathbf{l}_s = (l_1, l_2, l_3)$ and the imaged circular points $\mathbf{i} = \mathrm{conj}(\mathbf{j}) = (a + ib, c + id, 1)$, by solving for λ_s the equation $\mathbf{l}_s^{\mathrm{T}}\mathbf{v}(\lambda_s) = 0$ we get $\lambda_s = -\frac{1}{2}\,[1 + i\tan\phi_s]$, where

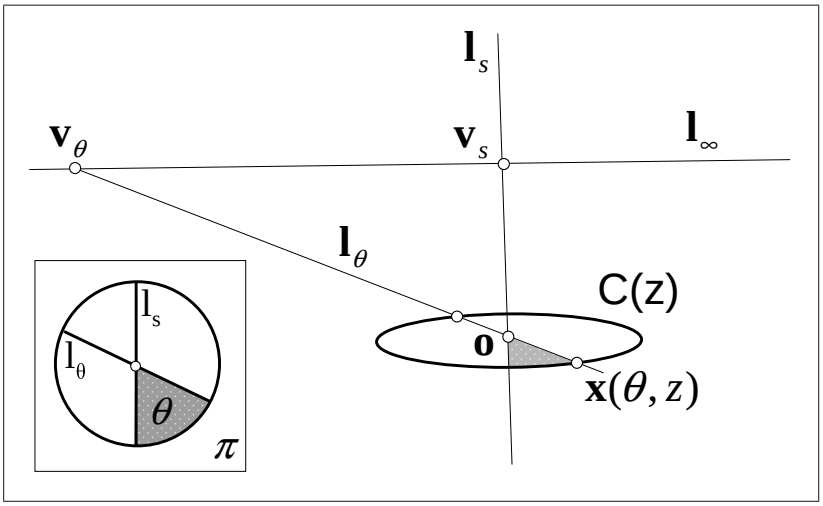
$$\phi_s = \arctan\left(-\frac{l_1 a + l_2 c + l_3}{l_1 b + l_2 d}\right) \ . \tag{2.4}$$

By replacing the above value of λ_s into Eq. A.12, the value of λ_θ can be computed as

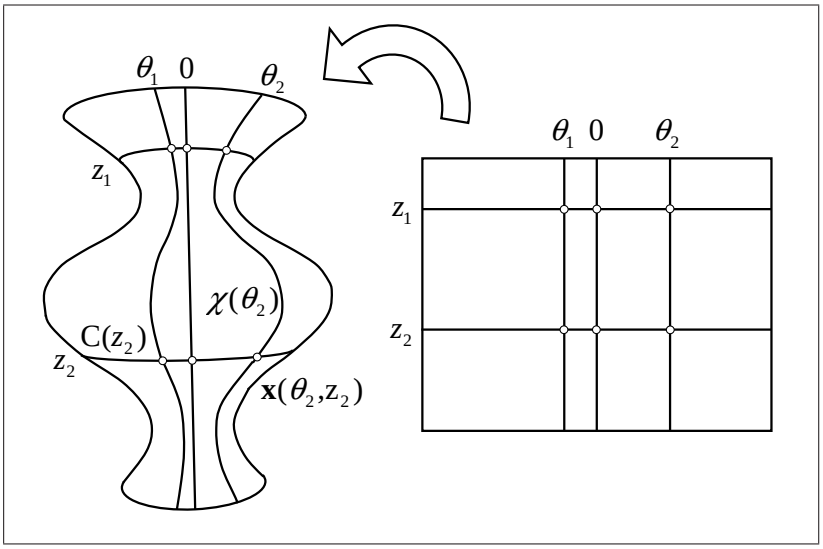
$$\lambda_\theta = -\frac{1}{2}\,[1 + i\tan(\phi_s + \theta)] \ , \tag{2.5}$$

which yields the vanishing point as $\mathbf{v}_\theta = \mathbf{i} + \lambda_\theta(\mathbf{i} - \mathbf{j})$. The image line $\mathbf{l}_\theta = \mathbf{v}_\theta \times \mathbf{o}$—where $\mathbf{o} = \mathtt{C}^{-1}(z)\mathbf{l}_\infty$ is the image of the cross section center—intercepts the imaged parallel $\mathtt{C}$ at two points, of which the required point $\mathbf{x}(\theta, z)$ on the visible imaged meridian $\chi(\theta)$ is the farthest one from $\mathbf{v}_\theta$ along the line $\mathbf{l}_\theta$. Tab. 2.2(a) summarizes the described algorithm.

Solving for the metric z. For any given θ, the algorithms of Tab. 2.1 and Tab. 2.2(a) can be used to obtain the whole imaged meridian $\chi(\theta)$. The z value associated to each of the points of this imaged meridian can be



(a)



(b)

Figure 2.3: (a) Sampling an imaged cross section $\mathtt{C}(\mathtt{z})$ at a given Euclidean angle θ. (b) The texture transformation.

obtained by rectification of the plane π_χ through the meridian. The rectifying homography [7] can be computed from the image of the absolute conic ω and the vanishing line $\mathbf{m}_\infty$ of the plane π_χ as 1.12, here reported for convenience:

$$\mathtt{M}_r = \begin{pmatrix} \beta^{-1} & -\alpha\,\beta^{-1} & 0 \\ 0 & 1 & 0 \\ m_1 & m_2 & 1 \end{pmatrix} ,$$

where $\mathbf{m}_\infty = (m_1, m_2, 1)$ and $\mathbf{i}_\chi = \mathrm{conj}(\mathbf{j}_\chi)$ is expressed as $\mathtt{M}_r^{-1}(1, i, 0) = (\alpha - i\beta, 1, -m_2 - m_1\alpha + im_1\beta)$. The vanishing line $\mathbf{m}_\infty$ is obtained as $\mathbf{m}_\infty = \mathbf{x}_\infty \times \mathbf{v}_\perp$ (see Fig. 2.2), where $\mathbf{x}_\infty$ and $\mathbf{v}_\perp$ are respectively the vanishing point of the θ-direction of all lines in π_χ that are orthogonal to the symmetry axis, and the vanishing point of the direction of the symmetry axis. The imaged circular points are computed as the intersection of $\mathbf{m}_\infty$ with ω. As the SOR symmetry axis lies by construction on π_χ, once the rectifying homography $\mathtt{M}_r$ for this plane is known, we are able to rectify both the imaged meridian χ and the imaged axis of symmetry $\mathbf{l}_s$, thus obtaining the required value of z.

The correspondence between a point of the apparent contour $\mathbf{x}'_\gamma \in \gamma$ at which the imaged cross section is tangent-contact and the (normalized) metric z where the 3D cross-section resides can be expressed in terms of a function $\zeta : \gamma \rightarrow [0, 1]$ such that

$$z = \zeta(\mathbf{x}'_\gamma) \ .$$

An algorithm for the computation of $\mathbf{x}'_\gamma = \zeta^{-1}(z)$ at the generic z by successive approximations is outlined in Tab. 2.2(b). This algorithm is essential for image warping by inverse texture sampling. The unknown $\mathbf{x}'_\gamma$ is denoted as $\gamma(t)$, where $t \in [0, 1]$ is any curve parameter on γ such that $\zeta(\gamma(0)) = 0$. Hence, the problem can be reformulated as to find the value of t which satisfies $\zeta(\gamma(t)) = z$. The algorithm exploits the fact that the function $(\zeta \circ \gamma)(t)$ is monotonic.

(a)

Objective: Given an angle θ in the world space, the fixed entities $\mathbf{l}_\infty$, $\mathbf{l}_s$, $\mathbf{v}_\infty$ $\mathbf{i}, \mathbf{j}$ and the conic $\mathtt{C}$. Compute the imaged point $\mathbf{x} \in \mathtt{C}$ subtending an angle θ with respect to the plane through the camera center and the 3D axis of revolution.

Algorithm:

1. Compute λ_θ as in Eq. 2.5
2. $\mathbf{v}_\theta = \mathbf{i} + \lambda_\theta(\mathbf{i} - \mathbf{j})$.
3. $\mathbf{v}_s = \mathbf{l}_s \times \mathbf{l}_\infty$
4. $\mathbf{l}_\theta = \mathbf{v}_\theta \times (\mathtt{C}^{-1}\mathbf{l}_\infty)$
6. Intersect $\mathtt{C}$ with $\mathbf{l}_\theta$ an choose from the two solution the farthest from $\mathbf{v}_\theta$.

(b)

Objective: Given z compute its corresponding point $\mathbf{x}'_\gamma \in \gamma$ at which the cross section at z is tangent contact to $\mathbf{x}'_\gamma$

Algorithm:

0. Set $t_- = 0$ and $t_+ = 1$.
1. Choose t as the midpoint of $[t_-, t_+]$ and set $\mathbf{x}'_\gamma = \gamma(t)$.
2. Compute $\hat{z} = \zeta(\mathbf{x}'_\gamma)$.
3. If $|z - \hat{z}| < \Delta z$ stop.
4. If $z > \hat{z}$ set $t_- = t$; else set $t_+ = t$; go to 1.

Table 2.2: (a) Imaged cross section sampling at a given θ. (b) Iterative evaluation of the point on the apparent contour γ that lies at a given z.

Objective:

- Compute the flattened imaged SOR region T

Algorithm:

1. Choose a reference imaged parallel $\mathtt{c}$.
2. Compute $\mathbf{x}'_\gamma = \zeta^{-1}(z)$ and the relative imaged cross section $\mathtt{c}(\mathbf{z})'$ with the algorithm of Tab. 2.2(b).
3. Sample $\mathtt{c}(\mathbf{z})'$ at $\theta = \theta_1, \ldots \theta_N$ with the algorithm of Tab. 2.2(a).
4. For each of the N points $\mathbf{x}'_{\chi(\theta)} = \mathbf{x}(\theta, z)$ thus obtained, set $T(\theta, z) = I(\mathbf{x}'_{\chi(\theta)})$.
5. Texture acquisition is achieved by repeating the steps 1 through 4 for all the rows of the texture image T, sampled at regular intervals of z.

Table 2.3: Texture transformation algorithm.

We are now finally in the position to perform image warping according to the algorithm of Tab. 2.3. The algorithm performs flattened texture acquisition by resampling the original image starting from an orthogonal grid of θ and z values in the reference plane. The image to reference plane transformation maps imaged SOR meridians and parallels onto mutually orthogonal straight lines (see also Fig. 2.3(b)). It is worth noticing that, to guarantee that texture details have the same size in all the warped images, a unique scaling factor for z must be specified for all the views. In order to achieve this, the portions of apparent contour used to warp each image must be chosen so that they are delimited by two imaged cross sections corresponding to the same 3D SOR parallels in all views. If the SOR has a top and a bottom, these two extremities can be conveniently selected to delimit the apparent contour in each view.

2.2.3 Texture Alignment

Thanks to the characteristics of the image warping algorithm just described, the subsequent image alignment phase is greatly simplified, and reduced to the problem of estimating rigid translations in the reference plane. The alignment procedure is very similar to that used for cylindrical panoramic mosaics [61]. Both a horizontal translation δ_θ and a vertical translation δ_z are estimated for each input image. Translations along z must be taken into account to compensate for misalignments due to slight uncertainties in the scaling factor.

Direct registration is employed to align at subpixel accuracy the warped images and recover the translation $\delta = (\delta_\theta, \delta_z)$. The intensity error $E(\delta) = \sum_{x_i} [I_1(x_i + \delta) - I_2(x_i)]^2$ between the two images I_1 and I_2 is minimized using the iterative method described in [66]. The algorithm starts from an initial guess lying close to the minimum.

2.3 Experimental Results

Fig. 2.4(a) shows an uncalibrated view of a vase taken from the web. In the figure, two ellipses were manually fitted by following two boundaries corresponding to imaged cross sections. The apparent contour was manually drawn and modeled by an interpolating cubic spline curve. The same figure shows six imaged meridians computed with the described algorithm at increasing angles of 10° measured from the imaged axis of revolution. The imaged meridians shown are part of the resampling grid. The imaged axis of revolution is also the image of a meridian, and specifically the one contained in the plane through the camera center and the 3D axis of revolution. It is worth noticing how the meridian gives the perception of the depth while

Figure 2.4: (a) shows the computed imaged meridians at θ=10°, 20°, 30°, 40°, 50°, 60° measured from the imaged axis of revolution. (b) shows the flattened texture obtained with the described algorithms. Here imaged cross sections and meridians are warped as mutually orthogonal straight lines. (c) Four complementary views of a japanese vase and the warped flattened textures obtained.

approaching the apparent contour. The regular sampling of the spline parameter clearly does not induce a regular sampling for the imaged meridians.

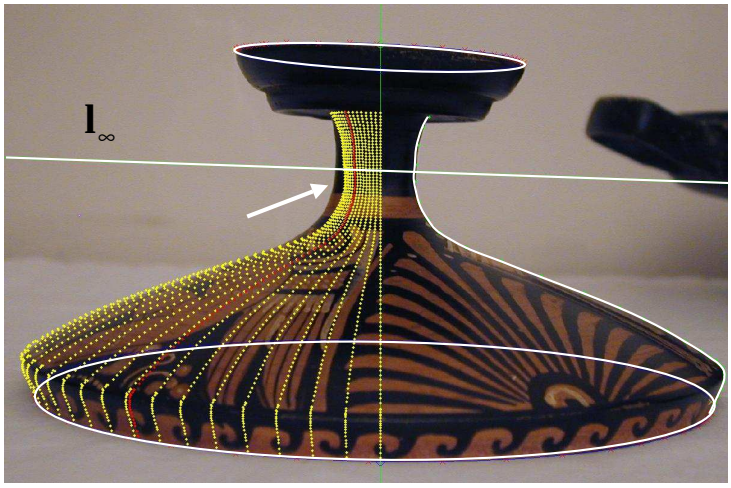
The flattened texture of Fig.2.4(b) is obtained by resampling the curves of the imaged SOR parameterization and warping them onto mutually orthogonal straight lines in the reference texture image (row and column). The largest amount of warping is required by the imaged surface regions close to the apparent contour and by high curvature surface parts. Fig. 2.4(c) shows four views of the vase of Fig. 2.4(a) having a common overlapping imaged surface pictorial content. Fig. 2.4(c) shows also the warped imaged surface in overlapping order. The resolution of the original images is 400×600. We chose a similar resolution for the warped texture images. Fig. 2.5(a) shows the manual initialization of the alignment step with three of the four warped images of Fig. 2.4. Notice how length ratios are maintained in all images, while the lighting is remarkably different from image to image. Fig. 2.5(b) shows the mosaic image resulting after image alignment and compositing. The mosaic represents a full 360° "vase panorama" obtained using five warped images, of which the leftmost image of Fig. 2.4(c) was used twice in order to close the visual texture loop. The effect of image compositing is to reduce the lighting gradient inside the mosaic. The registration fails when the SOR deviates from its ideal geometry. Fig. 2.6(a) shows a vase cover in which the axis of revolution is not straight. Since the top cross section is not perfectly coplanar with the bottom one, the generated imaged parameterization is not perfectly registered. The subsequent image alignment cannot be performed under this condition, since the common overlapping regions are different.

The transformations relating the different coordinate system of each imaged SOR region depends on camera calibration. Bad estimates of internal camera parameters can prevent the alignment with the translational motion

(a)

(b)

Figure 2.5: (a) Initial guess for the direct registration. (b) The complete mosaic obtained by image registration and compositing.



(a)

Figure 2.6: (a): A vase cover slightly deviating from an ideal SOR. The top cross section is not coplanar with respect to the bottom one. As indicated by the arrow the superimposed parameterization is not perfectly registered.

model. One of the main limitations of this approach is that the quality of the boundary fitting of the ellipses (imaged visible cross sections) strongly affects the accuracy of calibration results. Fig. 1.8, 1.9 and 1.10 shows the mean and the relative RMS errors in the computation of the principal point (upper) and focal length (lower) for different noise levels corrupting the imaged cross sections of a synthetic SOR view.

2.4 Summary

In this chapter we have discussed a novel method for mosaicing several uncalibrated views of a SOR. The proposed solution exploits the projective properties of SORs class and their relationships with camera geometry. The method uses as inputs two at least partially visible imaged cross sections and

the apparent contour (see [67] for an automatic method to extract simultaneously such curves and the imaged geometry).

The method gives good results especially with smooth SORs, and can be used reliably in all those cases in which uncalibrated photographs are available and structured light or other hardware solutions cannot be employed. In particular, the method can be applied for cultural heritage or archaeological objects that are either no more available as original, or cannot be moved from their site. It can also be applied in those cases where the nature of the object material makes it impossible or expensive the acquisition with laser-based techniques [64]. The flattened (rolled-out) representation can be regarded as a virtual painting drawn by the artist onto a curved support. Rolling out this kind of images facilitates the study and comparison of similar images. Specifically, existing image retrieval techniques can be applied for indexing databases of 3D objects by their pictorial surface content.

The main limitations of the method are related to self-occlusions and non smooth SORs, as only the texture portions corresponding to a differentialble apparent contour can be warped. Currently, images are warped separately, and then registered together: this means that calibration errors in one view can affect the final mosaic quality. To reduce the effect of calibration errors, future research will address performing the warping and alignment steps simultaneously. Further improvements will be the use of multiview calibration and the detection and removal of specular highlights.

Chapter 3

Automatic Reconstruction

In this chapter, we address the problem of the automatic metric reconstruction Surface of Revolution (SOR) from a single uncalibrated view. The apparent contour and the visible portions of the imaged SOR cross sections are extracted and classified. The harmonic homology that models the image projection of the SOR is also estimated. The special care devoted to accuracy and robustness with respect to outliers makes the approach suitable for automatic camera calibration and metric reconstruction from single uncalibrated views of a SOR. Robustness and accuracy are obtained by embedding a graph-based grouping strategy (Euclidean Minimum Spanning Tree) into an Iterative Closest Point framework for projective curve alignment at multiple scales. Classification of SOR curves is achieved through a 2-dof voting scheme based on a pencil of conics novel parametrization. The main contribution of this work is to extend the domain of automatic single view reconstruction from piecewise planar scenes to scenes including curved surfaces, thus allowing to create automatically realistic image models of man-made objects. Experimental results with real images taken from the internet are reported,

and the effectiveness and limitations of the approach are discussed[1].

3.1 Introduction and Related Work

The confluence of projective geometry and computer vision has produced recently impressive results in image based modeling. In particular, self-calibration methods have been developed to support metric 3D reconstruction even from single uncalibrated views. However, single view reconstruction is typically performed in a semi-automatic way, due to the difficulties arising in automatic image segmentation [20] [68] [7]. Automatic segmentation for reconstruction is even more challenging than for recognition, since accurate estimates of geometric features and their relationships are needed in order to get reasonable calibration results. The use of geometric models of the scene can ease the automatic segmentation task, by reducing the hypothesis search space during the process. In [69] [70] [71] [72], a piecewise planar scene model was used to support segmentation for reconstruction.

Models of non planar surfaces such as the Straight Homogeneous Generalized Cylinder (SHGC) were extensively used in the past for the specific problem of segmentation for recognition under affine view conditions (for a review, see [32]). Recent work on this problem is described in [41], where a bottom-up strategy is used to recognize SHGC models from B-spline interpolated image curves. Similar time consuming bottom-up strategies are also exploited in other approaches performing curve segmentation without a specific 3D model [73] [74] [53].

The most recent research in the domain of single view camera calibration and reconstruction has focused on Surfaces Of Revolution (SOR's), which are

[1]For this chapter i would like to acknowledge in particular Dario Comanducci.

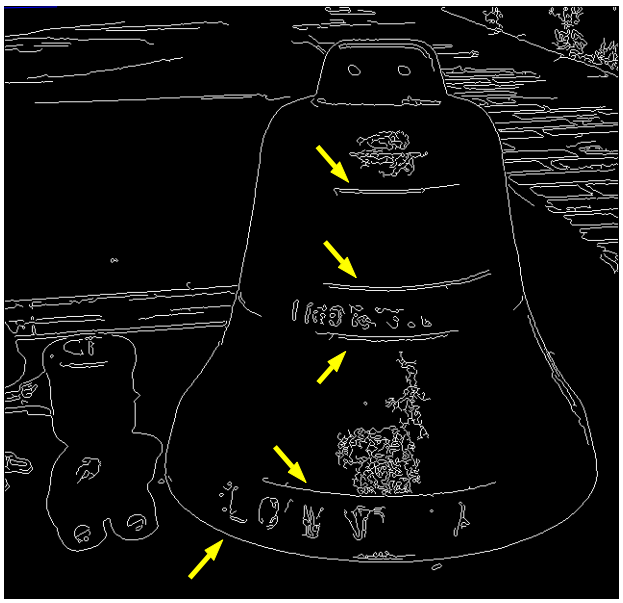
a subclass of SHGC's. In [6] it is shown that a single SOR view can provide two constraints to partially calibrate a pinhole camera. Semi-automatic approaches extending single view planar scene reconstruction to the SOR case were presented in [9] and [4]. In particular, the latter approach also shows how to perform texture acquisition from a single SOR view.

In this chapter, a method for the automatic segmentation of SOR models from a single uncalibrated perspective view is presented. The method is aimed at supporting single view reconstruction, and make it fully automatic. A top-down segmentation strategy is devised, in which a global projective model and the image curves consistent with it are estimated at multiple scales. The estimation interleaves robust curve alignment and graph-based curve grouping. The segmented image curves are then classified into apparent contour and imaged SOR cross sections by means of a voting scheme based on the projective properties of imaged SOR's. The chapter is organized as follows. In the following section, the segmentation problem is stated, and an outline of its solution is provided. In Sect. 3.3, model estimation and curve grouping is described. Curve classification is addressed in Sect. 3.4. In Sect. 3.5 experimental results are discussed; finally, conclusions and directions for further research are presented in Sect. 3.6.

3.2 Problem Statement and Overview of the Approach

The perspective projection of a SOR like the bell of Fig. 3.1(a) gives rise to two different kinds of image curves, namely the *apparent contour* and the *imaged cross sections*. The apparent contour is the image of the points at which the surface is smooth and the projection rays are tangent to the surface.

(a)



(b)

Figure 3.1: (a): original SOR image. (b): the edge image of (a), several cross-section fragments are visible (indicated with arrows); yet, these measurements are practically useless for classic ellipse detection and estimation algorithms.

The shape of this curve is view dependent. On the other hand, imaged cross sections are view independent elliptical curves, which correspond to parallel coaxial circles in 3D and arise from surface normal discontinuities or surface texture content. Both the apparent contour and the imaged cross sections of a SOR are transformed onto themselves by a 4-dof harmonic homology 2.2.1:

$$\mathtt{H} = \mathtt{I} - 2\frac{\mathbf{v}_\infty \mathbf{l}_s^{\mathrm{T}}}{\mathbf{v}_\infty^{\mathrm{T}} \mathbf{l}_s} \ ,$$

where $\mathbf{v}_\infty$ and $\mathbf{l}_s$ are respectively the vanishing point and the imaged axis of revolution of the normal direction of the plane passing through $\mathbf{l}_s$ and the camera center [17].

The problem addressed here is how to estimate automatically from a SOR image the harmonic homology of Eq. 2.2.1 together with the apparent contour and the visible imaged cross sections consistent with it. This geometric information is useful to perform both camera calibration and metric reconstruction of a SOR object based on a single uncalibrated view. This problem could hardly be solved without embedding the full geometric knowledge of the scene into the estimation process. For instance, as shown in Fig. 3.1(b), the visible segments of the imaged cross sections typically correspond to almost straight edgel chains, thus making their extraction highly inaccurate or even impossible with standard ellipse detectors. The approach followed in this work, sketched hereafter, attempts to exploit as much as possible the prior knowledge of the scene in order to tightly constrain the estimation process.

The overall approach can be divided into two phases: (1) estimation of the harmonic homology of Eq. 2.2.1 and of all image curves consistent with it (see Sect. 3.3); (2) classification of the image curves (see Sect. 3.4). The first phase uses an Iterative Closest Point (ICP) strategy [75][76] to align

sets of edgels related by the harmonic homology as the result of a nonlinear optimation problem. (Edgels are obtained by Canny edge detection.) The main advantage of using ICP is that it allows to represent curves globally as noisy point sets, thus avoiding any intermediate representation by local descriptors, which are typically cumbersome to extract and unreliable. In order to cope with a large fraction of outliers – i.e., edgels not belonging to the SOR –, the approach makes an extensive use of robust regression methods, such as the Least Median of Squares (LMedS) [77]. Outlier rejection also benefits from the use of a general curve grouping scheme based on continuity properties. Moreover, in order to avoid to get stuck in local minima during the nonlinear minimization, ICP is run at multiple scales and applied to subsequent levels of a Gaussian pyramid. ICP algorithms generally need an initial alignment in order to converge. At the beginning of the first phase, the RANdom SAmple Consensus (RANSAC) algorithm is run at the coarsest pyramid level so as to provide ICP with a reliable first guess solution. The second phase is devoted to classifying the image curves obtained in the first phase respectively into (a) apparent contour, (b) imaged cross sections and (c) clutter. To this end, ellipses are searched for over the image as particular instances of a conic pencils spanning the apparent contour.

3.3 Homology Estimation and Curve Grouping

3.3.1 RANSAC-based initialization.

To compute a weak estimate of the harmonic homology and use it as a first guess for the whole estimation process, bitangents to imaged SOR curves

could be used as shown in [46] [6]. However, bitangent estimation generally requires that the Canny edgels be interpolated by polynomials. To avoid this heavy computational step, an alternative way is to estimate the homology directly from the edgels. This is achieved by running RANSAC at the coarsest level of the pyramid. RANSAC is a random sampling procedure which is known to be slow in the presence of a large number of outliers and/or model parameters. In fact, the number N of RANSAC trials which guarantees the statistical convergence depends on the percentage of outlier to tolerate: $N = \frac{\log(1-p)}{\log(1-(1-\epsilon)^s)}$, where ϵ is the fraction of outliers present in the data and s is the number of points of the sample set (p is a probability value generally set to 0.99). If the sample set and the outlier percentage are low, a fast convergence is obtained. This is what happens at the coarsest level of the pyramid, since (*i*) due to low pass filtering, the number of edgels – and therefore, of possible outliers – is small; (*ii*) the harmonic homology reduces to Euclidean axial symmetry, which can be described by the 2-dof axis $\mathbf{l}_s$ only – the vanishing point $\mathbf{v}_\infty$ being the point at infinity in the direction orthogonal to the axis. Fig. 3.2(a) shows the estimated axial symmetry for the bell of Fig. 3.1.

3.3.2 Robust homology estimation.

At each level of the pyramid, the harmonic homology is estimated starting from the results obtained at the immediately coarser level by directly minimizing the registration error

$$\mathcal{F}(\mathbf{l}_s, \mathbf{v}_\infty) = \sum_i \|\mathbf{x}'_i - \mathtt{H}(\mathbf{l}_s, \mathbf{v}_\infty)\mathbf{x}_i\|^2 + \sum_i \|\mathbf{x}_i - \mathtt{H}^{-1}(\mathbf{l}_s, \mathbf{v}_\infty)\mathbf{x}'_i\|^2$$

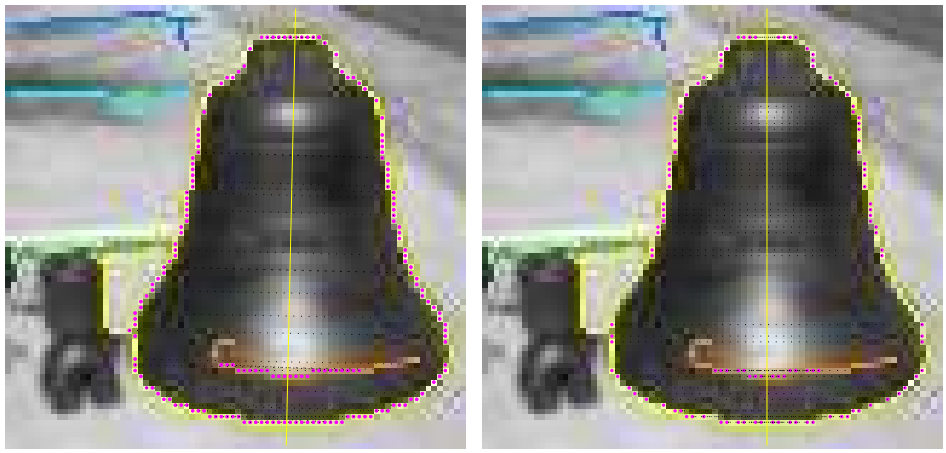
using nonlinear optimization. In Eq. 3.1, $\mathbf{x}_i$ and $\mathbf{x}'_i$ are edgel points corresponding under $\mathtt{H}$. In particular, since $\mathtt{H}^{-1} = \mathtt{H}$, $\mathcal{F}$ is a symmetric transfer

error – a distance measure with a remarkable geometric meaning. To reject outliers and also to improve convergence, the LMedS is used. Its effect is to reduce, by selecting them, the points involved in Eq. 3.1. Once the histogram of symmetric transfer distances is computed, only the points inside six times the standard deviation from the mode are retained.

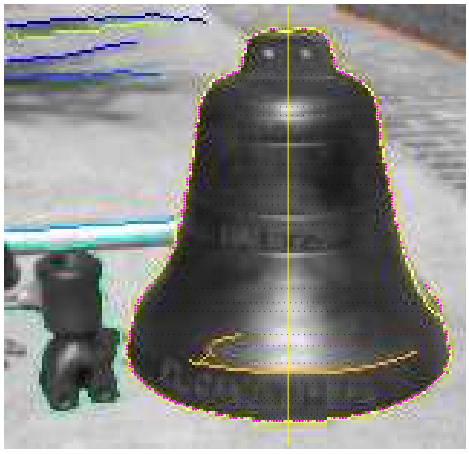
Figs. 3.2(b) and 3.2(c) show the estimation results obtained at the first and second pyramid levels, respectively. In particular, the result of Fig. 3.2(b) is an improvement of the RANSAC result shown in Fig. 3.2(a), and represents the initial guess for the result of Fig. 3.2(c). Note that although new edgels arise at each finer resolution level, the homology estimate (whose axis is shown in the figures) remains locked to the dominant SOR.

3.3.3 Curve Grouping.

Despite the fact that the multiresolution approach together with LMedS heavily contributes to regularize the error surface, the non linear minimization strategy above can get stuck in local minima in the presence of a huge number of outliers. This is typically the case when the background clutter and/or the SOR texture accidentally exhibit symmetric patterns which act as distractors. In order to cope with this situation, further prior knowledge about the geometric characteristics of imaged SOR curves is exploited. Indeed, only those image curves which (1) are long and dense, and (2) have tangents which also correspond under the homology should contribute to the minimization. To check the above criteria, curves have to be constructed from edgels. An efficient way to do this is to compute the Euclidean Minimum Spanning Tree (EMST) for the whole set of edgels. This is obtained by running the Kruskal algorithm on the Delaunay triangulation graph computed over the whole edgel set [78]. The assumption that the edgels lie on



(a) (b)



(c)

Figure 3.2: ICP homology RANSAC initialization (a) and estimation at the end of the 1st (b) and 2nd (c) levels of the pyramid. Each level uses as initial guess the results of the previous one.

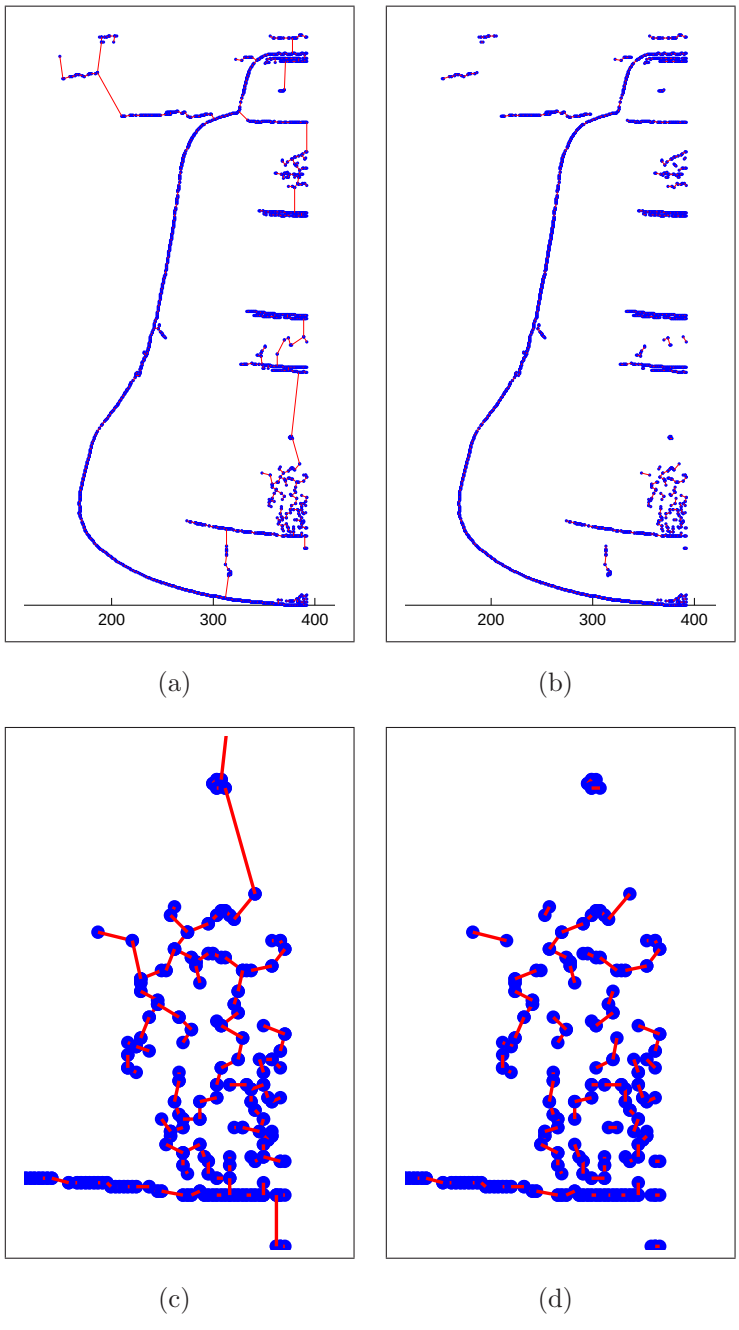
regular curves allows for the removal of inconsistent arcs from the EMST, whose remaining connected components are the required image curves. Arc removal is achieved by performing simple topological testing. Indeed, since each curve is regular, multiple branching from an edgel is not allowed. Therefore, all arcs but two at any branching point of the EMST must be removed. In particular, to meet the denseness requirement, the longest arcs are removed. The gradient direction information associated to each edgel is also exploited to check the local tangency requirement. In particular, given the point transformation $\mathtt{H}$, tangent lines must correspond under $\mathtt{H}^{-\mathrm{T}}$. Edgels whose tangent lines do not correspond are not put into correspondence in the ICP procedure.

Fig. 3.3(a) shows the EMST for the last level of the pyramid (bell image). Fig. 3.3(b) shows the result of arc removal at branching points. Fig. 3.4(a) shows the remaining connected components, i.e. the resulting image curves. The figures show the results for one of the two half planes generated by $\mathbf{l}_s$. Indeed, due to symmetry under the homology, the edgel set belonging to any half plane is sufficient for curve grouping purposes.

After arc removal, the obtained image curves $\mathcal{C}_k, k = 1, \ldots K$ are scored by means of the following saliency measure:

$$S(\mathcal{C}_k) = \frac{M}{\frac{1}{M-1}\sum_{j=1}^{M-1} d(\mathbf{x}_i, \mathbf{x}_{i+1})} \ , \tag{3.1}$$

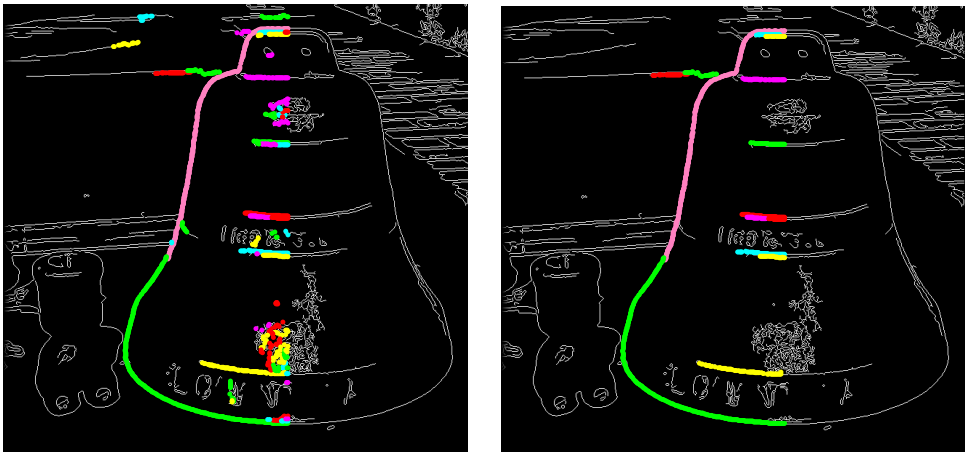
where M is the number of edgels of the curve and $d(\mathbf{x}_i, \mathbf{x}_{i+1})$ is the Euclidean distance between subsequent edgels. This measure assigns high scores to dense, long curves and low scores to short, fragmented ones. This allows for using LMedS again to keep only the curves which lie outside the mode. Fig. 3.4(b) shows the curves of Fig. 3.4(a) after saliency thresholding.



(a) (b)

(c) (d)

Figure 3.3: (a): the EMST of the original image (finer pyramid level). (b): arc removal from branching points. (c)(d) A zoomed view of the previous figures.



(a) (b)

Figure 3.4: (a): (a): the curves resulting from arc removal. (b): the curves resulting after saliency thresholding.

3.3.4 The algorithm

The homology estimation and curve grouping algorithm is summarized as:

0. Compute the Gaussian pyramid $\{G_l\}, l = 0 \ldots L-1$ and extract the edgels at all levels.

1. Set $l = 0$ (coarsest level). Compute $\mathtt{H}^{(l)}$ from the edgels of G_0 with RANSAC.

2. Repeat until convergence:

   – Transform each edgel $\mathbf{x}_i$ with $\mathtt{H}^{(l)}$ and compute the edgel closest to $\mathtt{H}^{(l)}\mathbf{x}_i$.

   – Discard all edgel pairs whose tangent lines do not correspond under $[\mathtt{H}^{(l)}]^{-\mathrm{T}}$.

   – Compute the inlier set $\{\mathbf{x}_i\}$ from the histogram of distances with LMedS.

   – Compute the EMST from the points of $\{\mathbf{x}_i\}$.

- Remove all but two arcs from the EMST branching points and obtain the set of curves $\{\mathcal{C}_k\}$.
- Compute the histogram of the saliency and the inliers from $\{\mathcal{C}_k\}$ with LMedS.
- Update the homology as $\mathtt{H}^{(l)} \leftarrow \arg\min_{\mathtt{H}^{(l)}} \mathcal{F}(\mathbf{l}_s, \mathbf{v}_\infty)$ by nonlinear optimization.

3. If $l < L$, set $l \leftarrow l + 1$ and go to 2.

3.4 Curve Classification

In this section a novel method is presented to classify curves extracted from a SOR image into three classes: apparent contour, imaged cross section (ellipse), clutter. The third class includes all the curves that are not part of the dominant SOR object in the scene. The classification strategy is based on the following result (see also Fig. 3.5):

Given two curves γ and γ' and two points on them, $\mathbf{x} \in \gamma$ and $\mathbf{x}' \in \gamma'$, which correspond under $\mathtt{H}$, then all the possible imaged SOR cross sections through $\mathbf{x}$ and $\mathbf{x}'$ are described by the 1-dof pencil of conics

$$\mathtt{C}(\lambda) = \mathbf{m}\mathbf{m}^{\mathrm{T}} + \lambda(\mathbf{l}\mathbf{l}'^{\mathrm{T}} + \mathbf{l}'\mathbf{l}^{\mathrm{T}}) \ , \tag{3.2}$$

where $\mathbf{m}\mathbf{m}^{\mathrm{T}}$ is the (rank 1) degenerate conic composed by the line $\mathbf{m} = \mathbf{x}' \times \mathbf{x}$ through $\mathbf{x}'$ and $\mathbf{x}$, and $\mathbf{l}\mathbf{l}'^{\mathrm{T}} + \mathbf{l}'\mathbf{l}^{\mathrm{T}}$ is the (rank 2) degenerate conic composed by the line pair $\mathbf{l}$ and $\mathbf{l}'$ tangent to γ and γ' (the two symmetric side of the apparent contour) respectively at $\mathbf{x}$ and $\mathbf{x}'$.

To prove this result, it is sufficient to recall that the apparent contour is tangent to an imaged SOR cross section at any point of contact [32]. This means that, at the point of contact, the apparent contour and the ellipse

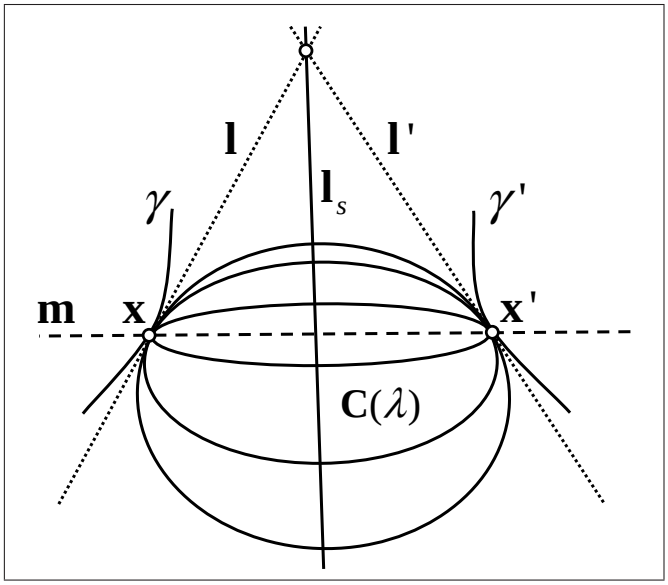


Figure 3.5: The pencil of conics at the tangent contact points $\mathbf{x}$ and $\mathbf{x}'$ of the apparent contour of a SOR. Three members of the pencil are reported for three distinct values of λ.

have the same tangent line. Fig. 3.5 shows two symmetric portions γ and γ' of the apparent contour, corresponding under the harmonic homology with axis $\mathbf{l}_s$. If the tangent lines $\mathbf{x} \in \gamma$ and $\mathbf{x}' \in \gamma'$ are $\mathbf{l}$ and $\mathbf{l}'$ respectively, then all the possible imaged cross sections $\mathtt{C}$ must include $\mathbf{x}$ and $\mathbf{x}'$ and have there as tangent lines $\mathbf{l}$ and $\mathbf{l}'$ respectively. From the point inclusion constraints $\mathbf{x}^{\mathrm{T}}\mathtt{C}\mathbf{x} = 0$ and $\mathbf{x}'^{\mathrm{T}}\mathtt{C}\mathbf{x}' = 0$ and the tangency constraints $\mathtt{C}\mathbf{x}' \times \mathbf{l}' = \mathbf{0}$ and $\mathtt{C}\mathbf{x} \times \mathbf{l} = \mathbf{0}$ the result follows immediately.

It is also straightforward to show that there exists a closed form solution for the member of the pencil passing through any assigned point $\mathbf{p}$, i.e. such that $\mathbf{p}^{\mathrm{T}}\mathtt{C}(\lambda_{\mathtt{p}})\mathbf{p} = 0$:

$$\lambda_{\mathtt{p}} = -\frac{\mathbf{p}^{\mathrm{T}}(\mathbf{m}\mathbf{m}^{\mathrm{T}})\mathbf{p}}{\mathbf{p}^{\mathrm{T}}(\mathbf{l}\mathbf{l}'^{\mathrm{T}} + \mathbf{l}'\mathbf{l}^{\mathrm{T}})\mathbf{p}} \quad . \tag{3.3}$$

This last result can be used to formulate the curve classification algorithm as a voting procedure similar to the Hough transform. The algorithm is as

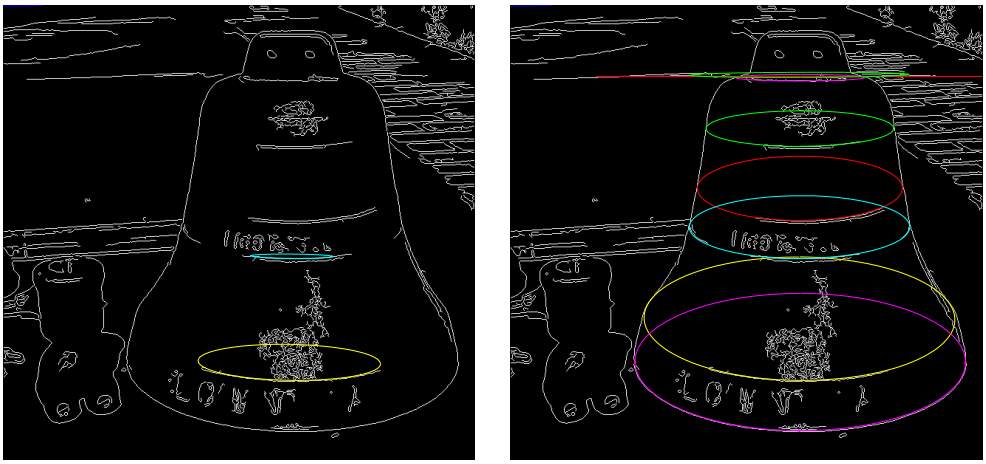
follows.

0. Repeat steps 1–3 for all the ordered pairs $(\mathcal{C}_i,\mathcal{C}_j)$.

1. Consider $\mathcal{C}_i$ from the curve set $\{\mathcal{C}_k\}$, assume that it belongs to the apparent contour, and parametrize it as $\mathbf{x}_m$, with m an integer spanning all the edgels of $\mathcal{C}_i$. For each value of m, a pencil $\mathtt{C}_m(\lambda)$ is obtained.

2. Consider a second curve $\mathcal{C}_j$, and parametrize it as $\mathbf{x}_n$. For all values of n, use Eq. 3.3 to find the value λ_n such that $\mathbf{x}_n^{\mathrm{T}}\mathtt{C}_m(\lambda_n)\mathbf{x}_n = 0$.

3. Let $v(m,n)$ be the fraction of points of $\mathcal{C}_j$ whose distance from $\mathtt{C}_m(\lambda_n)$ is below a predefined threshold δ (we use 1 pixel). If the peak value $v(m^\star,n^\star)$ of the matrix $\{v(m,n),\ \ \forall m\forall n\}$ is above 0.9, then classify $\mathcal{C}_i$ as a portion of the apparent contour, and $\mathcal{C}_j$ as a portion of the imaged cross section $\mathtt{C}_{m^\star}(\lambda_{n^\star})$.



(a) (b)

Figure 3.6: (a): Two ellipses fitted with the algorithm described in [79]. (b): The results of the proposed ellipse classification algorithm.

3.5 Experimental Results

In order to test the algorithm for SOR detection and homology estimation discussed in Sect. 3.3, a set of 37 SOR images was considered. In order to have a ground truth for experiments, for each image of the set the dominant SOR was identified, and the associated harmonic homology was computed from hand drawn apparent contours. The dominant SOR was correctly detected with the proposed algorithm in the 93% of the cases. Moreover, the average departure of the harmonic homology axis computed automatically from the ground truth axis is 0.6 degrees. The main source of error for the SOR detection algorithm is the presence of planar scene structures exhibiting a high degree of symmetry, giving rise to harmonic homologies which do not actually correspond to a SOR.

Concerning curve classification, results show that the algorithm of Sect. 3.4 is always correct in finding the SOR apparent contour, and also provides ellipse estimates which are far more reliable than the ones obtained with classical ellipse estimation algorithms. Fig. 3.6 shows the results of ellipse estimation obtained respectively with the algorithm described in [79] on two manually selected curves (Fig. 3.6(a)), and with the algorithm proposed in this chapter (Fig. 3.6(b)).

Curve classification takes about two minutes per image on a 1.5Ghz computer. Generally, most of the computation is spent on the last pyramid level (the original image), where a huge number of edgels is extracted. Timings are measured with images of medium size (640×480 or 800×600) and 4 or 5 pyramid levels.

The algorithms proposed here were used to support the method for camera calibration and metric reconstruction from single uncalibrated SOR images described in [4]. All images were taken from the internet. Results are

shown in Figs. 3.7, 3.8 and 3.9. Fig. 3.7(a) and 3.7(b) show the effectiveness of results for the metric reconstruction of the bell of Fig. 3.1. Estimation results are good enough to calibrate the camera and perform both metric reconstruction and texture acquisition as described in [4]. Fig. 3.7(c) shows the acquired texture, in which imaged cross sections and SOR meridians become orthogonal straight lines. Fig. 3.7(d) shows the image rectification of the plane the bell is laid. The detected ellipses furnish also the homography which reports to metric the planar geometry of the SOR ground plane. An example with a vase image is also presented. Fig. 3.8(c) shows the original image with the superimposed interpolated curves. Figs. 3.8(a) and (b) show the non interpolated curves before and after saliency thresholding. Finally, Fig. 3.8(d) reports the reconstructed 3D model. Fig. 3.9(a) is an image showing the Taj Mahal. Fig. 3.9(b) shows a synthetic view of the reconstructed dome of the Taj Mahal.

The algorithm may fail when the shadows is present nearly the apparent contour as shown in Fig.3.10(a). This problem arise since the detected edges of the two curves (the apparent contour and the shadow boundary) are very close to each other. This produce a slightly deviating symmetry which can be erroneously detected by the RANSAC initialization as shown in Fig.3.10(b). This also prevent the convergence of the nonlinear minimization to a global minimum (Fig.3.10(c)).

3.6 Summary

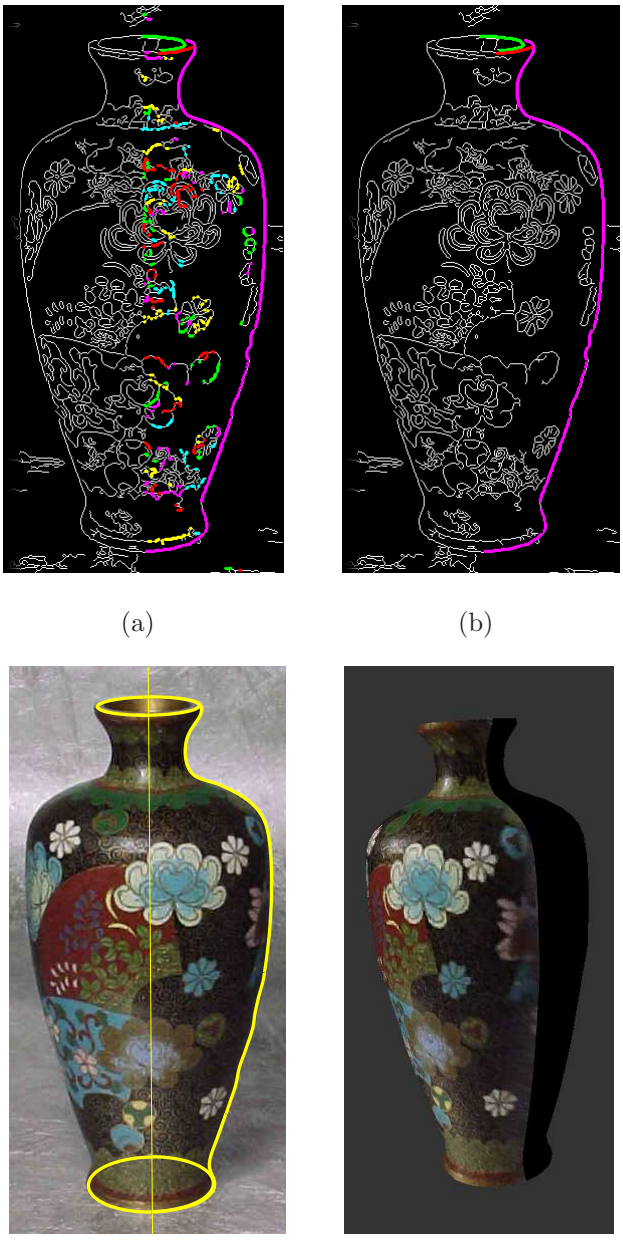
An original approach to automatically grouping and estimation of the projective geometry of single SOR views was presented. The approach is mainly devoted to camera autocalibration from a single SOR view and single view

(a) (b)

(c)

(d)

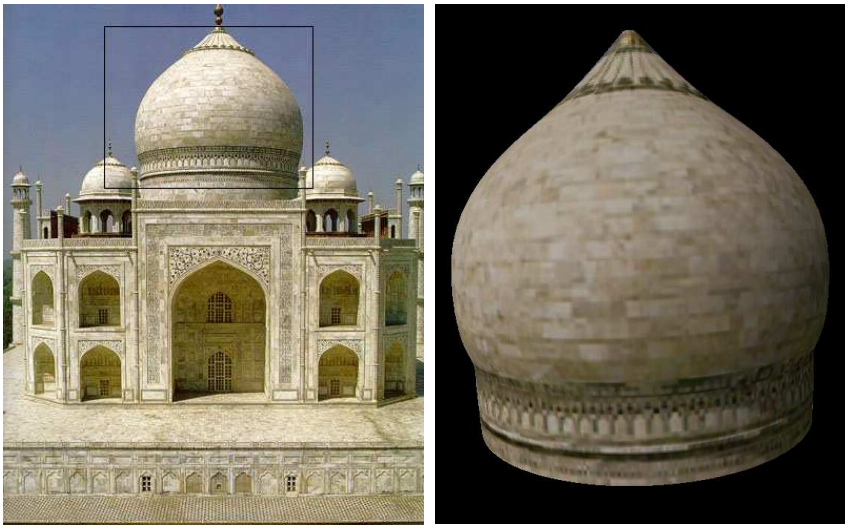
Figure 3.7: Applications. (a),(b): two synthetic views of the reconstructed bell of Fig. 3.1. (c): metric flattened texture of the bell surface (the imaged cross-section looks straight). (d): metric rectification of the floor inferred from the SOR.



(a) (b)

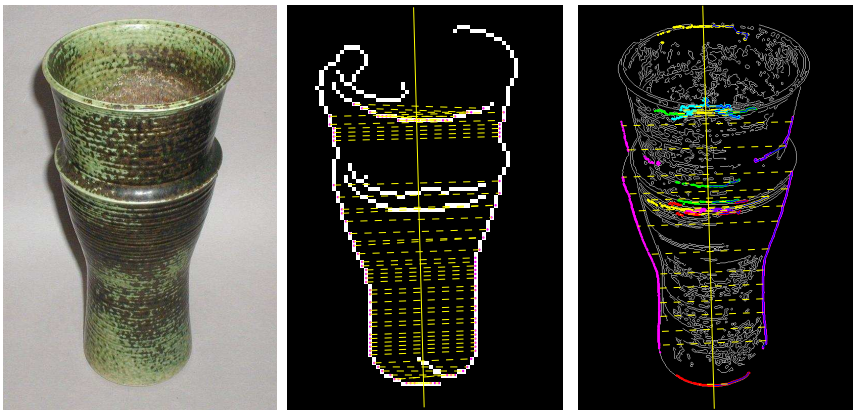
(c) (d)

Figure 3.8: (a), (b): Curve grouping before and after saliency thresholding for the vase shown in (c). (c): the original vase with superimposed the spline curve and the ellipses used for metric reconstruction. (d): A view for the reconstructed vase.



(a) (b)

Figure 3.9: (a): The Taj Mahal. (b): The reconstructed dome.



(a) (b) (c)

Figure 3.10: (a): A SOR view in which the algorithm may fail. (b): The RANSAC initialization at the coarsest level. The shadow boundary edges are erroneously classified as symmetric. The dotted lines indicates corresponding points. (c): The final result. The nonlinear minimization remain trapped in a local minima.

metric reconstruction of SOR objects.

The main limitation of the approach is related to the possibility of misdetections due to symmetric line patterns acting as distractors. To overcome this limitation, further research will be devoted to combining in a single general framework the SOR and planar cases. Another limitation is that curve differentiability is required in order to use Eq. 3.2 for the purpose of ellipse detection. This requirement prevents from estimating all ellipses corresponding to surface normal discontinuities.

Part II

Video Surveillance and Active Zooming Camera

Chapter 4

Saccade planning with PTZ camera

This chapter considers the problem of designing an active observer to plan a sequence of decisions regarding what target to look at, through a foveal-sensing action. We propose a framework in which a pan/tilt/zoom (PTZ) camera schedules saccades in order to acquire high resolution images (at least one) of as many moving targets as possible before they leave the scene. An intelligent choice of the order of sensing the targets can significantly reduce the total dead-time wasted by the active camera and, consequently, its cycle time. The grabbed images provide meaningful identification imagery of distant targets which are not recognizable in a wide angle view. We cast the whole problem as a particular kind of dynamic discrete optimization. In particular, we will show that the problem can be solved by modelling the attentional gaze control as a novel on-line Dynamic Vehicle Routing Problem (DVRP) with deadlines. Moreover we also show how multi-view geometry can be used for evaluating the cost of high resolution image sensing with a PTZ camera.

Congestion analysis experiments are reported proving the effectiveness of the solution in acquiring high resolution images of a large number of moving targets in a wide area. The evaluation was conducted with a simulation using a dual camera system in a master-slave configuration. Camera performances are also empirically tested in order to validate how the manufacturer's specification deviates from our model using an off-the-shelf PTZ camera.

4.1 Introduction

Equipping machines with even a limited version of our own visual abilities is proving a remarkable task. The image formation process of the eye alone can be emulated using a lens and CCD array; it is argued as a result of the 3D to 2D mapping of the real world to the visual retina that the major challenge for for machine vision is perception.

Making the biological comparison, it is widely reported that the human visual faculty is not a single comprehensive processing unit, but a series of small task-specific processors whose output can be combined. Image stabilization (fixation) occurs at a level below the brains main visual processors, having direct connections from the ear and retina to the eye muscles. Eye and head movements interact with the visual process, allowing maximum resolution to be focused on specific areas of the scene. In humans this is achieved by either repeated saccade-fixate cycles or by smooth motion tracking. So in reality, we do not scan a scene in raster fashion, our visual attention tends to jump from one point to another. These jumps are called saccade. Yarbus [80] demonstrated that the saccadic patterns depend on the visual scene as well as the cognitive task to be performed. The conclusion is that we do not see, we look [81]. In this chapter the focus is visual attention according to

task at hand and the scene content.

The lack of works addressing task-driven visual processing is mainly motivated by the fact that its studying seems, as a first sight, too specialized, non-generic, or bordering on hackery. But active vision demands such processes; it is founded in the idea of specialized processing for specialized tasks. Most of the active vision literature is limited to studying low-level subconscious reflexes. One wonders whether truly active and purposeful vision systems will be realized. In other words, while active tracking and visual attention was researched in the past years, purposeful zooming is (and probably will remain) a largely unexplored area in active vision [82]. Basically sensing was not a major issue for computer vision as for example was perception. However despite this for the particular task of object recognition notably works are reported in the literature [83] [84].

dire qui che active recognition is particular Our work is motivated by the goal of reproducing the ability of humans to recognize a person in a crowd of moving people for surveillance purposes. In humans, the process of recognizing a person and that of moving the eyes are served by almost two distinct subcortical brain areas: one specialized for recognizing faces and one specialized for making decisions on whom look at next. The eye acts as a foveal sensor that allows high resolution only at the point of interest, avoiding the cost of uniform high resolution. Indeed during a scan-path in a moving crowd of walking people it is normal to backtrack to a previous observed person thinking "oh that's my friend". This because the gaze planning task does not directly depend on the face recognition task. Visual attention in this particular task is more affected by the target position, the predicted time in exiting the scene and the effort made in moving the head and the eyes from one direction to another. In fact during a saccade, the redirection is so

rapid that the gaze lasts only a tenth of a millisecond. During that time the few images obtained are typically blurred because of the fast camera motion. As far as the deployment in sophistication in visual analysis is concerned, saccades are dead times. So our brain avoids doing large redirection of the gaze while undertaking this task, trying to minimize that dead time.

A direct application of that behavior of the human visual system can be applied in Visual Surveillance. Automated surveillance can be a powerful tool in deterrence of crime, but most of the solutions and implementations proposed so far are unnecessarily poor in evidential quality. In this sense, remote identification of targets is and will be an important mandatory capability for modern automated surveillance systems. In particular, recognizing a person or a car license plate requires that high resolution views must be taken before they leave the scene. Using a large number of static or active cameras that operate cooperatively is an expensive and impractical solution. One way to cope with this problem is to make better use of the capabilities of the sensor.

We argue that one active pan/tilt/zoom (i.e. a foveal sensor) camera (the slave camera) together with a wide angle camera (the master camera) and a good strategy for visiting the targets can be used instead. The fixed camera is used to monitor the scene estimating where targets are in the surveilled area. The active camera then follows each target to produce high resolution images. In this configuration, we show that the visual signal from the master camera provides the necessary information to plan the saccades sequence. Moreover, the introduction of an appropriate scheduling policy allows to maximize the number of targets that can be identified from the high resolution images collected. Indeed, this is achieved by continuously gazing at the most appropriate targets, where the appropriateness strongly depends

on the task considered. In fact, tasks may have conflicting requirements, as in the case where different tasks would direct the fovea to a different point in the scene. For systems with multiple behaviors, this scheduling problem becomes increasingly paramount.

The key contributions of this part of the thesis are: (1) We propose a novel formulation for the remote target identification problem in terms of saccadic gaze planning. (2) We give a general framework in which an active camera can be modelled. (3) The use of uncalibrated methods makes the proposed framework function in any planar scene. (4) We extend previous approaches on PTZ greedy scheduling proving through simulation that our framework yields better system performance.

4.2 Related Work

Recent years (especially after 9/11) have seen a continued increase in the need for and use of automatic video surveillance for remote identification problems. The few works addressing this subject do not address the planning problem or do not fully exploit all the information intrinsically present in the structure of the problem. In [85] the problem of deciding which camera should be assigned to which person was addressed and some general approaches are given. It should also be noted that there is no work except [86] on objectively evaluating the performance of multi-camera systems for acquiring high resolution imagery of people. Most results are presented in the form of video examples or a series of screen captures without explicit system performance evaluations. Very little attention is given to the problem of what to do when there are more people in the scene than active cameras available.

Many works in literature uses a master/slave camera system configura-

tion with two [87][86][88][89][90] or more cameras [91][92][93][85][94]. The remote target identification problem is also termed as distant human identification (DHID). In [87], a single person is tracked by the active camera. If multiple people are present in the scene, the person who is closest to the position of the previous tracked individual is chosen. In [86] the authors use greedy scheduling policies taken from the network packet scheduling literature. They are the first to describe the problem formally and propose a solution. In particular, in this work the authors, albeit mentioning that there is a transition cost measured in time to be paid whenever the camera switches from person to person, do not explicitly model this cost in their problem formulation. The consequence is that their analysis wrongly motivates an empirically determined watching time instead of at least a single video frame. Moreover the work uses greedy policies instead of policies with a time horizon. Also in [91] the authors propose a form of collective camera scheduling to solve surveillance tasks such as acquisition of multi-scale images of a moving target. They take into account the camera latency and model the problem as a graph weighted matching. In the paper there are no experimental results and no performance evaluation for the task of acquiring as many multi-scale images of many targets as possible in real time. In [89] another similar approach with a dual camera system was recently proposed in indoor scenes with walking people. No target scheduling was performed, targets are repeatedly zoomed to acquire facial images by a supervised learning approach driven by skin, motion and foreground features detection. In [95] a ceiling mounted panoramic camera provides wide-field plan-view sensing and a narrow-field pan/tilt/zoom camera at head height provides high-resolution facial images. The works in [96][88] concentrate on active tracking. In both works the respective authors propose a simple behavior (a policy) with a

finite state machine in order to give some form of continuity when the currently tracked target is changed. In [92] two calibration methods to steer a PTZ camera to follow targets tracked by another camera are proposed. The authors give some criteria of optimization leaving the formal optimization as future research. Though performing coarse registration the methods [92] and [87], generally suffice to bring the target object within a narrow zoomed field of view.

Another body of literature, concerning the mathematical optimization framework, comes from the motion planning literature and in particular from the context of rapid deployment automation. Specifically, those problems related to rearranging parts by a robot in an industrial assembly line setting. A representative work in this context is [97]. In that work the problem is: given n identical parts initially located on a conveyer belt, and a robot arm of capacity k parts, compute the shortest route for the robot arm to grasp and deliver the parts, handling at most at k a time. A PTZ-camera can be interpreted as a robot arm, we will use such analogy in our problem formulation.

The other important work related to our problem is [98], in which the authors study the problem in which a vehicle moves from point to point (customers) in a metric space with constant speed, and at any moment a request for service can arrive at a point in the space. The objective is to maximize the number of served customers. They analyze several policies showing that in such a problem lower bounds on system performance can be obtained analytically. This work is reminiscent of our problem, the main differences are that our customers (targets) are moving and have deadlines. A further important difference is that the nature of our particular vehicle (a PTZ-camera) does not allow us to model the cost of moving from target to

target in the euclidean space.

4.3 Problem Formulation

In this section we formulate and discuss the three main features that characterize this problem: targets motion, arrivals as a continuous process, and deadlines. Once a subset of moving target is selected the correct camera tour can be optimized as a Kinetic Travelling Salesperson Problem (KTSP). The problem of how choosing the best permutation subset from the currently tracked targets is an instance of the Time Dependent Orienteering (TDO) with deadlines.

4.3.1 Kinetic Travelling Salesperson Problem

As cameras can be calibrated with automatic or manual methods such as in [92] it is possible to associate to each point in the plane where targets are moving a vector of PTZ-camera parameters. According to this, at each point in the world plane it is possible to issue camera commands in order to bring a moving target in a close up view by giving to the camera the 3D vector (p, t, z), specifying pan, tilt and zoom values to be applied. In our formulation we model the PTZ-camera as an intercepter with restricted resources (e.g., limited speed in setting its parameter). The dynamics of the targets are assumed known or predictable (i.e., for each target one can specify its location at any time instant). The problem is expressed as that of finding a policy for the PTZ-camera which allows to "visually hit" (with a saccade sequence) as many targets as possible in accordance with the device speed. This allows to cast the problem as a Kinetic Travelling Salesperson problem (KTSP) [99]. In fig.4.1 are shown four targets A, B, C, D moving on a plane. The shortest-time tour is shown with the respective interception points. At

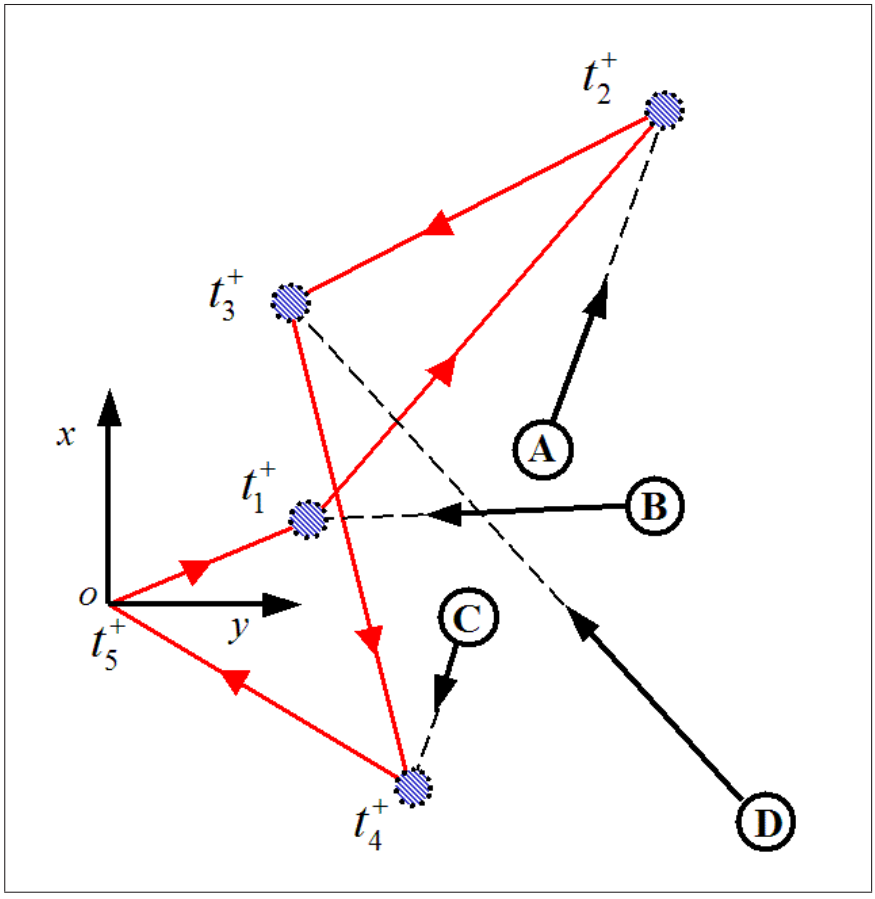


Figure 4.1: An instance of Kinetic-TSP with four targets. The shortest-time tour (light line).

each interception point is also shown the time instants of the sequence when the intercepter visually hits the targets. Formally this problem is formulated as follow:

KTSP : *Given a set* $S = \{s_1, s_2, ..., s_n\}$ *of moving targets, each* s_i *moving with known or predictable motion* $x_i(t)$*, and given an active camera intercepter starting at a given position and having maximum speed* $V_{ptz} \geq V_i \, \forall i$ *, find the shortest-time tour starting (and ending) at the origin, which intercepts all targets.* V_i *indicates the imaged speed of target i and* V_{ptz} *indicates the maximum speeds of the pan-tilt-zoom device. The solution is defined as the permutation of the discrete set S that has the shortest travel time.*

It is necessary that the intercepter run faster than the targets. This is not generally a problem even for slower PTZ-cameras. By imagining the PTZ-camera as a robot manipulator with two revolute (pan-tilt) and one prismatic (zoom) joint, it is possible to view the principal axis of the camera as a robot arm which rotates and move forward to reach a point in the space. In such settings, due to the typically high distance at which PTZ-cameras are mounted, the speeds of the virtual end-effector are generally higher than common moving targets such as cars or humans.

4.3.2 Time Dependent Orienteering (TDO)

In a typical surveillance application, targets arrive as a continuous process, so that we must collect "demands to observe", plan tours to observe targets, and finally dispatch the PTZ camera. In a such dynamic-stochastic setting there is a lot of interdependency between the state variables describing the system. Moreover, tours must be planned while existing targets move or leave the scene, and/or new targets arrive. Basically the whole problem can be viewed as a global dynamic optimization. Since for such a problem

no a-priori solution can be found, an effective approach is to determine a strategy to specify the actions to be taken as a function of the state of the system. In practice, we consider the whole stochastic-dynamic problem as a series of deterministic-static subproblems, with the overall goal of tracking the time progression of the objective function as close as possible. In our problem, targets are assumed to enter the scene at any time from a finite set of locations. The camera must steer its foveal sensor to observe any target before it leaves the scene. Assuming with no loss of generality that the paths of the targets are straight lines and that targets move at constant speeds, the time by which a target must be observed by the camera can be estimated. Moreover, real-time constraints may impose bounds on the total amount of time needed to plan the target observation tour. According to this, given a fixed reference time, KTSP can be reformulated as a Time Dependent Orienteering (TDO) problem [100]. In the classical formulation of the static orienteering problem there is a resource constraint on the length of the tour; the problem solution is the one that maximizes the number of sites visited. The time dependent orienteering problem for a single PTZ-camera can be formulated as follows:

TDO : *Given a set* $S = \{s_1, s_2, ..., s_n\}$ *of moving targets, each* s_i *moving with a known or predictable motion* $x_i(t)$ *, the deadline* t*, and a time-travel function* $l : S \times S \times N \longmapsto I\!R^+ \cup \{0\}$ *the salesperson's tour to intercept a subset* $T = \{s_1, s_2, ..., s_m\}$ *of* m *targets is a sequence of triples:* $(s_1, t_1^+, t_1^-), (s_2, t_2^+, t_2^-), ...(s_m, t_m^+, t_m^-)$*, such that: for* $i \in \{1, 2, ..., m\}$*,* $t_i^+, t_i^- \in N \cup \{0\}$ *with* $0 = t_1^+ \leq t_1^- \leq t_2^+ \leq ... \leq t_m^+ \leq t_m^- \leq t$ *. The subset* T *is composed by the maximum number of targets interceptable within the time* t*, imposed by the real-time constraint.*

Orienteering problems are classified as path-orienteering or cycle-orienteering

problems depending on whether the network to be induced by the set of pairs of consecutive targets visited is supposed take the form of a path or of a cycle, respectively. The deadline t breaks the dynamic problem into a sequence of static problems. Such a formulation has a great advantage which is computationally helpful. Since there is no polynomial time algorithms to solve the KTSP, it is impossible to solve an instance of the KTSP problem with more than eight or nine targets in a fraction of a second, by the exhaustive search. However even if such an algorithm did exist the time needed to switch to all the targets would be so large that novel targets would not be observed due to the time needed to complete the tour. *So, the brute force approach enumerating and evaluating all the subsets permutations perfectly fits with the nature of our dynamic incremental formulation.*

4.3.3 Deadlines

Based on the tracking predictions targets are put in a queue, according to their residual time to exit the scene. TDO is instantiated for the first k targets in the queue. If $\mathcal{A}_k$ is the set of the permutations of the subsets of k targets then it can be shown that:

$$|\mathcal{A}_k| = \sum_{i=0}^{k} \frac{k!}{(k-i)!} \tag{4.1}$$

where $|\mathcal{A}_k|$ is the cardinality of the set $\mathcal{A}_k$. So for example with a queue of $k = 7$ targets we have $|\mathcal{A}_7| = 13700$. In this case the exhaustive enumeration requires 13700 solutions evaluations. As remarked in the previous section, solutions with a large number of scheduled targets would not be practical for an incremental solution, since the time needed to switch to all the targets would be so large that novel targets would not be observed due to the time needed to complete the tour.

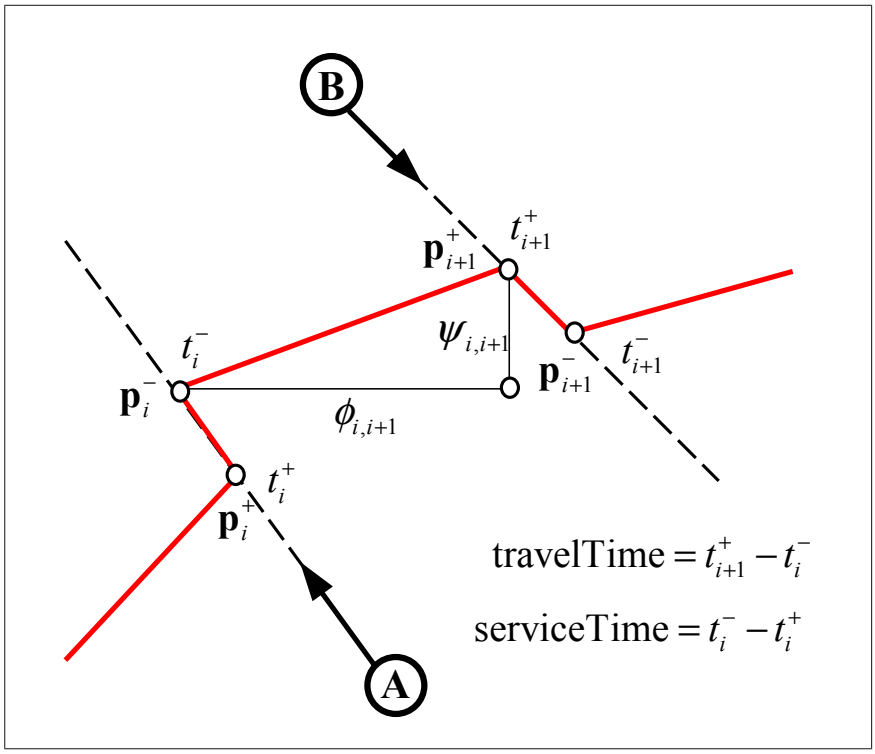


Figure 4.2: A symbolic scheme representing a saccade from the target A to the target B. The $\phi_{i,i+1}$, $\psi_{i,i+1}$ are respectively the pan and tilt angles as seen from the slave camera when the camera leaves target A at time t_i^- and intercepts B at time t_{i+1}^+.

The framework is fairly general and more elaborated policies can be estimated by changing optimization cost and/or the sorting used in the queue (priority in the queue can be specified according to some combined quality measure of the imagery of the targets, for example preferring targets moving in certain specified directions). Here we want to maximize the number of targets taken at high resolution. With the deadlines the TDO becomes a constrained combinatorial optimization, where the feasible set can be defined as follow (see the TDO definition in the previous section):

$$t_i^- < t_i^d \quad , \quad \forall i = 1..|T| \tag{4.2}$$

Where $T \in \mathcal{A}_k$ is an instance of the permutations of the subsets, and t_i^d is the deadline for the target at position i in T. That means the the camera must leaves the target i in T at time t_i^- before the target leaves the scene at time t_i^d.

The TDO solution is calculated by assuming a constant speed for the pan-tilt-zoom camera motors as specified by the manufacturer. There is no need for an exact specification of these speeds, in that they are used only for the prediction of the cost of the saccadic sequences. In order to keep the computation tractable the number of target in the queue k should not be greater than 8 (9 with optimized code). For example on a Pentium IV 2.0 GHz running Matlab, computing and evaluating the permutations of the subsets of 8 targets takes a fraction of a second.

4.4 Saccades Planning Geometry

In order to show the advantages of adopting this framework for our research objective, we consider the classic camera system in a master/slave configuration [87][86]. In this configuration a static, wide field of view master camera

is used to monitor a wide area and track the moving targets providing the position information to the foveal camera. The foveal camera is used to observe the targets at high resolution. We estimate the interception times of a target for each of the three foveal camera control signals (respectively t_ϕ, t_ψ, t_z for pan, tilt, zoom). Since the effects of the three control signals are independent from each other (i.e. the pan motor operates independently from the tilt motor) the time needed to conclude a saccade is dominated by the largest one. The largest time is taken as the time spent by the foveal camera to observe the target and is taken into account to derive the overall time needed to complete the tour in the TDO formulation.

With reference to fig.4.2 the estimated t_ϕ, t_ψ, t_z are assumed as the times needed to make the foveal camera gaze at the target at position $i+1$, leaving the target at position i in the sequence $S = \{s_1, ..., s_i, s_{i+1}, ..., s_m\}$ (in fig.4.2 the targets at position i and $i+1$ are respectively indicated as **A** and **B**). In other words they represent the times needed for changing the pan and tilt angles and zoom respectively by $\phi_{i,i+1}$, $\psi_{i,i+1}$ and $z_{i,i+1}$ (not shown in the figure) in order to intercept the new target at time t^+_{i+1} while leaving the old target at time t^-_i. The time $t^\star = \max\{t_{\phi_{i,i+1}}, t_{\psi_{i,i+1}}, t_{z_{i,i+1}}\}$ is the travel time needed to change the gaze.

By assuming targets moving on a calibrated plane, these times can be computed, at least in principle, by solving for t from each of the following equations:

$$\phi(t) = \omega_\phi t + \phi_{t^-_i} \qquad \psi(t) = \omega_\psi t + \psi_{t^-_i} \tag{4.3}$$

Where $\phi(t)$ and $\psi(t)$ are time varying functions, representing the angles between rays from the image points corresponding to the target trajectory w.r.t to a reference ray in the foveal camera. The ω_ϕ and ω_ψ are, respectively, the pan and tilt angular speeds and the angles $\phi_{t^-_i}$ and $\psi_{t^-_i}$ represent the

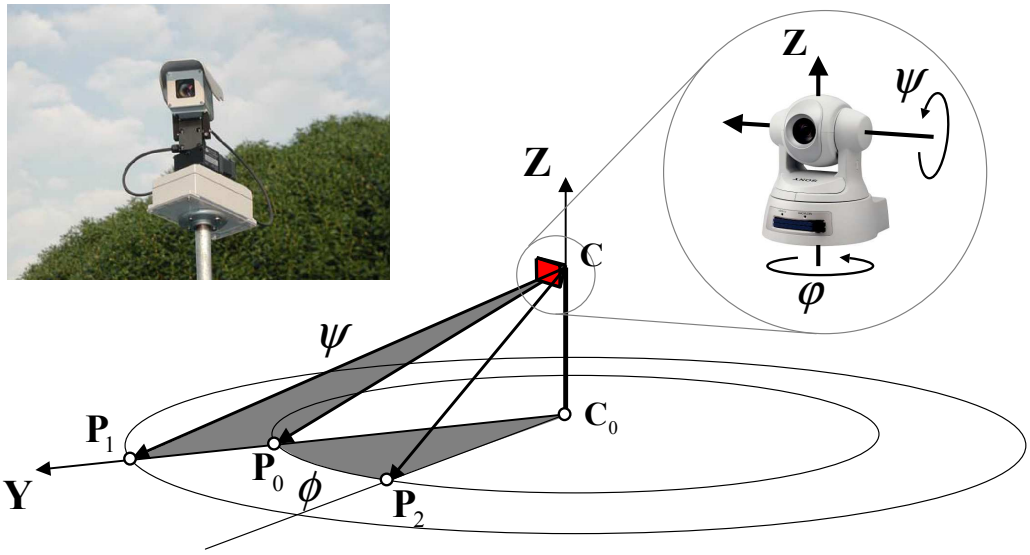


Figure 4.3: The geometry of a PTZ camera viewing a world plane in which the pan axis coincides with the normal of the plane. Also shown are the angles ϕ and ψ travelled by the pan-tilt device gazing from the target $\mathbf{P}_1$ to the target $\mathbf{P}_2$.

angle positions at time t_i^-. By separately solving the two equations in t we estimate the interception times t_ϕ and t_ψ, needed to intercept the target through pan and tilt camera motion. Each of the above equations is non-linear due to the image formation process. In order to make the TDO problem solvable, a closed form solution is obtained by assuming that during the camera interception process, the target motion is negligible. Now the TDO can be solved by exhaustive enumeration without an iterative root finder for the eq.4.3. With this assumption eq.4.3 becomes time independent and simplifies:

$$\phi_{t_{i+1}^+} = \omega_\phi t + \phi_{t_i^-} \qquad \psi_{t_{i+1}^+} = \omega_\psi t + \psi_{t_i^-} \tag{4.4}$$

defining the values for

$$t_{\phi_{i,i+1}} = \frac{\phi_{t_{i+1}^+} - \phi_{t_i^-}}{\omega_\phi} \qquad t_{\psi_{i,i+1}} = \frac{\psi_{t_{i+1}^+} - \psi_{t_i^-}}{\omega_\psi} \tag{4.5}$$

In order to keep tractable the estimate of the angles of the targets as seen

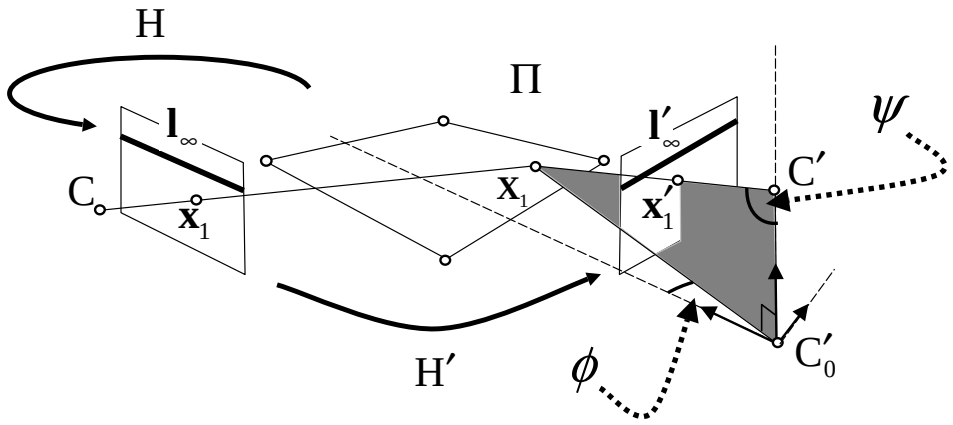
by the slave camera we assume that the PTZ-camera is not mounted oblique w.r.t. the world plane. The camera pan axis it is approximately aligned with the normal of the world plane. This is generally the case when PTZ-cameras are mounted on top of a pole (see fig.4.3). This means that during continuous panning while keeping a fixed angle for the tilt, the intersection of the optical axis with the 3D plane approximately describes a circle. The principal axis sweeps a cone surface so its intersection with the 3D world plane is in general an ellipse with an eccentricity close to one. In the same sense during continuous tilting while keeping a fixed angle for the pan, the intersection of the optical axis with the 3D plane describes approximately a line. The swept surface is a plane (see fig.4.3). In such conditions the tilt angle between a reference ray and the ray emanating from the image point corresponding to a target trajectory can be measured once the intrinsic internal camera parameters for the slave camera are known as [101]:

$$cos(\psi) = \frac{\mathbf{x'_1}^T \omega \mathbf{x'_0}}{\sqrt{\mathbf{x'_1}^T \omega \mathbf{x'_1}} \sqrt{\mathbf{x'_0}^T \omega \mathbf{x'_0}}} \tag{4.6}$$

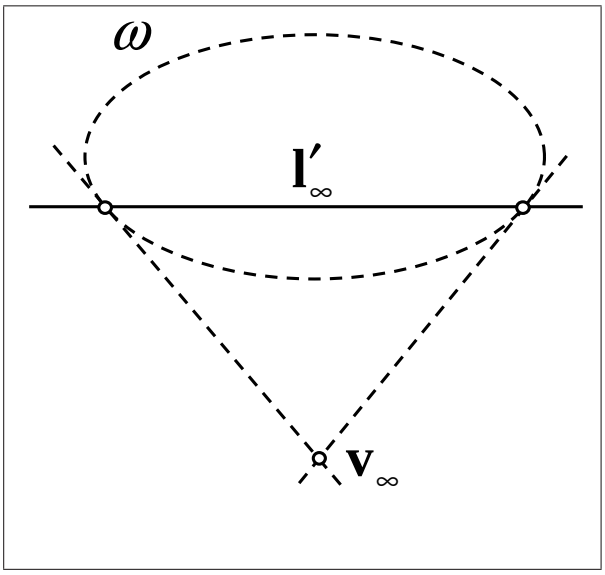
where $\omega = \mathtt{K}^{-\mathrm{T}}\mathtt{K}^{-1}$ is the image of the absolute conic an imaginary point conic directly related to the internal camera matrix $\mathtt{K}$. While $\mathbf{x'_1}$ and $\mathbf{x'_0}$ (as also shown in fig.4.4) are, respectively, the projection of the world point $\mathbf{X}_1$ as seen by the master camera and transformed through $\mathtt{H}'$ to the slave camera, and the projection of the point $\mathbf{C}'_0$.

$\mathbf{C}'_0$ is the orthogonal projection of the camera center of the slave camera $\mathbf{C}'$ onto the world plane. By choosing as reference ray to represent tilt angles of the ray passing through $\mathbf{C}'_0$ and $\mathbf{C}_0$ as shown in fig.4.4, the value of $\mathbf{x'_0}$ can be computed directly using the pole-polar relationship as:

$$\mathbf{x'_0} = \omega^{-1}\mathbf{l}'_\infty \tag{4.7}$$



(a)



(b)

Figure 4.4: (a) The geometry used for computing the pan ϕ and tilt ψ angles of a target $\mathbf{X}_1$ as seen from the slave camera $\mathbf{C}'$ in its home position between. (b) Pole-polar relationship between vanishing point $\mathbf{v}_\infty$ of the plane normal and its the vanishing line $\mathbf{l}'_\infty$ used to compute the tilt angle ψ. The IAC is shown dashed to remind that it is a pure imaginary conic.

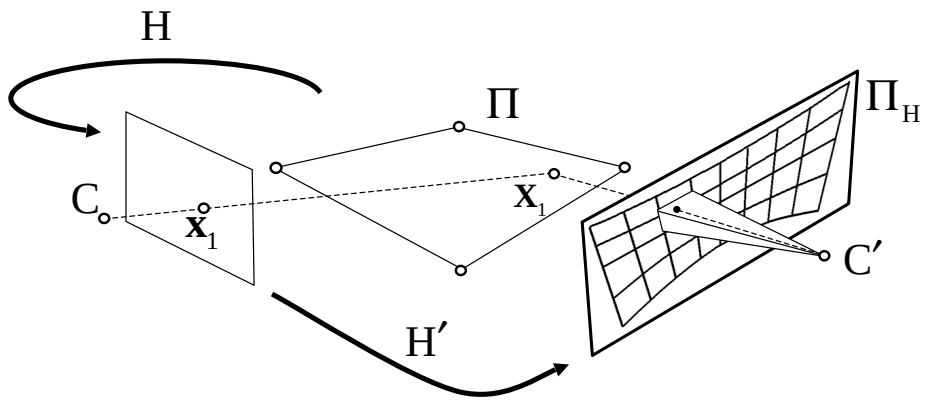


Figure 4.5: The slave camera is internally calibrated and the inter-image homography $\mathtt{H}'$ between the master camera $\mathbf{C}$ and the slave camera $\mathbf{C}'$ is computed in its home position (image plane Π_H). We can consider the slave camera as an angle measurement device using the extended image plane composed of the planar image mosaic having Π_H. as a reference plane.

Where $\mathbf{l}'_\infty$ is the vanishing line of the plane Π as seen from the slave camera and it can be computed by transferring the vanishing line $\mathbf{l}_\infty$ in the master camera to the slave camera as $\mathbf{l}'_\infty = \mathtt{H}'^{-\mathtt{T}}\mathbf{l}_\infty$. The above formula can be applied because $\mathbf{x}'_\mathbf{0}$ coincides with the vanishing point of the directions normal to the plane Π (see [65]). Summarizing, in this configuration the slave camera, in addition to its foveal capability also uses calibration (in its home position) as angle measurement device. Internal camera parameters necessary for the PTZ-camera can be computed very accurately as recently shown in [102] using the method originally described in [103].

The pan angle of a world point in the plane can be computed directly from the master camera once the world to image homography $\mathtt{H}$ is known and the point $\mathbf{C}'_0$ is measured from the master camera. If that point cannot be measured because it is not visible from the master camera, it can also be computed using the inter-image homography $\mathtt{H}'$. In fact since the slave

camera is internally calibrated at its home position, it is possible to obtain its pose and so its camera center w.r.t. the world reference once the world to image homography $\mathtt{H}'_0$ is known from the slave camera. This can be computed as: $\mathtt{H}'_0 = \mathtt{H}'\mathtt{H}$ (see fig.4.4).

The same approach of eq.4.4 is followed to obtain the zoom control, once the amount of zoom needed to obtain the desired close-up is calibrated for each point in the world plane. A look-up table using an equispaced grid of points can be used to perform this calibration manually or automatically as shown in [92]. The equation for the estimation of the time needed for changing the zoom to intercept the new target can be written similarly as for pan and tilt:

$$z_{t^+_{i+1}} = v_z t + z_{t^-_i} \tag{4.8}$$

where v_z is the zooming speed and $z_{t^-_i}$ is the zooming value at time t^-_i, when the target is left and $z_{t^+_{i+1}}$ is the zooming value at time t^+_{i+1} when the next target is intercepted.

4.5 Simulation Results

4.5.1 Estimating Camera Speeds

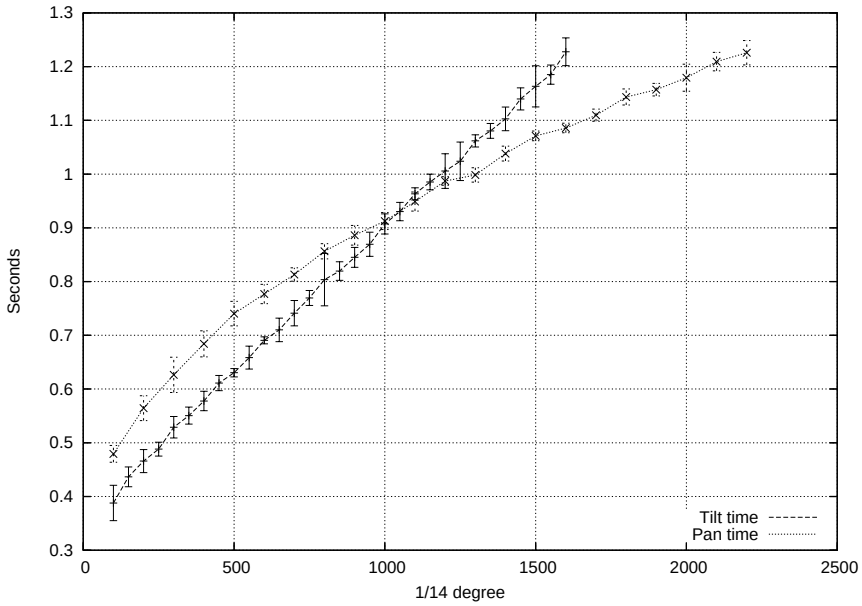
We ran several experiments to empirically estimate the pan/tilt/zoom speeds of our cameras in order to validate the constant velocity kinematic models used in the eq.4.4 and eq.4.8. The results of these experiments are shown figure 4.6. In particular we have conducted several trials and then we have averaged the results in fig.4.6(a) are shown the pan and tilt speeds while in fig.4.6(b) are reported the zoom speeds. Worthy of note is the fact that, contrary to manufacturer specification, the cameras do not move at a constant speed. Indeed, there are situations in which either panning or tilting

might be the slowest of motions, as indicated by the crossover point of the two curves in figure. When moving such short distances, camera motion is nearly instantaneous and we found that assuming a constant camera velocity when planning a saccade sequence worked just as well as the more complex camera performance model.

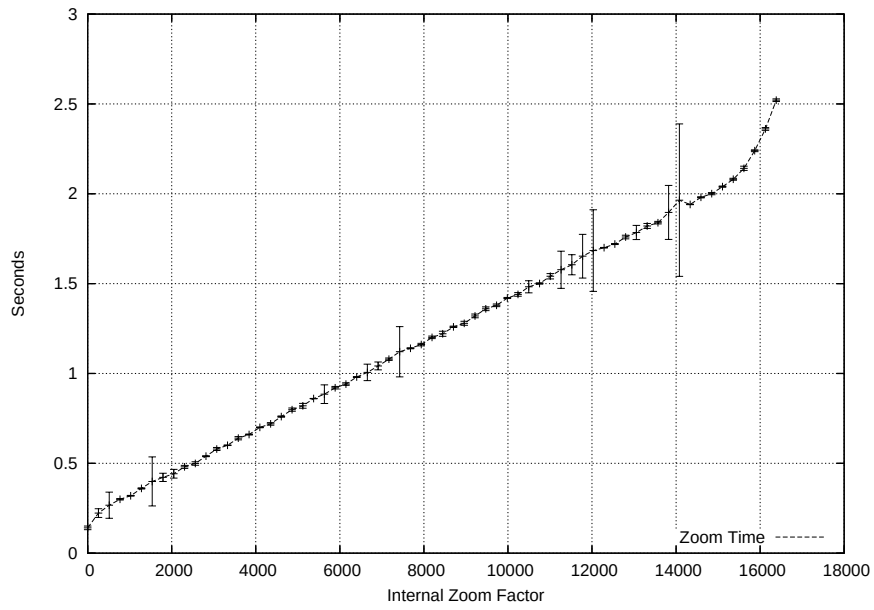
4.5.2 Congestion Analysis

Evaluating different planning strategies using a video surveillance system installed in a real context is a very complicated task. In fact, while we can easily collect video from a static camera, and use it for target tracking, it is almost impossible to collect all the information needed to plan tours in a master-slave camera configuration with a foveal slave camera. To address these difficulties, we have created a Monte Carlo simulation for evaluating scheduling policies using randomly generated data. But there is also another main reason for using randomly generated data. The use of randomly generated data often enables more in-depth analysis, since the datasets can be constructed in such a way that other issues could be addressed. For example the arrival rate parameter, generally denoted λ, describes the "congestion" of the system. This is basically the only important parameter which is worth of testing in a similar scenario. We stress the importance of this kind of testing: real data testing cannot evaluate the algorithm performance in this context.

We performed a Monte Carlo simulation that permits evaluating the effects of different scheduling policies in a congestion analysis setting. We used in our simulator a particular scene in which our framework could be of invaluable benefit. A large area of approximatively 50x60 meters (half of a soccer field) is monitored with the slave camera placed as shown in fig.4.7



(a)



(b)

Figure 4.6: Empirically estimated pan-tilt (a) and zoom (b) times for the Sony SNC-RZ30, averaged over thirty trials.

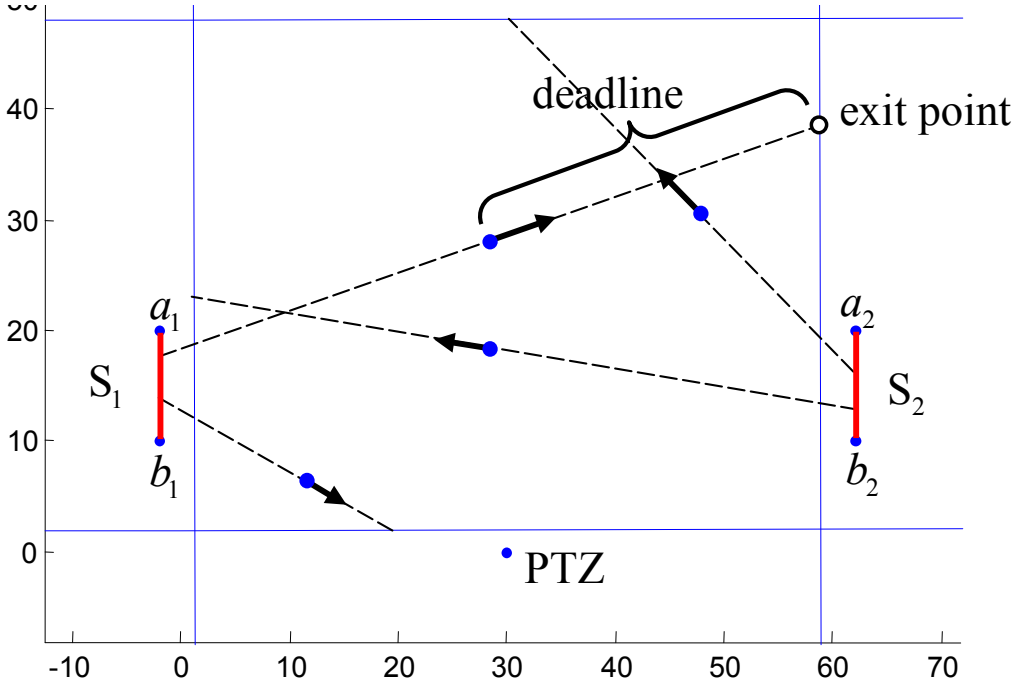


Figure 4.7: The simulated surveillance scene. The segment $\mathbf{S_1}$ and $\mathbf{S_2}$ represent sources where targets originates with a given statistics.

at position $(30, 0, 10)$. The master camera views the monitored area at a wide angle from above (more suitable for tracking due low occlusion between target). Arrivals of targets are modelled as a Poisson process. The scene is composed of two target sources situated at opposite positions in the area. Targets originate from these two sources S_1 and S_2 from initial positions that are uniformly distributed in given ranges of length 10 meters positioned as shown in fig.4.7. The starting angles for targets are also distributed uniformly with the range $[-40, 40]$ degrees. Target speeds are generated from a truncated Gaussian with a mean of 3.8 meter/sec and standard deviation of 0.5 meter/sec. (typical of a running person) and are kept constant for the duration of target motion. Targets follow a linear trajectory. This is not a restrictive assumption since each TDO has in this simulation a deadline of $t = 5$ seconds, and the probability of maneuvering for targets with a running-human dynamic in an interval of five seconds is very low. So the overall performance of the system is not generally affected. The deadline t has a role similar to a sampling time for traffic behavior and can be generally tuned depending on the speeds of the targets. In our simulated scene it is quite improbable that a target enters and exits the scene before five seconds are elapsed.

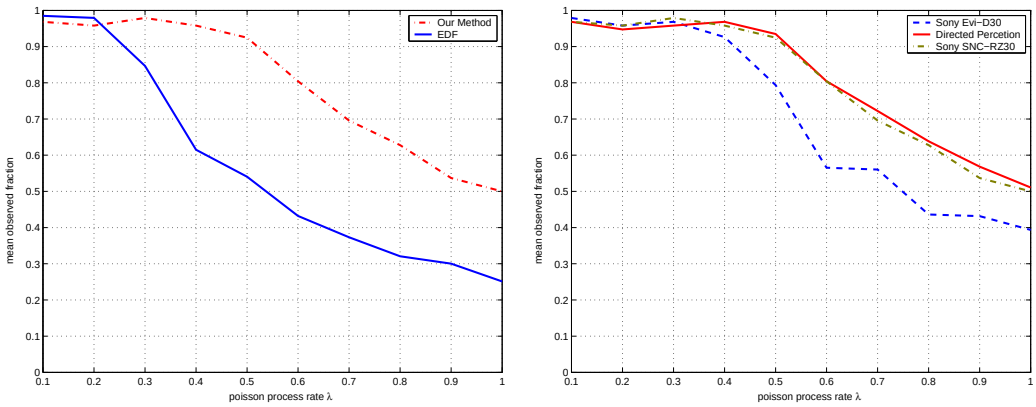
The used scene can represent a continuous flow of people, in a crisis situation. An example is people exiting from a stadium or from the subway stairs. It can be interesting, for crime detection purposes, to acquire as many high resolution images of such running people as possible before they leave the scene.

We assume that all targets have the same size in the scene (average humans height) and a specific size is fixed at which the target must be observed by the foveal camera. For pinhole cameras, as the focal length of the cam-

	Pan Speed deg/sec	Tilt Speed deg/sec	Zoom Speed #mag/sec
Sony EVI-D30	80	50	0.6
Sony SNC-RZ30	170	76.6	8.3
Directed Perception	300	300	11.3

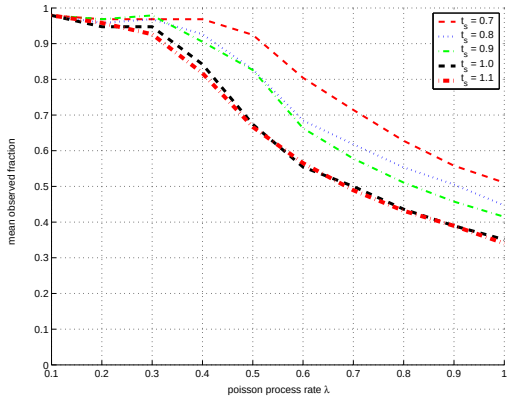
Table 4.1: Off the shelf PTZ-cameras performance. The #mag means magnification factor per second and is calculated dividing the maximum optical zoom (for example 25X) by the zoom movement time from wide to tele (for example 2.2 seconds).

era changes, the pinhole model predicts that the images will scale in direct proportion to the focal length [104]. By assuming a constant speed for the zooming motor and a linear mapping of focal length to zoom it is possible to build a look-up table in the simulator as: $\text{Zoom}[x, y] = M \cdot \text{dist}(\mathbf{C}', \mathbf{X})$ where x and y are the imaged coordinates of the world plane point $\mathbf{X}$ as seen by the master camera, $\mathbf{C}'$ is the camera center of the slave camera and M is the constant factor which depends on the size at which targets are imaged and on the target size in the scene. We want to collect human imagery with an imaged height of approximatively 350 pixels using an image resolution of 720×576. In fig.4.8, plots indicate the number of targets that are observed by the foveal camera (ordinates) as a function of the arrival rate λ (abscissa) for three different situations. Since there are two sources with the same arrival rate, λ actually refers to half the number of arrivals per second. The size of the queue is six elements which guarantees that the enumeration of all the subsets with their permutations is generated in a fraction of a second (basically a negligible time). Performance is measured by running a scenario in which 500 targets are repeatedly generated one hundred times and the performance metric was estimated by taking the mean. The metric corre-



(a)

(b)



(c)

Figure 4.8: Policy performance versus arrival rate λ. (a) Our methods and simple earliest deadline first policy. (b) Three different PTZ-camera under test with different pan-tilt-zoom speed. (b) Performance variation at varying service time t_s (the specified time to watch a target).

sponds to the fraction of people observed in the scene. In particular we take the mean (over the experiments) of the number of observed target divided by number of all the targets.

Fig.4.8(a) shows a comparison of our methods with the earliest deadline first policy studied in [86]; it evident that our policy, using long term planning plus the cost of moving the sensor, outperforms a simple greedy strategy. While there is no need for planning in very modest traffic scenes, traffic monitoring, in large, wide areas would receive an invaluable great advantage of more than 40% by adopting the proposed techniques. Fig.4.8(b) shows experiments conducted using different speeds for PTZ motors typical of off-the-shelf active cameras. Three cameras were selected using their respective performance as indicated by the technical specification (see tab.4.1). Using this performance values in the simulator produce the plots of fig.4.8(b). Although the three models are very different in performance, such differences are less evident for the observing task under test. This is mostly caused by the camera position w.r.t. the scene plane; the performance in tilt speed was practically never employed because of the latency of the other controls w.r.t. the imaged motion pattern of targets. The control which delayed most of the saccades, employing the largest setup time, was the zoom control (mostly caused by the scene depth). This explains why the two fastest cameras exhibit similar performances. This type of analysis can be useful for determining the type of cameras and ultimately the cost needed to monitor an area with a multi-camera system.

Fig.4.8(c) shows the performance degradation w.r.t. the service time (or the watching time) t_s. This time is directly related to the quality of the acquired images and can potentially affect recognition results. The figure also shows that varying t_s does not affects the performances in direct proportion.

4.6 Summary

Automated high resolution imaging of targets using PTZ cameras is an important and mandatory capability for modern automated surveillance. In such systems, and especially in the case of wide area surveillance applications, to view multiple moving targets each camera must share observation time. We have presented a solution for planning saccade sequences using a single foveal camera in a master-slave camera system configuration. The system models the attentional gaze planning, with a novel approach combining ideas from Dynamic Vehicle Routing Problem (DVRP) and multiview geometry. Results are presented using a simulator that indicates how many targets are missed as a function of the arrival rate, camera speed parameters and watching time. Results have been derived under realistic assumptions in a challenging scene. We proved that our framework gives good performance in monitoring wide areas with little extra effort with respect to other cumbersome approaches coordinating a large number of cameras doing the same task.

The same principles presented here can also be applied to camera-networks to build large surveillance systems; the framework is open and may be extended easily in several different ways; e.g. a real-time face recognition/detection can be incorporated in the optimization.

One main limitation of the presented method is that it does not take advantage of persistent motion patterns generally present in common scenes, for example an intersection with moving cars. Such knowledge would be of invaluable benefit in cases where targets are following pre-defined paths. Ongoing research will address on-line learning algorithms capable of finding

more long-term policies. Moreover, further research can apply supervised machine learning methods to a simulated data set (as generated by our approach) to understand the behaviors of complex saccadic patterns for the task under consideration.

Chapter 5

Conclusion

In conclusion of the thesis, this chapter describes what are believed to be the novel contributions of the work and suggests areas of further research.

5.1 Novel Contribution

The main contributions of the thesis with reference to the recent literature can be summarized as follows:

- Single-view reconstruction based on planar rectification, originally introduced in [49] [7] for planar surfaces, has been extended to deal also with the SOR class of curved surfaces.

- Self-calibration of a natural camera (3 dofs) is obtained from a single image of an SOR. This improves the approach presented in [6], in which the calibration of a natural camera requires the presence of two different SORs in the same view. Moreover, since self-calibration is based on two visible elliptical segments, it can also be used to calibrate turntable sequences and remove the 1D projective reconstruction ambiguity due

to underconstrained calibration experienced so far in the literature of motion-constrained reconstruction [5].

- The invariant-based description of the SOR scaling function discussed in [8] is extended from affine to perspective viewing conditions. 4. Since the approach exploits both the geometric and topological properties of the transformation that relates the apparent contour to the scaling function, a metric localization of occluded parts can be performed and the scaling function can be reconstructed piecewise. In this regard, the method improves the SOR reconstruction approach described in [105] [26].

- Texture acquisition does not require the explicit computation of external camera parameters; therefore, the results developed in [7] and [20] for planar surfaces are extended to the SOR class of curved surfaces. Moreover, since SORs are a superclass of the SUGC class of curved surfaces, texture acquisition extends the solution presented in [106] [37].

- Image mosaicing from single views of surfaces of revolution is firstly proposed and a two step method for registering images is given.

- An original approach to automatically grouping and estimation of the projective geometry of single SOR views was presented. The approach is mainly devoted to camera autocalibration from a single SOR view and single view metric reconstruction of SOR objects.

- We have presented a very novel problem and a solution for planning saccade sequences using a single foveal camera in a master-slave camera system configuration. The system models the attentional gaze plan-

ning, with a novel approach combining ideas from Dynamic Vehicle Routing Problem (DVRP) and Projective Geometry of multiple images.

5.2 Future Research

Multiview geometry is now a very mature subject, the book [11] and the software Boujou© [107] bring close an era in Computer Vision. Indeed papers regarding the geometry of multiple images are becoming infrequent in these two last years. This is mainly due to the rigorous framework of Projective Geometry. So it seems to be difficult to extend the proposed framework for reconstructing Surfaces of Revolution. While it is worth extending the approach of mosaicing images of SORs. For example further investigation can be done in anti-aliasing filtering during the resampling step.

Automated high resolution imaging of targets using PTZ cameras is an important and mandatory capability for modern automated surveillance. In such systems, and especially in the case of wide area surveillance applications, to view multiple moving targets each camera must share observation time. Active cameras in this sense are an under-researched topic in computer vision. The lack of works addressing task-driven visual processing is mainly motivated by the fact that its studying seems, as a first sight, too specialized, non-generic, or bordering on hackery. But active vision demands such processes; it is founded in the idea of specialized processing for specialized tasks. Most of the active vision literature is limited to studying low-level subconscious reflexes. One wonders whether truly active and purposeful vision systems will be realized. In other words, while active tracking and visual attention was researched in the past years, purposeful zooming is (and prob-

ably will remain) a largely unexplored area in active vision [82]. Basically sensing was not a major issue for computer vision as for example was perception. However despite this for the particular task of object recognition notably works are reported in the literature.

The same principles presented in the second part of this thesis can also be applied to camera-networks to build large surveillance systems; the framework is open and may be extended easily in several different ways; e.g. a real-time face recognition/detection can be incorporated in the optimization. Ongoing research will address on-line learning algorithms capable of finding more long-term policies. Moreover, further research can apply supervised machine learning methods to a simulated data set (as generated by our approach) to understand the behaviors of complex saccadic patterns for the task under consideration.

Appendix A

Appendix of the Part A

A.1 Computing the harmonic homology from the vertices of the complete quadrangle

In this appendix, we give the formal proof of Eqs. 1.5 and 1.6 used to compute the fixed entities of the harmonic homology $\mathtt{H}$ from the four intersections $\mathbf{x}_k$, $k = 1, \dots 4$ of two imaged cross sections $\mathtt{C}_1$ and $\mathtt{C}_2$. As explained in section 1.4.1, we can always assume that $\mathbf{x}_1$ and $\mathbf{x}_2$ are complex conjugate, so that either of the pairs ($\mathbf{x}_1$,$\mathbf{x}_2$) or ($\mathbf{x}_3$,$\mathbf{x}_4$) must be equal to ($\mathbf{i}$,$\mathbf{j}$), and therefore either of the lines $\mathbf{l}_{12} = \mathbf{x}_1 \times \mathbf{x}_2$ or $\mathbf{l}_{34} = \mathbf{x}_3 \times \mathbf{x}_4$ must be equal to $\mathbf{l}_\infty = \mathbf{i} \times \mathbf{j}$.

By property 1.3.2 of section 1.3.2, the conics $\mathtt{C}_1$ and $\mathtt{C}_2$ are fixed as a set under the harmonic homology: $\mathtt{C}_h = \mathtt{H}^{\mathrm{T}}\mathtt{C}_h\mathtt{H}, \quad h = 1, 2$. A consequence of this is that if $\mathbf{x}_k$ is an intersection point of $\mathtt{C}_h$, so is the point $\mathtt{H}\mathbf{x}_k$: $(\mathtt{H}\mathbf{x}_k)^{\mathrm{T}}\mathtt{C}_h(\mathtt{H}\mathbf{x}_k) = 0$. By expressing $\mathtt{H}$ according to the parametrization

$$\mathtt{H} = \mathtt{I} - 2\frac{\mathbf{v}_\infty \mathbf{l}_s^{\mathrm{T}}}{\mathbf{v}_\infty^{\mathrm{T}} \mathbf{l}_s} \tag{A.1}$$

obtained from eq. 1.1 with $\mu = -1$, we can write

$$\mathtt{H}\mathbf{x}_k = \mathbf{x}_k - 2\,\frac{\mathbf{l}_s^{\mathrm{T}}\mathbf{x}_k}{\mathbf{l}_s^{\mathrm{T}}\mathbf{v}_\infty}\,\mathbf{v}_\infty \ . \tag{A.2}$$

Now, since the line $\mathbf{x}_k \times \mathtt{H}\mathbf{x}_k$ must contain the fixed point $\mathbf{v}_\infty$, recalling that $(\mathbf{i} \times \mathbf{j})^{\mathrm{T}}\mathbf{v}_\infty = 0$ and that no three intersection points can be collinear, it follows that

$$\mathbf{x}_2 = \mathtt{H}\mathbf{x}_1 \ \text{ and } \ \mathbf{x}_4 = \mathtt{H}\mathbf{x}_3 \ . \tag{A.3}$$

This proves Eq. 1.5, as the lines $\mathbf{l}_{12}$ and $\mathbf{l}_{34}$ can be written respectively as $\mathbf{x}_1 \times \mathtt{H}\mathbf{x}_1$ and $\mathbf{x}_3 \times \mathtt{H}\mathbf{x}_3$. Using Eq. A.3, we can also write $\mathbf{l}_{13} \times \mathbf{l}_{24} = (\mathbf{x}_1 \times \mathbf{x}_3) \times (\mathtt{H}\mathbf{x}_1 \times \mathtt{H}\mathbf{x}_3)$ and $\mathbf{l}_{14} \times \mathbf{l}_{23} = (\mathbf{x}_1 \times \mathtt{H}\mathbf{x}_3) \times (\mathtt{H}\mathbf{x}_1 \times \mathbf{x}_3)$. By using again the parametrization of Eq. A.1 and the basic equality $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a}^{\mathrm{T}}\mathbf{c})\mathbf{b} - (\mathbf{a}^{\mathrm{T}}\mathbf{b})\mathbf{c}$, it follows easily that $\mathbf{l}_{13} \times \mathbf{l}_{24} = \mathbf{l}_{13} \times \mathbf{l}_s$ and $\mathbf{l}_{14} \times \mathbf{l}_{23} = \mathbf{l}_{14} \times \mathbf{l}_s$. This proves Eq. 1.6.

A.2 Parametrizing the image of the absolute conic

In this appendix, we demonstrate that the linear system of Eq. 1.7 has only three independent constraints, and provide a parametrization for the ∞^2 conics that satisfy these constraints.

The third of Eqs. 1.7 provides two independent linear constraints on ω. We will show that the first two equations of the system, i.e. $\mathbf{i}^{\mathrm{T}}\omega\mathbf{i} = 0$ and $\mathbf{j}^{\mathrm{T}}\omega\mathbf{j} = 0$, add to $\mathbf{l}_s = \omega\mathbf{v}_\infty$ only one independent constraint. Indeed, the family of ∞^3 conics $\widetilde{\omega}$ satisfying $\mathbf{l}_s = \widetilde{\omega}\mathbf{v}_\infty$ can be written as

$$\widetilde{\omega}(\lambda_1, \lambda_2, \lambda_3) = \Lambda_0 + \sum_{k=1}^{3} \lambda_k \Lambda_k \ , \tag{A.4}$$

where the λ_k's are scalars and the Λ_k's are four linearly independent conics such that

$$\mathbf{l}_s = \Lambda_k \mathbf{v}_\infty \ . \tag{A.5}$$

Now, in Appendix A.1 we have shown that $\mathbf{j} = \mathtt{H}\mathbf{i}$. Therefore, we can write

$$\mathbf{j}^{\mathrm{T}} \Lambda_k \mathbf{j} = \mathbf{i}^{\mathrm{T}} (\mathtt{H}^{\mathrm{T}} \Lambda_k \mathtt{H}) \mathbf{i} = \mathbf{i}^{\mathrm{T}} \Lambda_k \mathbf{i} \ , \tag{A.6}$$

where the last equality follows from the fact that, as it satisfies Eq. A.5, each of the Λ_k's is transformed onto itself by the homology $\mathtt{H}$—this can also be directly verified by using for $\mathtt{H}$ the parametrization of Appendix A.1. From Eq. A.6 it also follows that $\mathbf{j}^{\mathrm{T}} \widetilde{\omega} \mathbf{j} = \mathbf{i}^{\mathrm{T}} \widetilde{\omega} \mathbf{i}$: this means that the inhomogeneous linear system in the three unknowns λ_k's

$$\begin{cases} \mathbf{i}^{\mathrm{T}} \, \widetilde{\omega}(\lambda_1, \lambda_2, \lambda_3) \, \mathbf{i} = 0 \\ \mathbf{j}^{\mathrm{T}} \, \widetilde{\omega}(\lambda_1, \lambda_2, \lambda_3) \, \mathbf{j} = 0 \end{cases} \tag{A.7}$$

has ∞^2 solutions. This proves our assert that the solution set of Eq. 1.7 is composed of ∞^2 conics.

It can be easily verified that a valid parametrization for these conics is

$$\widetilde{\omega}(p, q) = \omega + p \, \mathbf{l}_\infty \mathbf{l}_\infty^{\mathrm{T}} + q(\mathbf{l}_{is} \mathbf{l}_{js}^{\mathrm{T}} + \mathbf{l}_{js} \mathbf{l}_{is}^{\mathrm{T}}) \ , \tag{A.8}$$

where ω is the (unknown) true image of the absolute conic, $\mathbf{l}_\infty \mathbf{l}_\infty^{\mathrm{T}}$ is a degenerate (rank 1) conic composed by the line $\mathbf{l}_\infty$ taken twice, and $\mathbf{l}_{is} \mathbf{l}_{js}^{\mathrm{T}} + \mathbf{l}_{js} \mathbf{l}_{is}^{\mathrm{T}}$ is a degenerate (rank 2) conic composed by the two lines $\mathbf{l}_{is} = \mathbf{i} \times \mathbf{x}_s$ and $\mathbf{l}_{js} = \mathbf{j} \times \mathbf{x}_s$ meeting at any point $\mathbf{x}_s \in \mathbf{l}_s$ different from $\mathbf{v}_s = \mathbf{l}_s \times \mathbf{l}_\infty$.

If the vanishing point $\mathbf{v}_\perp \in \mathbf{l}_s$ of the direction parallel to the SOR symmetry axis is known, the independent constraint $\mathbf{v}_s^{\mathrm{T}} \omega \mathbf{v}_\perp = 0$ can be added to the system of Eq. 1.7, thus fixing one of the two degrees of freedom left for $\widetilde{\omega}$. A parametrization for these ∞^1 conics is then

$$\widetilde{\omega}(r) = r \, \mathbf{l}_\infty \mathbf{l}_\infty^{\mathrm{T}} + (\mathbf{l}_{i\perp} \mathbf{l}_{j\perp}^{\mathrm{T}} + \mathbf{l}_{j\perp} \mathbf{l}_{i\perp}^{\mathrm{T}}) \ , \tag{A.9}$$

where $\mathbf{l}_{i\perp} = \mathbf{i} \times \mathbf{v}_\perp$ and $\mathbf{l}_{j\perp} = \mathbf{j} \times \mathbf{v}_\perp$. This last result is in accordance with the fact, discussed in [54], that the self-calibration equations involving the imaged circular points $\mathbf{i}$ and $\mathbf{j}$ bring only one independent constraint if the line $\mathbf{i} \times \mathbf{j}$ goes through any of the points of a self-polar triangle for ω—which, in our case, is $\mathbf{v}_\infty$,$\mathbf{v}_s$,$\mathbf{v}_\perp$.

A.3 Computing the point of an imaged SOR parallel for a given Euclidean angle

In this appendix, we derive a closed form solution to the general problem of finding the vanishing point $\mathbf{v}_\theta$ of the line l_θ that intersects, in the world plane π, a reference line l_0 with a given angle θ. The imaged circular points $\mathbf{i}$ and $\mathbf{j}$ of π are supposed to be known, together with the vanishing point $\mathbf{v}_0$ of l_0. We then use this result to obtain the intersection point $\mathbf{x}(\theta, z)$ between the image $\mathtt{C}(z)$ of the SOR parallel on π and the visible imaged meridian $\chi(\theta)$.

The basic relation between the angle θ and the vanishing point $\mathbf{v}_\theta$ is provided by the Laguerre's formula [12]

$$\theta = \frac{1}{2i} \log(\{\mathbf{v}_\theta, \mathbf{v}_0, \mathbf{i}, \mathbf{j}\}) \ , \tag{A.10}$$

where $\{\}$ denotes the usual cross ratio of four points. By expressing the generic point on the vanishing line $\mathbf{l}_\infty$ of π as

$$\mathbf{v}(\lambda) = \mathbf{i} + \lambda(\mathbf{i} - \mathbf{j}) \ , \tag{A.11}$$

Eq. A.10 can be rewritten as

$$e^{i2\theta} = \{\lambda_\theta, \lambda_0, \lambda_i, \lambda_j\} \ , \tag{A.12}$$

where λ_θ, λ_0, $\lambda_i = 0$ and $\lambda_j = -1$ are the values of the complex parameter λ respectively for the points $\mathbf{v}_\theta$, $\mathbf{v}_0$, $\mathbf{i}$ and $\mathbf{j}$. Taken any image line $\mathbf{l}_0 = (l_1, l_2, l_3)$

through $\mathbf{v}_0$ and distinct from $\mathbf{l}_\infty$, and set $\mathbf{i} = \text{conj}(\mathbf{j}) = (a + ib, c + id, 1)$, solving for λ_0 the equation $\mathbf{l}_0^{\mathrm{T}}\mathbf{v}(\lambda_0) = 0$ we get $\lambda_0 = -\frac{1}{2}\,[1 + i\tan\phi_0]$, where the angle

$$\phi_0 = \arctan\left(-\frac{l_1 a + l_2 c + l_3}{l_1 b + l_2 d}\right) \tag{A.13}$$

embeds in a compact way all the information about the reference line l_0 and the circular points.

Replacing the above value of λ_0 of into Eq. A.12, the value of λ_θ can be easily computed as

$$\lambda_\theta = -\frac{1}{2}\,[1 + i\tan(\phi_0 + \theta)] \quad , \tag{A.14}$$

which eventually yields the required vanishing point as $\mathbf{v}_\theta = \mathbf{i} + \lambda_\theta(\mathbf{i} - \mathbf{j})$.

In the particular case of a SOR image, the vanishing point $\mathbf{v}_\theta$ can be computed as above with the point $\mathbf{v}_s = \mathbf{l}_s \times \mathbf{l}_\infty$ and the image line $\mathbf{l}_s$ as the reference $\mathbf{v}_0$ and $\mathbf{l}_0$, respectively (see Fig. 2.3). The image line $\mathbf{l}_\theta = \mathbf{v}_\theta \times \mathbf{o}$—where $\mathbf{o} = \mathtt{C}^{-1}(z)\mathbf{l}_\infty$ is the image of the parallel's center—intercepts the imaged parallel $\mathtt{C}$ at two points, of which the required point $\mathbf{x}(\theta, z)$ on the visible imaged meridian $\chi(\theta)$ is the farthest one from $\mathbf{v}_\theta$ along the line $\mathbf{l}_\theta$.

A.4 Linear Constraints on the IAC

This section describes single view calibration constraints as geometric constructions involving the IAC and the fixed entities computed as shown in the first part of this thesis. The ω is represented as a symmetric matrix with elements:

$$\omega = \begin{pmatrix} \omega_1 & \omega_2 & \omega_4 \\ \omega_2 & \omega_3 & \omega_5 \\ \omega_4 & \omega_5 & \omega_6 \end{pmatrix} \tag{A.15}$$

The constraints can be expressed linearly as

$$\kappa^T \omega_{\mathbf{v}} = 0 \tag{A.16}$$

where $\omega_{\mathbf{v}} = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6)^T$ is the vector of the elements of ω and $\kappa = (\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6)^T$ is the vector of the coefficient of the constraint.

A.4.1 The constraint of the circular points

The constraint of the circular points:

$$\mathbf{I}^T \omega \mathbf{I} = 0 \qquad \mathbf{J}^T \omega \mathbf{J} = 0 \tag{A.17}$$

can be rewritten as

$$I_1 I_1 \omega_1 + (I_1 I_2 + I_2 I_1)\omega_2 + I_2 I_2 \omega_3 + (I_1 I_3 + I_3 I_1)\omega_4 + (I_2 I_3 + I_3 I_2)\omega_5 + I_3 I_3 \omega_6 = 0 \tag{A.18}$$

and in the vector representation it becomes:

$$\kappa_{\mathbf{IJ}} = (I_1 I_1, I_1 I_2 + I_2 I_1, I_2 I_2, I_1 I_3 + I_3 I_1, I_2 I_3 + I_3 I_2, I_3 I_3)^T \tag{A.19}$$

where $\mathbf{I} = (I_1, I_2, I_3)^T$; in practice, all the circular points information is contained in one of the complex conjugate points, say $\mathbf{I}$. In this case also $\kappa_{\mathbf{IJ}}$

is a complex vector, writing out the real and imaginary parts of the eq.A.19 yields two linear expressions in the elements of ω denoted in vector form

$$\begin{pmatrix} \mathrm{Re}[\kappa_{\mathbf{IJ}}] \\ \mathrm{Im}[\kappa_{\mathbf{IJ}}] \end{pmatrix}^T \omega_{\mathbf{v}} = 0 \tag{A.20}$$

A.4.2 The constraint of the pole-polar relationship

The last important constraint is expressed by the pole-polar relationship

$$\mathbf{l}_\infty = \omega \mathbf{v} \tag{A.21}$$

where $\mathbf{v}$ is the vanishing point of the direction orthogonal to the plane with the vanishing line $\mathbf{l}_\infty$. Eq.A.21 place two linear constraint on the elements of the IAC. The third constraint is dependent on the other two. The eq.A.21 can be transformed using the cross product as

$$\mathbf{l}_\infty \times \omega \mathbf{v} = \mathbf{0} \tag{A.22}$$

Indicating $\mathbf{l}_\infty = (l_1, l_2, l_3)^T$ and $\mathbf{v} = (v_1, v_2, v_3)^T$ and expanding eq.A.22 the constraints have the following form

$$-l_3v_1\omega_2 - l_3v_2\omega_3 + l_2v_1\omega_4 + (l_2v_2 - l_3v_3)\omega_5 + l_2v_3\omega_6 = 0$$

$$l_3v_1\omega_1 + l_3v_2\omega_2 + (l_3v_3 - l_1v_1)\omega_4 - l_1v_2\omega_5 - l_1v_3\omega_6 = 0$$

$$-l_2v_1\omega_1 + (l_1v_1 - l_2v_2)\omega_2 + l_1v_2\omega_3 - l_2v_3\omega_4 + l_1v_3\omega_5 = 0$$

in matrix form:

$$\kappa_{\mathbf{lv}} = \begin{pmatrix} 0 & -l_3v_1 & -l_3v_2 & l_2v_1 & (l_2v_2 - l_3v_3) & l_2v_3 \\ l_3v_1 & l_3v_2 & 0 & (l_3v_3 - l_1v_1) & -l_1v_2 & -l_1v_3 \\ -l_2v_1 & (l_1v_1 - l_2v_2) & l_1v_2 & -l_2v_3 & l_1v_3 & 0 \end{pmatrix}^T \tag{A.23}$$

or

$${\kappa_{\mathbf{lv}}}^T \omega_{\mathbf{v}} = 0$$

A.4.3 The constraint of the zero skew and unit aspect ratio

The general five parameter internal camera model may be simplified in many cases by a priori knowledge of some of the parameters. That camera skew is zero is the most commonly applied internal parameter constraint. It specifies that the vertical and horizontal imaging axes are orthogonal. This is a reasonable assumption for most cameras. In terms of the IAC this mean $\omega_2 = 0$ [108] and the corresponding constraint is

$$(1, 0, 0)\omega(0, 1, 0)^T = 0$$

In practice CDD cameras can often be considered to have zero skew and unit (or known) aspect ratio to have square pixels. A known aspect ratio r yields the following constraint [108]

$$\frac{\omega_3}{\omega_1} - \frac{\omega_2^2}{\omega_1^2} = \frac{1}{r^2} \tag{A.24}$$

In the case when also the camera skew is zero, i.e. $\omega_2 = 0$ the constraint of eq.A.24 becomes:

$$(1, r, 0)\omega(1, -r, 0)^T = 0$$

or

$$\omega_1 - r^2\omega_3 = 0$$

A.5 Closed form solution for the intersection of two ellipses

Given two ellipses $\mathtt{C}_1$ and $\mathtt{C}_2$ is there an analytical solution to solve for the points of intersection with Gröbner basis [47]. These solutions are the roots of a quartic polynomials whose coefficient are calculated in terms of the conic coefficient matrix $\mathtt{C}_1$ and $\mathtt{C}_2$.

Gröbner bases [47] provide a systematic computational method to find the intersection between two ellipses without testing for false solution as [109] or [110] and without relying in numerical methods.

Given $\mathtt{C}_1$ and $\mathtt{C}_2$ the coefficient matrix of two ellipses in homogeneous representation we want to find their intersection. If we consider the dual problem (i.e. $\mathtt{C}_1^*$ and $\mathtt{C}_2^*$) the problem becomes to find all the lines tangents to both the ellipses. In general this problem can be formulated as

$$\begin{cases} \mathbf{p}^T \mathtt{C}_1 \mathbf{p} = 0 \\ \mathbf{p}^T \mathtt{C}_2 \mathbf{p} = 0 \end{cases} \tag{A.25}$$

where $\mathbf{p} = (x, y, 1)^T$ defines a point in $\mathbb{R}^2$ and $\mathtt{C}_1$, $\mathtt{C}_2$ are the coefficients matrix of the two ellipses, respectively:

$$\mathtt{C}_1 = \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix} \tag{A.26}$$

and

$$\mathtt{C}_2 = \begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{pmatrix} \tag{A.27}$$

By substituting eq.A.26 and eq.A.27 in the system of eq.A.25 the system becomes

$$\begin{cases} x^2 a + x\,y\,b + y\,d + y^2 c + y\,e + f = 0 \\ x^2 A + x\,y\,B + y\,D + y^2 C + y\,E + F = 0 \end{cases} \tag{A.28}$$

by using the Gröbner basis the system becomes:

$$\begin{cases} p_4 x^4 + p_3 x^3 + p_2 x^2 + p_1 x + p_0 = 0 \\ q_1 y + q_0 = 0 \end{cases} \tag{A.29}$$

where the polynomials coefficients $p_0, p_1, p_2, p_3, p_4, q_0, q_1$ can be calculated for example using a computer algebra system such as Mathematica© or Maple©. Using Maple© i have obtained the following results:

$$p_0 = f^2C^2 - fCEe - 2\,fCFc + fE^2c + CFe^2 - EFce + F^2c^2 \tag{A.30}$$

$$\begin{aligned} p_1 = {} & -2\,CfcD - bCfE + 2\,C^2df - CeBf + 2\,EcBf + CDe^2 \\ & - cEeD + 2\,c^2FD + 2\,bCFe - bcEF - CdEe \\ & - 2\,CcFd + cE^2d - cBeF \end{aligned} \tag{A.31}$$

$$\begin{aligned} p_2 = {} & 2\,C^2fa - bCBf - 2\,CfcA + cB^2f - CeEa - 2\,CcFa + cE^2a \\ & + c^2D^2 + 2\,bCDe - bcED - 2\,CcDd - cBeD + b^2CF - bCdE \\ & - bFcB + Ce^2A - cEeA + 2\,c^2FA + C^2d^2 - CeBd + 2\,EcBd \end{aligned} \tag{A.32}$$

$$\begin{aligned} p_3 = {} & -2\,CacD - bCEa + 2\,C^2ad - CeBa + 2\,EcBa + b^2CD \\ & - bDcB + 2\,c^2AD + 2\,bCeA - bcEA - bCBd - 2\,CcAd \\ & - cBeA + cB^2d \end{aligned} \tag{A.33}$$

$$p_4 = ACb^2 - bBaC - bAcB + a^2C^2 - 2\,acAC + cB^2a + c^2A^2 \tag{A.34}$$

$$\begin{aligned}
q_0 = &-b^3xC^2F - b^3Ax^3C^2 - Db^3x^2C^2 + 2\,DbxCceB + 2\,bAx^2CceB \\
&- b^2C^2Fe + faC^3e - fbC^3d - bxC^3d^2 + fb^2C^2E - 2\,abx^2C^3d \\
&+ 2\,Db^2x^2CcB + 2\,b^2xCFcB - b^2Ax^2C^2e - Db^2xC^2e - bAxCFc^2 \\
&+ fbAxC^2c - fabxC^3 + 2\,bAx^2C^2dc - 2\,DbAx^2Cc^2 + 2\,aDbx^2C^2c \\
&+ abxC^2Fc - bA^2x^3Cc^2 - 2\,abx^3CcB^2 + xC^2d^2cB - a^2bx^3C^3 \\
&- 2\,bx^2CdcB^2 - 2\,fbxCcB^2 + fb^2xC^2B + 2\,abAx^3C^2c - bAx^3c^2B^2 \\
&+ 2\,DAx^2c^3B - 2\,aDx^2Cc^2B - Dxc^2eB^2 - Ax^2c^2eB^2 - Dbx^2c^2B^2 \\
&- Fc^2eB^2 + 2\,ax^2C^2dcB - axCFc^2B + faxC^2cB + x^2dc^2B^3 \\
&+ fxc^2B^3 - 2\,aAx^3Cc^2B + a^2x^3C^2cB + AxFc^3B - fAxCc^2B \\
&- 2\,Ax^2Cdc^2B - bxFc^2B^2 + fC^2dcB - CFdc^2B + A^2x^3c^3B \\
&+ ax^3c^2B^3 + b^2xC^2Ed + D^2xc^3B + 2\,DbxC^2dc + fEc^2B^2 \\
&- 2\,fbCEcB - fDCc^2B + DFc^3B - 2\,abx^2CEcB + ab^2x^2C^2E \\
&+ ax^2Ec^2B^2 + xEdc^2B^2 - 2\,DxCdc^2B + bC^2Fdc - fAC^2ce \\
&+ ACFc^2e - D^2bxCc^2 + fDbC^2c - DbCFc^2 - 2\,bxCEdcB \\
&+ AxCEdc^2 - aDxC^2ce - a^2x^2C^2Ec - faC^2Ec + 2\,aAx^2CEc^2 \\
&- A^2x^2Ec^3 + fACEc^2 + A^2x^2Cc^2e - AxC^2dce + a^2x^2C^3e \\
&+ aCEFc^2 - axC^2Edc + aDxCEc^2 - aC^2Fce - DAxEc^3 \\
&- AEFc^3 + axC^3de + 2\,bCFceB + 2\,b^2Ax^3CcB + b^2x^2C^2dB \\
&+ ab^2x^3C^2B - 2\,aAx^2C^2ce + DAxCc^2e
\end{aligned} \tag{A.35}$$

$$
\begin{aligned}
q_1 = {} & fb^2C^3 - 2\,fbC^2cB + fCc^2B^2 + aC^3e^2 - 2\,aC^2Ece + aCE^2c^2 \\
& + DbC^2ce - DbCEc^2 - DCc^2eB + DEc^3B - b^2C^2Fc - bC^3de \\
& + bC^2Edc + 2\,bCFc^2B - AC^2ce^2 + 2\,ACEc^2e - AE^2c^3 + C^2dceB \\
& - CEdc^2B - Fc^3B^2
\end{aligned}
\tag{A.36}
$$

The roots of the quartic polynomial are the x coordinates of the intersection points, the corresponding y coordinates can be computed by plugging the roots x_i, $i = 1..4$ in the linear equation in eq.A.29 obtaining:

$$
y_i = \frac{q_0(x_i)}{q_1} \quad i = 1..4 \tag{A.37}
$$

Appendix B

Appendix of the Part B

B.1 Absolute Conic and rotating cameras

The image of the absolute conic is the projection of the absolute conic Ω. This is an imaginary point conic that lies on the plane at infinity Π_∞ in 3D and has the property that it is invariant to similarity transformations of space [101]. The conic relevant to calibration is its projection onto the image plane, i.e. the image of the absolute conic ω (IAC). The IAC is related to the camera calibration parameter by $\omega = \mathtt{K}^{-\mathrm{T}}\mathtt{K}^{-1}$. The calibration matrix $\mathtt{K}$ may be computed from ω according to the Cholesky decomposition.
One important property of the IAC is that it can be transferred from one image to another through the infinite homography $\mathtt{H}^\infty$ as:

$$\omega_i = \mathtt{H}_{ij}^{\infty -T}\omega_j\mathtt{H}_{ij}^{\infty -1} \tag{B.1}$$

Once we have $\mathtt{H}^\infty$ the equation above can be used to impose constraint on ω. Points at infinity (like for example vanishing point) are mapped between views by the infinite homography $\mathtt{H}^\infty$ and for this is independent on translation between view. In particular when there is no translation between the

views, the infinite homography $\mathtt{H}^{\infty}$ relates points of any depth. This simplification can be exploited when images are taken with cameras having a common center. The $\mathtt{H}^{\infty}$ coincide with the inter-image homographies, so we have a convenient method of measuring $\mathtt{H}^{\infty}$ directly from images.

B.2 Computing the Slave internal internal camera parameter

Internal camera parameters necessary for the PTZ-camera can be computed very accurately as recently shown in [102] using the method originally described in [103]. When images are taken with cameras all located at the same camera center point in space camera matrices can be simplified. It is possible to analyze this situation by representing each camera as a 3×3 matrix instead of a general 3×4 camera matrix. A point in the i-th image, represented by a homogeneous 3-vector $\mathbf{X}_i$ corresponds o a ray in space consisting of points of he form $\lambda \mathtt{P}_i^{-1}\mathbf{x}_i$. Points on this ray are mapped into the j-th image to a point $\mathbf{x} = \mathtt{P}_j\mathtt{P}_i^{-1}\mathbf{x}_i$. Denoting the transformation $\mathtt{H}_{ij} = \mathtt{P}_j\mathtt{P}_i^{-1}$ one sees that he i-th and j-th images are related by a projective planar transformation $\mathtt{H}_{ij}$. Clearly this can be estimated by at least four matched points. Each transformation estimated by point correspondences is related to the internal camera parameter as eq.B.1. Once the homographies are known the equation above can be expressed linearly in the terms of ω. If the skew is zero which is usual in modern cameras, there are four unknown in the internal camera parameters: focal length (1 DOF), principal point (2 DOF) and aspect ratio (1 DOF). Four homographies suffices to compute the minimal solution. Infact each equation provides a single constraint on ω. Since we need the internal camera parameter in a specified home position only for computing angles we

don't need using the zoom. This means that the internal camera parameter does not variyng while panning and tilting so the in the eq.B.1 $\omega_i = \omega_j = \omega$ and becomes

$$\omega = \mathtt{H}_{ij}^{-T} \omega \mathtt{H}_{ij}^{-1} \tag{B.2}$$

Bibliography

[1] M. Pollefeys, *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*, Ph.D. thesis, K.U.Leuven, 1999. 3

[2] S. Bougnoux, "From projective to euclidean space under any practical situation, a criticism of self-calibration," in *ICCV 98*, 1998, pp. 790–796. 3

[3] R.I. Hartley and R. Kaucic, "Sensitivity of calibration to principal point position," in *ECCV02*, 2002, p. II: 433 ff. 3

[4] C. Colombo, A. DelBimbo, and F. Pernici, "Uncalibrated 3D metric reconstruction and flattened texture acquisition from a single view of a surface of revolution," *3DPVT02*, pp. 277–284, 2002. 4, 8, 54, 56, 70, 83, 84

[5] G. Jiang, H.T. Tsui, L. Quan, and A. Zisserman, "Single axis geometry by fitting conics," in *ECCV02*, 2002, p. I: 537 ff. 4, 6, 15, 19, 120

[6] Kwan-Yee Kenneth Wong, P. R. S. Mendonça, and R. Cipolla, "Camera calibration from surfaces of revolution," *PAMI*, vol. 25, no. 2, pp. 147–161, February 2003. 4, 8, 15, 70, 74, 119

[7] D. Liebowitz, A. Criminisi, and A. Zisserman, "Creating architectural models from images," in *Proc. EuroGraphics*, September 1999, vol. 18, pp. 39–50. 5, 6, 59, 69, 119, 120

[8] S. M. Abdallah and Zisserman, "Grouping and recognition of straight homogeneous generalized cylinders.," *In Proc. Asian Conference on Computer Vision*, vol. II, pp. 850–857, 2000. 5, 8, 13, 120

[9] Kwan-Yee Kenneth Wong, P. R. S. Mendonça, and R. Cipolla, "Reconstruction of surfaces of revolution from single uncalibrated views," in *Proc. British Machine Vision Conference 2002*, P. L. Rosin and D. Marshall, Eds., Cardiff, UK, September 2002, British Machine Vision Association, vol. 1, pp. 93–102. 5, 8, 70

[10] G. Slabaugh, W. B. Culbertson, T. Malzbender, and R. Schafer, "A survey of volumetric scene reconstruction methods from photographs," in *Volume Graphics 2001, Proc. of Joint IEEE TCVG and Eurographics Workshop*, K. Mueller and A. Kaufman, Eds., Stony Brook, New York, USA, June 2001, pp. 81–100, Springer Computer Science. 5

[11] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521623049, 2000. 5, 14, 22, 51, 54, 121

[12] Olivier Faugeras and Quan-Tuan Luong, *The Geometry of Multiple Images*, The MIT press Cambridge, Massachusetts London, England, 2001. 5, 28, 57, 127

[13] R. Szeliski, "Rapid octree construction from image sequences," *CVGIP*, vol. 58, no. 1, pp. 23–32, July 1993. 5

[14] Steven M. Seitz and Charles R. Dyer, "Photorealistic scene reconstruction by voxel coloring," in *Proc. Image Understanding Workshop*, 1997, pp. 935–942. 5

[15] D. Scharstein and R. Szeliski., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.," *International Journal of Computer Vision*, vol. 47(1/2/3):7-42, April–June 2002. 5

[16] O.D. Faugeras and R. Keriven, "Variational-principles, surface evolution, pdes, level set methods, and the stereo problem," *IP*, vol. 7, no. 3, pp. 336–344, March 1998. 5

[17] J. Mundy and A. Zisserman, "Repeated structures: Image correspondence constraints and ambiguity of 3D reconstruction," in *Applications of invariance in computer vision*, J. Mundy, A. Zisserman, and D. Forsyth, Eds., pp. 89–106. Springer-Verlag, 1994. 5, 54, 72

[18] L. Van Gool, M. Proesmans, and A. Zisserman, "Planar homologies as a basis for grouping and recognition," *Image and Vision Computing*, vol. 16, pp. 21–26, January 1998. 6

[19] C. J. Taylor P. E. Debevec and J. Malik., "Modeling and rendering architecture from photographs: A hybrid geometry and image based approach.," *In Proceedings, ACM SIGGRAPH*, pp. 11–20, 1996. 6

[20] Peter Sturm and Steve Maybank, "A method for interactive 3d reconstruction of piecewise planar objects from single images," in *British Machine Vision Conference, Nottingham, England*, Sep 1999, pp. 265–274. 6, 39, 69, 120

[21] P. Sturm and S.J. Maybank., "On plane-based camera calibration: A general algorithm, singularities, applications.," *CVPR - IEEE Interna-*

tional Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, pp. 432–437, June 1999. 6

[22] G. Cross and A. Zisserman, "Quadric surface reconstruction from dual-space geometry," *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pp. 25–31, Jan 1998. 6

[23] M. Armstrong, *Self-Calibration from Image Sequences*, Ph.D. thesis, University of Oxford, England, 1996. 6, 14

[24] A. W. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3D model construction for turn-table sequences," in *3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506*, R. Koch and L. Van Gool, Eds. June 1998, pp. 155–170, Springer-Verlag. 6, 15

[25] G. Cross and A. Zisserman, "Surface reconstruction from multiple views using apparent contours and surface texture," in *NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics, Ljubljana, Slovenia*, A. Leonardis, F. Solina, and R. Bajcsy, Eds., 2000, pp. 25–47. 6

[26] Paulo R. S. Mendonça, Kwan-Yee Kenneth Wong, and Roberto Cipolla, "Epipolar geometry from profiles under circular motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 604–616, 2001. 6, 15, 16, 120

[27] A. Rappoport M. Werman E. Ofek, E. Shilat, "Multiresolution textures from image sequences," *IEEE Computer Graphics and Applications*, vol. 17, no. 2, pp. 18–29, March–April 1997. 7

[28] P. J. Neugebauer and Klein, "Texturing 3d models of real world objects from multiple unregistered photographic views.," *Computer Graphics Forum*, vol. 3, no. 18, pp. 245–256, September 1999. 7

[29] Cignoni P. Rocchini, C. and C. Montani, "Multiple textures stitching and blending on 3d objects.," *Eurographics RenderingWorkshop*, june 1999. 7

[30] W. Heidrich H. Lensch and H.-P. Seidel., "Automated texture registration and stitching for real world models.," *Pacific Graphics 00*, pp. 317–326, October 2000. 7

[31] Jean Ponce, David Chelberg, and Wallace B. Mann, "Invariant properties of straight homogeneous generalized cylinders and their contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-11, no. 9, pp. 951–966, 1989. 7

[32] Samer M. Abdallah, *Object Recognition via Invariance*, Ph.D. thesis, The University of Sydney, Australia, June 2000. 8, 9, 13, 22, 39, 69, 80

[33] Ari D. Gross and Terrance E. Boult, "Recovery of SHGCs from a single intensity view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 161–180, 1996. 8

[34] K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla, "Camera calibration from symmetry," in *Proc. 9th IMA Conference on the Mathematics of Surfaces*, R. Cipolla and R. Martin, Eds., Cambridge, UK, September 2000, Institute of Mathematics and its Applications, pp. 214–226, Springer–Verlag. 8

[35] Sven Utcke and Andrew Zisserman, "Projective reconstruction of surfaces of revolution," in *25. DAGM-Symposium Mustererkennung*,

Magdeburg, Germany, Sept. 2003, DAGM, pp. 265–272, Springer-Verlag, Berlin. 8

[36] Hui Zhang, Kwan-Yee Kenneth Wong, and Mendonça P., "Reconstruction of surface of revolution from multiple uncalibrated views: a bundle-adjustment approach," *ACCV*, Jan 2004. 8

[37] I. Pitas J.-M. Chasssery W. Puech, A.G. Bors, "Projection distortion analysis for flattened image mosaicing from straight uniform generalized cylinders," *Pattern Recognition*, vol. 34, no. 8, pp. 1657–1670, Aug 2001. 8, 10, 51, 53, 120

[38] H. Sato and T.O. Binford, "Finding and recovering SHGC objects in an edge image," *CVGIP: Image Understanding*, vol. 57, pp. 346–358, 1993. 9

[39] H. Sato and T.O. Binford, "On finding the ends of shgcs in an edge image," *Proceedings of Computer Vision and Pattern Recognition*, pp. 695–698, 1992. 9

[40] J. Liu, J. Mundy, D. Forsyth, A. Zisserman, and C. Rothwell, "Efficient recognition of rotationally symmetric surfaces and straight homogeneous generalized cylinders," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1993. 9

[41] M. Zerroug and R. Nevatia, "Volumetric descriptions from a single intensity image.," *International Journal of Computer Vision*, vol. 20(1/2), pp. 11–42, 1996. 9, 69

[42] Fausto Bernardini and Holly Rushmeier, "The 3d model acquisition pipeline," *Computer Graphics forum*, vol. 21, no. 2, pp. 149, 172 2002. 9

[43] D.W. Eggert and K.W. Bowyer, "Computing the perspective projection aspect graph of solids of revolution," *PAMI*, vol. 15, no. 2, pp. 109–128, February 1993. 11, 37

[44] J.J. Koenderink, "What does the occluding contour tell us about solid shape?," *Perception*, vol. 13, pp. 321–330, 1984. 11

[45] J. Semple and G. Kneebone, *Algebraic projective geometry*, Oxford University Press, 1952. 12, 16, 24, 57

[46] D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. A. Rothwell, "Recognising rotationally symmetric surfaces from their outlines," in *Proc. European Conference on Computer Vision*. 1992, LNCS 588, Springer-Verlag. 13, 74

[47] T. Sederberg and J. Zheng, *Handbook of Computer Aided Geometric Design*, chapter Algebraic methods for Computer Aided Geometric Design, 2001. 16, 31, 132

[48] E. Hayman L. de Agapito and I. Reid, "Self-calibration of rotating and zooming cameras," Tech. Rep. OUEL 0225/00, Department of Engineering Science, University of Oxford, September 2000. 21

[49] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 1998, pp. 482–488. 26, 119

[50] Luiz Henrique de Figueiredo, "Adaptive sampling of parametric curves," in *Graphics Gems V*, Alan Paeth, Ed., pp. 173–178. Academic Press, Boston, 1995. 27, 30

[51] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976. 28

[52] Paul Heckbert, "Fundamentals of texture mapping and image warping," M.S. thesis, CS Division, U.C. Berkeley, June 1989, UCB/CSD 89/516. 28, 30

[53] R.B. Fisher M. Pilu, "Model-driven grouping and recognition of generic object parts from single images," *Proc. 4th Int. Symp. on Intelligent Robotic Systems, Lisbon*, pp. 147–154, July 1996. 31, 69

[54] D. Liebowitz, *Camera Calibration and Reconstruction of Geometry from Images*, Ph.D. thesis, University of Oxford, Dept. Engineering Science, June 2001, D.Phil. thesis. 39, 127

[55] A. Zisserman, J. Mundy, D. Forsyth, J. Liu, N. Pillow, C. Rothwell, and S. Utcke, "Class-based grouping in perspective images," in *Proc. International Conference on Computer Vision*, 1995. 40

[56] M. Irani and P. Anandan, "All about direct methods," in *International Workshop on Vision Algorithms: Theory and Practice*, W. Triggs, A. Zisserman, and R. Szeliski, Eds. 1999, pp. 267–277, Springer Verlag. 51

[57] P. H. S. Torr and A. Zisserman, "Feature based methods for structure and motion estimation," in *International Workshop on Vision Algorithms: Theory and Practice*, W. Triggs, A. Zisserman, and R. Szeliski, Eds. 1999, pp. 278–295, Springer Verlag. 51

[58] R.Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Application*, pp. 22–30, March 1996. 51

[59] L. McMillan and G. Bishop., "Plenoptic modeling: An image-based rendering system.," *Computer Graphics (SIGGRAPH'95)*, pp. 39–46, August 1995. 51

[60] S. E. Chen., "Quicktime vr – an image-based approach to virtual environment navigation.," *Computer Graphics (SIGGRAPH'95)*, pp. 29–38, August 1995. 51

[61] H.-Y. Shum and R. Szeliski., "Systems and experiment paper: Construction of panoramic mosaics with global and local alignment," *International Journal of Computer Vision*, vol. 36, no. 2, pp. 101–130, February 2000. 51, 62

[62] A.Rav-Acha and A.Zomet S.Peleg, B.Rousso, "Mosaicing on adaptive manifolds," *ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1144 1154, October 2000. 51

[63] Hong Shen, Charles V. Stewart, Badrinath Roysam, Gang Lin, and Howard L. Tanenbaum, "Frame-rate spatial referencing based on invariant indexing and alignment with application to online retinal image registration," *IEEE PAMI*, vol. 25, no. 3, pp. 379–384, march 2003. 51

[64] Beraldin-J-A. Godin G.-Cournoyer L. Baribeau R. Rioux M. Taylor, J. and J. Domey, "Nrc 3d imaging technology for museum and heritage applications," *Journal of Visualization and Computer Animation*, vol. 14, no. 3, pp. 121–138, July 2003. 52, 67

[65] C. Colombo, A. Del Bimbo, and F. Pernici, "Metric 3d reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 99–114, 2005. 54, 108

[66] B. Lucas and T. Kanade., "An iterative image registration technique with an application to stereo vision.," *In Proc. of IJCAI*, pp. 674–679, 1981. 62

[67] C. Colombo, A. Del Bimbo D. Comanducci, and F. Pernici, "Accurate automatic localization of surfaces of revolution for self-calibration and metric reconstruction," *IEEE Workshop on Perceptual Organization in Computer Vision (POCV '04). In association with the IEEE CVPR '04*, June 2004. 67

[68] Heung-Yeung Shum, Rick Szeliski, Simon Baker, Mei Han, and P. Anandan, "Interactive 3d modeling from multiple images using scene regularities," in *Proceedings of the European Workshop on 3D Structure from Multiple Images of Large-scale Environments (at ECCV '98)*, 1998. 69

[69] F. Schaffalitzky and A. Zisserman, "Planar grouping for automatic detection of vanishing lines and points," vol. 18, pp. 647–658, 2000. 69

[70] A. Bartoli., "Piecewise planar segmentation for automatic scene modeling," *CVPR'01 - In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. vol. II, pp. 283–289, December 2001. 69

[71] C. Baillard and A. Zisserman, "Automatic reconstruction of piecewise planar models from multiple views," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, jun 1999, pp. 559–565. 69

[72] A.R. Dick, P.H.S. Torr, S.F. Ruffle, and R. Cipolla, "Combining single view recognition and multiple view stereo for architectural scenes," *In*

Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, 2001. 69

[73] P.L. Rosin and G.A. West., "Non-parametric segmentation of curves into various representations.," *PAMI*, vol. 17, no. 12, pp. 1140–1153, December 1995. 69

[74] D. S. Chen, "A data-driven intermediate level feature extraction algorithm," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 11, no. 7, pp. 749–758, 1989. 69

[75] N.D. McKay P.J. Besl, "A method for registration of 3-d shapes," *IEEE Trans. on Pattern Analisys and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992. 72

[76] R Fisher, "Projective icp and stabilizing architectural augmented reality overlays," *Proc. Int. Symp. on Virtual and Augmented Architecture (VAA01), Dublin, Ireland,*, pp. 69–80, Jun 2001. 72

[77] Annick M. Leroy Peter J. Rousseeuw, *Robust Regression and Outlier Detection*, John Wiley and Sons, ISBN: 0-471-85233-3, December 1987. 73

[78] Luiz Henrique de Figueiredo and Jonas Gomes, "Computational morphology of curves," *The Visual Computer*, vol. 11, no. 2, pp. 105–112, 1995. 75

[79] Andrew W. Fitzgibbon, Maurizio Pilu, and Robert B. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, 1999. 82, 83

[80] A. Yarbus, *Eye Movements and Vision*, Plenum Press, 1967. 91

[81] R.K. Bajcsy., "Active perception," *Proc of the IEEE*, 1988. 91

[82] Markus A. Stricker Michael J. Swain, "Promising directions in active vision," *International Journal of Computer Vision*, 1993. 92, 122

[83] J. Denzler and C.M. Brown., "Information theoretic sensor data selection for active object recognition and state estimation.," *Pattern Analysis and Machine Intelligence*, 2002. 92

[84] M. Prantl H. Borotschnig, L. Paletta and A. Pinz, "Appearance based active object recognition," *Image and Vision Computing*, 2000. 92

[85] R. Tanawongsuwan S.Stillman and I. Essa., "A system for trackng and recognizing multiple people with multiple cameras.," *Technical Report GIT-GVU-98-25 Georgia Institute of Technology, Graphics, Visualization, and Usability Center*, 1998. 94, 95

[86] A. Banerjee H. Fisher C. J. Costello, C. P. Diehl, "Scheduling an active camera to observe people," *Proceedings of the 2nd ACM International Workshop on Video Surveillance and Sensor Networks*, 2004. 94, 95, 103, 116

[87] T. Kanade X. Zhou, R. Collins and P. Metes., "A master-slave system to acquire biometric imagery of humans at a distance.," *ACM SIGMM 2003 Workshop on Video Surveillance*, 2003. 95, 96, 103

[88] P. Peixoto J. Batista and H. Araujo., "Real-time active visual surveillance by integrating peripheral motion detection with foveated tracking.," *In Proceedings of the IEEE Workshop on Visual Surveillance*, pp. 18–25, 1998. 95

[89] Y. Hou M. Sizinstev S. J. D. Prince, J. H. Elder, "Pre-attentive face detection for foveated wide-field surveillance," *IEEE Workshop on Applications on Computer Vision*, 2005. 95

[90] L. Marcenaro L. Marchesotti and C. Regazzoni, "Dual camera system for face detection in unconstrained environments," *ICIP*, 2003. 95

[91] L.S Elgammal A. Ser-Nam Lim, Davis, "Scalable image-based multi-camera visual surveillance system," *Proceedings IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 205–212, 2003. 95

[92] Arun Hampapur Andrew Senior and Max Lu, ""acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration"," *IEEE Workshop on Applications on Computer Vision*, 2005. 95, 96, 97, 109

[93] A. Senior Y.-L. Tian L. Brown A. Hampapur, S. Pankanti and R. Bolle., "Face cataloger: Multi-scale imaging for relating identity to location.," *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003. 95

[94] H. Fujiyoshi R. Collins, A. Lipton and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1456–1477, 2001. 95

[95] D. Comaniciu Heinrich Niemann M. Greiffenhagen, V. Ramesh, "Statistical modeling and performance characterization of a real-time dual camera surveillance system," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00)*, 2000. 95

[96] P. F. McLauchlan I. D. Reid D. W. Murray, K. J. Bradshaw and P.M. Sharkey, "Driving saccade to pursuit using image motion," *Int. Journal of Computer Vision*, 1995. 95

[97] R. Motwani P. Chalasani and A. Rao., "Approximation algorithms for robot grasp and delivery," *In Proceedings of the 2nd International Workshop on Algorithmic Foundations of Robotics*, 1996. 96

[98] Dimitris Bertsimas and Garrett Van Ryzin., "A stochastic and dynamic vehicle routing problem in the euclidean plane," *Operations Research*, 1991. 96

[99] G. Robins Helvig, C.S. and A. Zelikovsky, "The moving-target traveling salesman problem," *Journal of Algorithms*, 2003. 97

[100] F.V. Fomin and A. Lingas, "Approximation algorithms for time-dependent orienteering," *Information Processing Letters*, 2002. 100

[101] R. I. Hartley and A. Zisserman., "Multiple view geometry in computer vision.," *Cambridge University Press, second edition*, 2004. 106, 136

[102] S.N. Sinha and M. Pollefeys., "Towards calibrating a pan-tilt-zoom cameras network," *P. Sturm, T. Svoboda, and S. Teller, editors, OMNIVIS*, 2004. 108, 137

[103] R. Hartley L. de Agapito and E. Hayman., "Linear selfcalibration of a rotating and zooming camera.," *In Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1999. 108, 137

[104] B.J. Tordoff, "Active control of zoom for computer vision," *DPhil thesis, Univ. of Oxford*, 2002. 114

[105] P. R. S. Mendonça, *Multiview Geometry: Profiles and Self-Calibration*, Ph.D. thesis, University of Cambridge, Cambridge, UK, May 2001. 120

[106] W. Puech, A. Bors, J. Chassery, and I. Pitas, "Mosaicing of paintings on curved surfaces," 1996. 120

[107] 2d3 Ltd. Boujou, "http://www.2d3.com," . 121

[108] A. Zisserman, D. Liebowitz, and M. Armstrong, "Resolving ambiguities in auto-calibration," *Philosophical Transactions of the Royal Society of London, SERIES A*, vol. 356, no. 1740, pp. 1193–1211, 1998. 131

[109] Kenneth J. Hill, "Matrix-based ellipse geometry," *Graphics Gems V*, vol. 5, pp. 72–77, 1994. 132

[110] David Eberly, "Intersection of ellipses," Magic Software, Inc. www.magic-software.com. 132