Information Engineering Research Group
Department of Engineering Science
University of Oxford

# Modelling and Using Uncertainties in Video Metrology

First Year Report


Antonio Criminisi
The Queen's College

July 27, 1997

# Abstract

Using vision to measure world distances requires that both measurements and their uncertainties can be determined and modelled. The work described in this report develops the theory of computing distances using only images and an uncertainty analysis which includes both the errors in image localisation and the uncertainty in the imaging transformation.

We present different methods of estimating the geometry of the imaging transformation between world and camera plane using image to world point correspondences. A general expression is then derived for the uncertainties in the measurements.

We mainly focus our attention on computing measures and related uncertainties on world planes and then we extend these results to a complete 3D reconstruction of the scene.

While the research presented here is of a general nature, a particular application is presented showing how to use the results obtained to build a plane measuring device. This device finds its application in several commercial and industrial fields.

# Acknowledgements

I am very grateful to my family for their great help and support.

I would like to thank Dr.Andrew Zisserman and Dr.Ian Reid for giving me the opportunity to work in this exciting area of research, and for their active supervision and enthusiasm.

I am very grateful to Dr.Roberto Cipolla for his suggestions relating to the Parallax Theory and its uses.

I would like to thank Dr.Andrew Fitzgibbon for his help with Target Junior libraries and Philipp Pritchett for his suggestions on image warping software.

Thanks to David Liebowitz for his help with Camera Self-Calibration Theory, to Nic Pillow, Geoff Cross and David Capel for conversations about geometric vision problems.

This work is supported financially by the Technology Foresight Challenge.

# Contents

# Chapter 1

# Introduction

## 1.1 Introduction

The aim of this report is working out the most suitable way to use a camera as a world measurement device. We want to find out how to use cameras for taking measurements of distances between two points in a scene.

The process of measuring objects is traditionally an engineering task, and as all the engineering tasks it must be accurate and robust. A proper treatment of the sources of error and its propagation during the chain of computations is needed.

We analyse methods for taking measurements of world distances using images, and for accurately predicting the related uncertainty.

The research presented here finds its application in several scientific, commercial and industrial fields.

## 1.2 Why use vision?

Several different types of distance measurement devices have been used in the past. We can distinguish them in two main categories: the active devices and the passive ones. Active devices use to send signals into the environment and get them back. Information related to distances is carried by their phase or echoing time. But unexpected reflections or interferences can affect the output measurement. Passive devices do not suffer this problem.

### 1.2.1 Active devices

**Ultrasonic devices.** Systems for measuring distances have been built up using ultrasonic technology. It is possible to buy relatively cheap ultrasonic devices able to measure the

distance of the operator from another object such as a wall relying on an echoing reflection time measuring system.

Ultrasonic scanners have, for instance, been successfully used in medical imaging such as for 3D reconstruction of the structure of bones [53] or other internal tissues. They have been employed also in robotics problems as autonomous vehicles navigation [23, 24], where accuracy and speed of the localisation system are required.

The main problem of such an approach is that the measure returned is affected by strange and almost unpredictable phenomena like multiple reflections of the ultrasound waves on walls with misleading estimation of the reflection time.

Problems arise also if the operator tries to take planar measurements on the plane opposite to him. It should be possible to improve the capabilities of the devices by placing some sensors on the object we want to measure, but this will increase the cost of the system and the difficulty in the whole operation.

**Laser range finders.** Another possible approach for measuring depths is the use of laser range finders. Those systems are accurate enough but they suffer problems similar to those of ultrasonic devices. Furthermore, laser-based devices are usually quite expensive.

Laser range finders are being applied for 3D shape metric reconstruction of relatively small objects [7]. They have also been applied to solve other common computer vision problems as autonomous navigation [32].

**Structured light.** Other active devices use two or more cameras to acquire different images of an object illuminated by a regular light pattern. Some auxiliary devices, in fact, project a light pattern or a set of patterns onto the object in order to improve its texture and make its features easier to be detected on the images.

Structured light-based algorithms have been used for accurate measurement of surfaces of close range objects which do not show enough texture [40]. Those methods have also been used to capture facial expressions. In [15] coloured dots of light on a regular grid are projected onto the examined face; wrinkles and other deformations of the skin are identified from the deformation of the projected grid.

### 1.2.2 Passive devices

**Tape.** This is the traditional way of measuring distances. It consists in a series of actions which involve the use of metric tapes. This approach can be prone to errors and invasive. Furthermore it is necessary that the operator takes manually all the distances; in the case one ore more are forgotten then a new visit of the place is necessary.

**Cameras.** From all the above discussion the need of a more flexible and friendly measuring device arises. It is straightforward hence to think of a system which creates a digital model of the scene seen in order to take measurements on it, interacting with it.

The idea is:

- An operator takes some images of the object we want to measure.
- A computer processes those images and builds up a 3D metric model of the viewed scene [5].
- The model is at any time queried for measurements via a friendly and easy to use graphical interface (GUI).
- The 3D model is visited from any point of view using virtual reality approaches.

Such a device presents several interesting features.

Firstly the system is very friendly and easy to use, in fact, once the images are taken and the model built an operator can virtually walk through it, watch the scene from different points of view, take measurements querying the software interface, interact with the objects of the scene, print out images of the scene with measurements superimposed.

Furthermore the capture process is very quick, easy and minimal invasive since it involves just a camera to take pictures of the environment we want to measure.

The acquired data are then stored digitally into a disk in order to be reused at any time without the need of going back to the place in the case some new measurements are needed.

The hardware involved is cheap and easy to use, there is no need of new dedicated hardware.

All the work presented in this report finds its motivation in the necessity of exploring the mathematical theory at the basis of such a visual measurement device and the technical problems arising when a working prototype is built.

## 1.3 Possible applications

The theory analysed in this report finds its application in solving many of the engineering and architectural problems which involve measuring objects (buildings, rooms, walls, doors etc...) [2].

Photogrammetry researchers could find it interesting for measuring building dimensions from aerial pictures.

Furthermore useful application can also be found in video compression techniques. If a 3D model of a scene is known then it is possible to eliminate all the unnecessary and

redundant information obtaining a high rate compression of the data space we want to store or transfer.

Other possible uses can be found in surveying, insurance and emergency services.

## 1.4 The AMVIR project

The work presented finds it's main motivation in the development of the AMVIR project (Automated Measurements in VIsual Retailing). The leading idea is replacing conventional measurement tools (tapes and rules) with standard solid state cameras.

The AMVIR project is mainly oriented to the improvement of services offered by companies in the fitted and soft furnishing market sector.

The traditional process of taking measurements using tapes can be prone to errors and unreliable, in fact involving a number of tiring and frustrating movements . Furthermore it can be found too invasive (within possible customers rooms) and too slow.

These problems lead to deficiencies and a general bad quality of the specification and delivery service of fitted and soft furniture companies. A solution for the enhancement of the customer service quality is required which offers: speedy, unobtrusive and accurate data capture where the furnishing has to be installed; systematic and permanent representation of the site; integration into the retailers' related systems such as CAD, estimating, stock control, ordering.

More technically the project is based on results in a branch of mathematics known as *projective geometry* and it is developed in steps, from the easiest to the hardest one.

Firstly the geometry related to two-dimensional projectivities is analysed. A well known theorem states that we can define an unique transformation between any two planes given the locations of four (at least) corresponding points or lines. In our case these planes are the wall and the image plane. Given, say, the edges of the wall we can then transform image measurements, made in pixel units, into real length measurements of the wall.

Secondly the use of 3D-2D projective geometry is required. A single view of a point does not provide enough information to locate the point in 3D, it only constrains the point to lie on a ray containing the camera's optic center and the image point. A second view of the same point from a different viewpoint provides a second ray, and the 3D location is determined as the intersection of the two rays.

Furthermore in this project we believe that a homography-based approach to the complete 3D reconstruction can be used, i.e. using the metric information we retrieve for each planar surface and combining it with that computed for the others we can model the environment as a simple shoe-box room.

We can also model the protrusions (columns, pipes, windows sills) from planar surfaces using two or more views of the same plane.

A further extension to larger-scale variations from the basis box model has to be developed (bay windows, L-shaped rooms) until we are able to make a complete 3D metric representation of the scene captured [6].

## 1.5 Research issues

The basis of the mathematical theory involved in this project is projective geometry, in particular 2D-2D homographic transformations and more general 3D-2D projectivities. While much of the underlying theory exists, there remain a number of research areas which must be addressed to achieve our goal.

Firstly we want to explore different methods for computing scene quantities using only uncalibrated cameras. The camera, in fact, is modelled as a pure projective device (*pinhole model*) whose internal parameters are unknown (focal length, optical center etc...). Mainly we want to explore methods for estimating distances in planar surfaces using a single view and then estimating depths using a multiple-view, uncalibrated framework.

Secondly we need to create a mathematical theory to estimate uncertainties in output measurements and also an adequate test system to check its goodness.

Particular attention has to be paid to this point, in fact, every time we try to take any kind of measurement we make an error which depends on several factors such as: the accuracy of the used device, the error affecting the data acquisition process, the operator, the chain of computations which connect the input data to the final output measurement.

Therefore taking a measurement without knowing the related uncertainty does not make any sense. Once we measure a distance we need to know how accurate it is, i.e. what is the uncertainty range and what is the likelihood that the ground truth falls in that range.

A proper treatment of the sources of error and a modelisation of the accuracy in the measurements is then necessary.

## 1.6 Overview

The report begins with a literature survey of the most important past research conducted on the field of video metrology, 3D reconstruction and error propagation analysis (chapter 2).

In chapter 3 different methods regarding how to accurately estimate plane to plane homographies are investigated.

Chapter 4 proves in different ways that the first order theory is good enough to model the error propagation process for practical situations. It also describes how to estimate the uncertainties in computed distances accounting all error sources. In this chapter we make three novel contributions: first, it is shown in section 4.2 that first order uncertainty analysis is sufficient for typical imaging arrangements. This is achieved by developing the analysis to second order and obtaining a bound on the truncation error. Second, it is shown that the first order analysis is exact for the affine part of the homography, and that an approximation is only involved for the non-linear part. Third, and most significant, in section 4.3.1 an expression is obtained for the covariance of the estimated `H` matrix by using first order matrix perturbation theory [26].

The uncertainty analysis developed here builds on and extends previous analysis of the uncertainty in relations estimated from homogeneous equations, for example homographies [45, 46] and epipolar geometry [13, 17]. It extends these results because it covers the cases both of when the matrix is exactly determined *and* also when over-determined by the world-image correspondences, and furthermore it is not adversely affected when the estimation matrix is nearly singular. This is explained in more detail in section 4.3.1.

The correctness of the uncertainty predictions has been extensively tested analytically, by Monte Carlo simulation (see fig. 4.3a) and by several experiments on real images.

In chapter 5 a device for computing distances of points lying on planes is described. This is a working example of one of the many possible applications of the developed theory.

Finally chapter 6 summarises all the results obtained in this paper and lists possible future extensions. We show possible ways to extend the results we have obtained for plane to plane analysis into a complete three-dimensional reconstruction of the viewed scene.

# Chapter 2

# Related works

## 2.1 Introduction

In this chapter a survey of all the most significant works in the field of 3D reconstruction from 2D images and uncertainty estimation is presented.

Those papers are arranged from mono-view systems to bi and multi-view ones. Parallax-based approaches are also investigated. The works related to the estimation of the accuracy in metric reconstruction are described in the last paragraph.

## 2.2 Using images for measuring

One of the main aims of Computer Vision in the past few years has been taking measurements of the environment just using images and reconstruct it in a digital three-dimensional model.

There are two problems in doing this:

- Reconstructing the scene from images is a hard task because fundamentally it involves solving the "correspondence" problem, i.e. which set of features in the images is genuinely the projection of a feature in world.
- Errors inevitably propagating along the computation chain cause a loss of accuracy in the final 3D structure.

Many researchers in the past have been interested in Euclidean or more simply affine or projective reconstruction, but only a few have investigated the problem of conducting a proper uncertainty analysis in order to assess the accuracy of the final structure.

In this section we analyse different ways of recovering structure using one or more cameras and different methods for analysing the involved uncertainty.

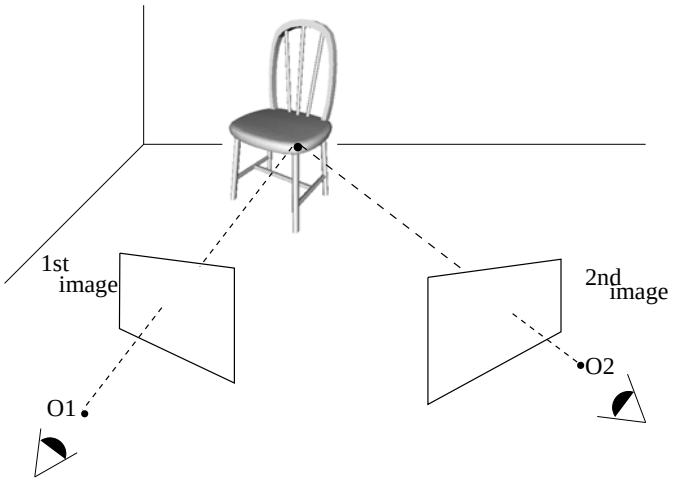



Figure 2.1: Stereo acquisition system scheme

### 2.2.1 Using one view

Only one view does not provide enough information for a complete 3D reconstruction. However some metrical quantities can be computed from the knowledge of some geometrical information such as the relative position of points, lines and planes in the scene.

Several methods, using a single image only, have been developed in order to compute the unknown intrinsic parameters of the camera. This task is known as *camera calibration*.

The work of Tsai [57] is one of the first in the field of camera calibration. From a single image of a known planar calibration grid it estimates the internal parameters of the camera such as the focal length and the image scanning parameters as well as the external position and orientation. An analysis of the accuracy in the parameter estimation is also reported in the paper.

In [17] Faugeras discusses the problem of calibrating a camera in order to retrieve the projection matrix and eventually its internal parameters by using one view. He analyses linear and non linear methods for estimating the 3D-2D projection matrix, the robustness of the estimate and the best location of the reference points. Furthermore the author addresses new methods of self-calibration which eliminate the necessity of using calibration objects.

### 2.2.2 Using two views

The classical methods for 3D reconstruction use stereo vision systems [16] which consists in capturing some images of a scene, taken from two different points of view and then analysing them to estimate the depth of the scene (see fig 2.1). This methodology finds its basis in the human binocular vision system.

The main steps in reconstructing a scene from two stereo images are:

- Matching points on the two images.
- Intersecting the rays in the 3D space for each pair of matching points.

In the following we show a comparison of how a reconstruction can be achieved using calibrated or uncalibrated images.

**Calibrated route.**

The typical route for computing the structure from a pair of calibrated images is as follows.

Once we take two images of a scene with a calibration grid in it, it's possible to compute the 3D-2D projection matrices for each camera [22, 57], and then the epipolar geometry. Some interesting features are extracted in both images, mainly corners or edges of the objects in the scene. Popular approaches to feature detection include the Harris detector [27] for retrieving corners and the Canny detector [8] for edges. Then the process of matching features in the two images is achieved using the computed epipolar constraint. Computing the 3D depth [17] is, now, straightforward. In fact the scene structure is computed via a ray triangulation task, achieved using a robust algorithm such as the one described by Hartley and Sturm [30].

A different method for 3D reconstruction from calibrated stereo systems is described in [62]. Zhongfei Zhang and Allen R. Hanson investigate the problem of recovering the metric structure from the knowledge of two internally calibrated cameras. When the two cameras are internally and non externally calibrated, then the 3D structure can be retrieved mapping one image into the other using homographies inducted by planar surfaces in the images. Using at least four coplanar correspondences the 3D structure can be achieved in Euclidean space up to a scale factor and two real solutions. In order to disambiguate the two solutions a third view is required.

But using a calibrated stereo rig has some disadvantages since the camera internal parameters may not always be constant. We might need to adjust the focal length; or accidental mechanical and thermal events can affect it. Therefore, new algorithms to compute the 3D structure from images with unknown internal camera parameters have been achieved in the past few years (uncalibrated route).

**Uncalibrated route.**

If two cameras used during the acquisition process are uncalibrated (unknown internal parameters), then only a projective structure can be obtained [16]. In fact, we can still match points and lines on the two images, retrieve from those the epipolar geometry and then compute the two projection matrices but only up to a projectivity.

But in order to use a pairs of cameras as distance measuring device we need a complete metric reconstruction. In order to extend a projective structure to an *Euclidean* one some more geometric information about the viewed scene are necessary. Several methods have been developed [19].

In [65] Zisserman et al. describe a method to determine affine and metric calibration for a stereo rig that does not involve calibration objects but simply a single, general motion of a rig with fixed parameters. The internal camera parameters and the relative orientation of the cameras are retrieved robustly and automatically.

Another example can be found in [14] where Devernay and Faugeras show a way for self-calibrate a stereo rig. The stereo rig is weakly calibrated, i.e. the epipolar geometry is known. From one rigid displacement of the rig the three-dimensional Euclidean structure of the scene is uniquely recovered up to a similitude transformation. This method has the advantage that does not require solving the nonlinear Kruppa equations [18, 34].

Related to the previous work is the one from Zhang et al. [63]. It describes a method for the self-calibration of a stereo rig and the metric scene reconstruction using, again, the motion of the entire stereo rig but in this case a simplified camera model is used (the center of the camera is known). Because of the exploitation of information redundancy in the stereo system, this approach yields to a more robust calibration result than only considering a single camera.

### 2.2.3 Using three or more views

Using two views is enough for the reconstruction of the scene, but adding one more image, taken from a third point of view can constrain more the reconstruction problem reducing the uncertainty in the estimated structure. This is particularly true if a line matching process has been used instead of a point matching one (line matching is not possible using only two views).

Furthermore the use of three or more views allows us to check the consistency of the features matched using the first two views.

Faugeras has been one of the first to investigate the problem of using more than two images. In [18] Faugeras et al. present a method for self-calibration of a camera using a sequence of images. The authors, in fact, prove that it is possible to calibrate a camera just pointing it at the environment, selecting points of interest and then tracking them in the image as the camera moves (see also [20]).

Works in this field have been conducted by Shashua in [4, 51, 52], where the geometry involved in a 3D reconstruction task from 2D images is analysed in a very systematic and

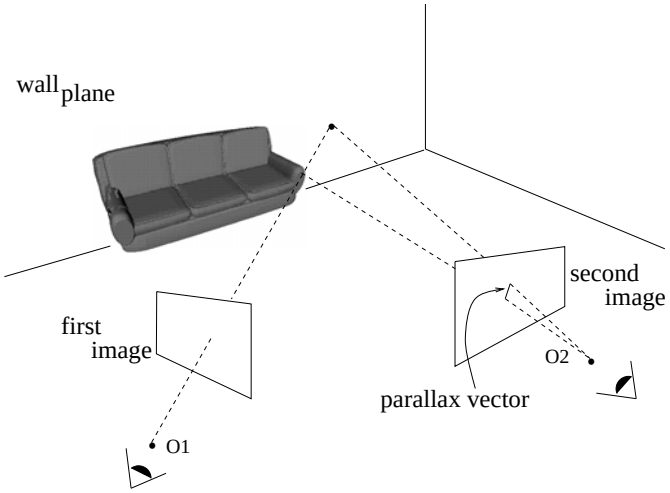



Figure 2.2: Parallax scheme

algebraic way. The concept of "trifocal tensor" which encapsulates the geometric relationship between three views is used. This theory is found useful also for solving the "rendering" problem, i.e. from the knowledge of two images predicting what a third one will be if we consider the camera placed in a new position [21].

The problem of calibrating the camera and estimate the Euclidean structure using three or more uncalibrated images has been investigated by Armstrong et al. in [1, 3]. The authors, using only point matches between images, compute the internal camera parameters and the *scaled* Euclidean structure of the viewed scene. This method has the advantage that it can be applied in active vision tasks, the Euclidean reconstruction can be obtained during normal operation with initially uncalibrated cameras; i.e. the cameras do not need to be calibrated off-line. The trifocal tensor is used here too [29].

In [28] another approach for camera calibration using three or more images is presented by Hartley. This method does not require the knowledge of the camera orientation and position and does not involve the epipolar geometry. In fact the images are taken from the same point of view and therefore the epipolar geometry is useless. The calibration process is based on image correspondences only.

### 2.2.4 Parallax-based approaches

Newer methods for 3D reconstruction make use of the concept of parallax.

From one image we can only say what the projection ray of each image point is, but, in order to get the depth of a point in 3D with respect to a reference plane (or a generic surface) a second view is necessary. The parallax vector seen in the second image (see fig. 2.2) is, in fact, function of the depth of the point from the reference plane.

The parallax approach has been used by Cipolla in [9] as a way to robustly estimate curvature of surfaces. He shows that when we want to recover depth from motion normally we need to know accurately the rotational velocity. But this dependence can be removed if, instead of using raw motion, the difference of the image motions of points (*parallax*) is used. The deletion of the rotational component of the motion field from the process of depth estimation can improve considerably the accuracy of the reconstruction.

Cipolla et al. used the concept of *affine* parallax also in [38] for epipolar geometry computation using the parallax field, in [39] for the estimation of the ego-motion of a mobile robot and in [10] for 3D structure estimation; assessing once more how powerful this approach can be.

In [48] Sawhney presents a formulation for the 3D motion and structure recovery that uses the parallax concept with respect to an arbitrary plane in the environment (real or virtual dominant plane). He shows that if an image coordinate system is warped using a plane-to-plane homography with respect to a reference view, then the residual image motion is dependent only on the epipoles and has a simple relation to the 3D structure.

Directly related to the work of Sawhney is the one of Kumar et al. [35] where the authors show that the parallax vector of a point is directly proportional to its height from a reference plane and inversely proportional to its depth from the camera. Furthermore in [36] they show how scenes can be represented implicitly as a collection of images. In this paper a parallax-based approach is proved to be very useful for independent motion detection, pose estimation and construction of mosaic representations.

As we have already noticed the geometry of the reconstruction problem has been deeply discussed and used by several researchers. In fact, understanding well the geometric properties of the acquisition system can be very useful for computing invariant quantities which increase the robustness and accuracy of the process [17, 43, 64].

The idea of using parallax to compute some useful geometric invariants, has been investigated by Irani and Anandan in [31]. They present an analytical relationship between the image motion of pairs of points and several invariant quantities which can be derived from that relationship. This is shown to be useful for 3D scene structure recovery using more than two images as well as for detection of moving objects and synthesis of new camera views.

### 2.2.5 Investigation of accuracy

In all the above mentioned works we have seen how researchers, using several different methods try to reconstruct a 3D world environment from the analysis of some planar images.

But metric information is of no use without an appropriate estimation of the uncertainties

on the computed distances. In fact, Video Metrology as a precision engineering task, must be treated as such; i.e. in a reliable, efficient and accurate way [17].

Csurka et al. in [13, 60] compute the uncertainty related to the epipolar geometry between two cameras. They use the 8-point algorithm to compute the fundamental matrix and then they work out the related covariance matrix. In order to compute the uncertainty in the fundamental matrix two approaches are used: the first one is statistical and therefore as accurate as required but time consuming, the second is analytical and much simpler.

The problem of conducting an accurate metrology analysis with an estimation of the involved uncertainty has been discussed by Reid and Zisserman in [45, 46]. The authors here analyse an incident which occurred during the 1996 World Cup football final. They, using an uncalibrated video sequence of that event, estimate the position of the ball with respect to the posts and work out whether a goal has occurred or not. Mainly the accuracy in the computation of the ball position is investigated. The structure of the football pitch is reconstructed using geometric information such as the pitch marking, the goal posts position, their parallelism, and no internal camera calibration is involved.

In [11] a tutorial showing general analytical methods for error analysis propagation, based on well known ideas by Faugeras [17] and Kanatani [33], is presented by Clarke. In this paper formulae to compute uncertainties of matrices, vectors and points are discussed. The first order estimation theory is analysed and used, including examples of propagating uncertainty through explicit transformations, transformations defined by implicit functions and even implicit functions with constraints. Some working examples are presented showing how the developed theory can be applied to some well known computer vision problems.

In [37] Kumar and Hanson, using the parallax approach try to estimate robustly the camera 3D location and orientation (*camera pose*) from a matched set of 3D model and 2D landmark features. Robust techniques for pose determination are developed to handle data contamined by fewer than 50.0 % of outliers. The sensitivity of pose determination to incorrect estimates of camera parameters is analysed.

Methods for accurate projective reconstruction are discussed by Mohr et al. in [41]. In this paper the internal camera parameters are unknown and the calibration process is made via the knowledge of some image-world matching points only. Two reconstruction methods are shown and a comparison of their accuracy conducted. The need for a subpixel image point detection is proved to be necessary in order to get an accurate 3D structure.

# Chapter 3

# Estimating the homography

## 3.1 Introduction

Points on a plane are mapped to points on another plane by a *plane to plane homography*, also known as a plane projective transformation. A homography is described by a $3 \times 3$ matrix $\mathtt{H}$.

It is then possible to model the image acquisition process as a perfect projectivity of world plane points onto the image plane, i.e. as a homography mapping function.

Once the world-to-image homography matrix is determined the back-projection of an image point to a point on the world plane is straightforward. The distance between two points on the world plane is simply computed from the Euclidean distance between their back-projected images.

The problem then is to estimate the homography.

## 3.2 The camera model

### 3.2.1 Plane-to-plane camera

The camera model for perspective images of planes, mapping points on the image to points on a world plane is well known [49]. It consists of central projection specialised to planes.

Figure 3.1 shows the imaging process. The notation used is that points on the world plane are represented by upper case vectors, $\mathbf{X}$, and their corresponding images are represented by lower case vectors $\mathbf{x}$. Under perspective projection corresponding points are related by [42, 49]:

$$\mathbf{X} = \mathtt{H}\mathbf{x} \tag{3.1}$$

where $\mathtt{H}$ is a $3 \times 3$ homogeneous matrix, and "=" is equality up to scale. The world and image

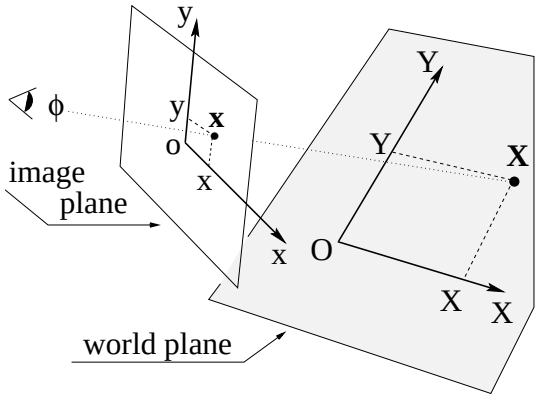



Figure 3.1: Plane-to-plane camera model: a point $\mathbf{X}$ on the world plane is imaged as $\mathbf{x}$. Euclidean coordinates $X-Y$ and $x-y$ are used for the world and image planes, respectively. $\phi$ is the observer.

points are represented by homogeneous 3-vectors as $\mathbf{X} = (X, Y, W)^\top$ and $\mathbf{x} = (x, y, 1)^\top$. The scale of the matrix does not affect the equation, so only the eight degrees of freedom corresponding to the ratio of the matrix elements are significant.

The camera model is completely specified once the matrix is determined. The matrix can be computed from the relative positioning of the two planes and camera centre. However, it can also be computed directly from image-to-world point correspondences. This computation is described in section 3.3.

### 3.2.2 Line-to-line camera

A one-dimensional version of the plane to plane homography is described here. This model is used for the second order analysis in section 4.2. Equation (3.1) reduces to

$$\begin{pmatrix} X \\ 1 \end{pmatrix} = \mathtt{H}_{2\times 2} \begin{pmatrix} x \\ 1 \end{pmatrix}$$

where $\mathtt{H}_{2\times 2}$ is a $2 \times 2$ homography matrix. For the geometry shown in figure 3.2 the matrix is given by

$$\mathtt{H}_{2\times 2} = \begin{pmatrix} \alpha & t \\ \mu & 1 \end{pmatrix}$$

with parameters $\alpha = \frac{d}{f cos^2(w)} - \frac{t}{f} tan(w)$ and $\mu = -\frac{tan(w)}{f}$.

## 3.3 Computing the plane-to-plane homography

From equation (3.1) each image to world point correspondence provides two equations which are linear in the $\mathtt{H}$ matrix elements. For $n$ correspondences we obtain a system of $2n$ equation

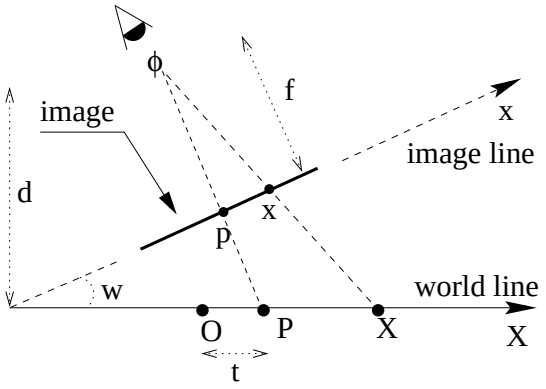



Figure 3.2: One-dimensional camera model: the camera centre is a distance $f$ (the focal length) from the image line. The ray at the principal point $p$ is perpendicular to the image line, and intersects the world line at $P$, with world ordinate $t$. $w$ is the angle between the world and image lines.

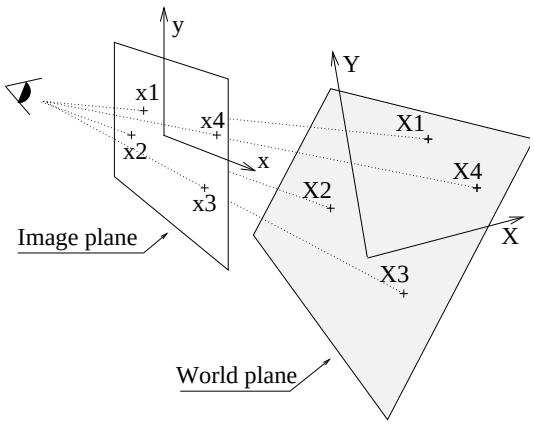



Figure 3.3: Correspondences between computation points

in 8 unknowns. If $n = 4$ (see figure 3.3) then an exact solution is obtained. Otherwise, if $n > 4$, the matrix is over determined, and `H` is estimated by a suitable minimisation scheme.

The covariance of the estimated `H` matrix depends both on the errors in the position of the points used for its computation *and* the estimation method used.

### 3.3.1 Computation methods

There are three standard methods for estimating `H`:

#### Non-homogeneous linear solution.

One of the 9 matrix elements is given a fixed value, usually unity, and the resulting simultaneous equations for the other 8 elements are then solved using a pseudo-inverse. This is the most commonly used method. It has the disadvantage that poor estimates are obtained if the chosen element should actually have the value zero.

**Homogeneous solution.**

The solution is obtained using SVD. This is the method used here and is explained in more detail below. It does not have the disadvantage of the non-homogeneous method.

**Non-linear geometric solution**

In this case the summed Euclidean distances between the measured and a mapped point is minimised.

i.e. minimising the following $2 \times 1$ cost function $C$:

$$C = \sum_{i=1}^{n} \left\{ \left[ \begin{pmatrix} X_i \\ Y_i \end{pmatrix} - \frac{1}{(\mathtt{H}\mathbf{x}_i)_3} \begin{pmatrix} (\mathtt{H}\mathbf{x}_i)_1 \\ (\mathtt{H}\mathbf{x}_i)_2 \end{pmatrix} \right]^2 + \left[ \begin{pmatrix} x_i \\ y_i \end{pmatrix} - \frac{1}{(\mathtt{H}^{-1}\mathbf{X}_i)_3} \begin{pmatrix} (\mathtt{H}^{-1}\mathbf{X}_i)_1 \\ (\mathtt{H}^{-1}\mathbf{X}_i)_2 \end{pmatrix} \right]^2 \right\} \quad (3.2)$$

This returns the maximum likelihood estimation of the $\mathtt{H}$ matrix.

This method has the advantage, over the above two algebraic methods, that the quantity minimised is meaningful and corresponds to the error involved in the measurement (similar minimisations are used in estimate the fundamental matrix and trifocal tensor [56, 61]). There is no closed form solution in this case and a numerical minimisation scheme, such as Levenberg-Marquardt [44], is employed. Usually an initial solution is obtained by method 2, and then "polished" using this method.

### 3.3.2 Homogeneous estimation method

Writing the $\mathtt{H}$ matrix as a 9-vector $\mathbf{h} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33})^\top$ the homogeneous equations (3.1) for $n$ points become $\mathtt{A}\mathbf{h} = \mathbf{0}$, with $\mathtt{A}$ the $2n \times 9$ matrix:

$$\mathtt{A} = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 & -X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1Y_1 & -y_1Y_1 & -Y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -y_2X_2 & -X_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2Y_2 & -y_2Y_2 & -Y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_nX_n & -y_nX_n & -X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_nY_n & -y_nY_n & -Y_n \end{pmatrix} \quad (3.3)$$

We can reduce the problem of computing the $\mathbf{h}$ vector to the constrained minimisation of the following cost function:

$$C = \mathbf{h}^\top \mathtt{A}^\top \mathtt{A}\mathbf{h} \quad (3.4)$$

with the constraint that $|\mathbf{h}| = 1$. The corresponding Lagrange function is :

$$\mathcal{L} = \mathbf{h}^\top \mathtt{A}^\top \mathtt{A}\mathbf{h} - \lambda(\mathbf{h}^\top \mathbf{h} - 1) \tag{3.5}$$

Differentiating it with respect to the $\mathbf{h}$ and setting these derivatives equal to zero we obtain:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{h}} = 2\mathtt{A}^\top \mathtt{A}\mathbf{h} - 2\lambda\mathbf{h} = 0 \tag{3.6}$$

i.e.

$$\mathtt{A}^T \mathtt{A}\mathbf{h} = \lambda\mathbf{h} \tag{3.7}$$

Therefore the solution $\mathbf{h}$ is an unit eigenvector of the matrix $\mathtt{A}^\top\mathtt{A}$ and $\lambda = \mathbf{h}^\top\mathtt{A}^\top\mathtt{A}\mathbf{h}$ is the corresponding eigenvalue.

In order to minimise the $C$ function we should consider only the eigenvector $\mathbf{h}^*$ corresponding to the minimum eigenvalue $\lambda^*$. This eigenvector can be obtained directly from the SVD of $\mathtt{A}$. In the case of $n = 4$, $\mathbf{h}$ is the null-vector of $\mathtt{A}$ and the residuals are zero.

On the other hand the previous is a standard result of linear algebra.

# Chapter 4

# Estimating uncertainties

## 4.1 Introduction

As we have already said (sect.3.2.1) if $\mathbf{x}$ is a point on the image plane then its backprojection onto the world plane is given by the equation (3.1) (in homogeneous coordinates).

However the input image point $\mathbf{x}$ is uncertain because of localisation errors on the image. Furthermore the homography is also uncertain because it is obtained from $n \geq 4$ pairs of uncertain computation points. These two sources of error affect the output $\mathbf{X}$ localisation on the world plane and consequently the distance between two world points.

The goal of this chapter is then to model the uncertainty in the distance between two world points once we know the uncertainty in the localisation of image points and computation matches.

We model the uncertainty in the localisation of an image point $\mathbf{x}$ as a bidimensional gaussian homogeneous and isotropic noise centred on the point itself, with standard deviation $\sigma_x = \sigma_y$. Similarly for world computation points we have $\Sigma_x = \Sigma_y$ (see fig. 4.1). It is not strictly necessary to have such idealised distributions but this has not been found to be a restriction in practice.

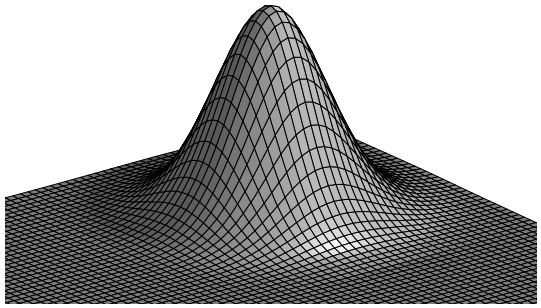


Figure 4.1: Typical isotropic gaussian density probability function. This is used to model the uncertainty in the localisation of image and world-plane points.

If we want to work out how the noise propagates from the input point to the output world-point then we could use two different methods: a statistical one or an analytical one.

**Statistical method.** This consists in using the well known Law of Large Numbers. We compute the homography from a set of $n$ randomly generated matches, generate an image point $\mathbf{x}$ and backproject it onto the world plane using the homography. This produces a world point. We repeat the process $N$ times obtaining a distribution of world points $X_i$ around a mean position $\bar{X}$.

If the number of iteration $N$ is large enough then we can approximate the mean position by:

$$E_N[\mathbf{X}_i] = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{4.1}$$

and its bidimensional $2 \times 2$ covariance matrix by:

$$\Lambda_{\mathbf{X}} = E_N[(\mathbf{X}_i - E_N[\mathbf{X}_i])(\mathbf{X}_i - E_N[\mathbf{X}_i])^\top] \tag{4.2}$$

This method is very costly in terms of CPU time, in fact it returns reliable results only for a very large number of iterations.

**Analytical method.** The analytical method takes into account the fact that the $\mathbf{X}$ point is computed from another random vector $\mathbf{x}$ using a mapping function that is itself uncertain (the homography).

The advantage of using such analytical method is that it gives us a non iterative and therefore fast algorithm. The main disadvantage is that this theory introduces an approximation of the non-linear mapping function. Therefore a further analysis to check how much this approximation affects our final results is needed.

In the past only the first order theory has been used. In the present work we still use the first order approximation but only after having proved that this is good enough for our purposes. The check of goodness of the first order analysis is done by comparing the first order results with the second order ones and with a number of Monte-Carlo simulations, i.e. we use the statistical method described above as a test of goodness of the analytical one.

## 4.2 First and second order uncertainty analysis

In order to avoid unnecessarily complicated algebra the comparison between first and second order analysis is developed for a line-to-line homography. The one-dimensional case illustrates all the ideas involved, and the algebraic expressions are easily interpreted. The

generalisation to $3 \times 3$ matrices is straightforward and does not provide any new insights here.

This simpler case is modelled by a homogeneous $2 \times 2$ matrix as described in section 3.2.2. Under back-projection an image point $x$ maps as

$$x \to X = \frac{h_{11}x + h_{12}}{h_{21}x + h_{22}} = \frac{\alpha x + t}{\mu x + 1}$$

This non-linear mapping (on inhomogeneous coordinates) can be expanded in a Taylor series. Statistical moments of $X$, such as the variance, are then computed in terms of the Taylor coefficients and the moments of $x$ [12, 47]. It is assumed here that the homography is exact (no errors) and the measurement of the image test point $x$ is subject to Gaussian noise with standard deviation $\sigma_x$. The Taylor series is developed about the point's mean position denoted as $\bar{x}$.

### 4.2.1 First order

If the Taylor series is truncated to first order in $(x - \bar{x})$ then the mapping is linearised.

$$X \approx \frac{\alpha\bar{x} + t}{\mu\bar{x} + 1} + \frac{\alpha - \mu t}{(\mu\bar{x} + 1)^2}(x - \bar{x}) \tag{4.3}$$

The mean of $X$ is

$$\bar{X} \approx \frac{\alpha\bar{x} + t}{\mu\bar{x} + 1} \tag{4.4}$$

since the mean is a linear operator.

The variance of $X$ can be shown to be

$$\sigma_X^2 \approx \frac{(\alpha - \mu t)^2}{(\mu\bar{x} + 1)^4}\sigma_x^2 \tag{4.5}$$

### 4.2.2 Second order

Usually only the first order approximation is used for error propagation [11, 13, 17, 45, 46]. Here we extend the Taylor expansion to second order so that the approximation involved in truncating to first order can be bounded. It can be shown that to second order in $(x - \bar{x})$

$$X \approx \frac{\alpha\bar{x} + t}{\mu\bar{x} + 1} + \frac{\alpha - \mu t}{(\mu\bar{x} + 1)^2}(x - \bar{x}) - \frac{\mu(\alpha - \mu t)}{(\mu\bar{x} + 1)^3}(x - \bar{x})^2 \tag{4.6}$$

The mean of $X$ is

$$\bar{X} \approx \frac{\alpha\bar{x}+t}{\mu\bar{x}+1} - \frac{\mu(\alpha-\mu t)}{(\mu\bar{x}+1)^3}{\sigma^2}_x \tag{4.7}$$

and its variance is

$$\sigma^2_X \approx \frac{(\alpha-\mu t)^2}{(\mu\bar{x}+1)^4}\sigma^2_x\left(1+\frac{2\mu^2}{(\mu\bar{x}+1)^2}\sigma^2_x\right) \tag{4.8}$$

### 4.2.3 Comparison

We want now to compare the first order approximation of the perspectivity function with the second order one. We are interested in the error we make in the estimation of the mean $\bar{X}$ and of the variance ${\sigma^2}_X$ of the world point $X$.

**Bias on mean.** In order to find out whether the first order approximation mean is biased or not we define a $bias_X$ quantity as follow:

$$bias_X = \bar{X}^{1^{st}} - \bar{X}^{2^{nd}} = \frac{\mu(\alpha-\mu t)}{(\mu\bar{x}+1)^3}{\sigma^2}_x \tag{4.9}$$

Then the bias is just the difference between the mean of $X$ computed from the first order series truncation and that computed from the second order one. The bias has dimensions of length.

The Lagrange remainder of the Taylor series [55] provides an upper bound on the error on the mean if the series is truncated to first order instead of using the complete expansion.

Using the first order truncation of the Taylor expansion with the Lagrange remainder we obtain:

$$\bar{X} = \frac{\alpha\bar{x}+t}{\mu\bar{x}+1} - \frac{\mu(\alpha-\mu t)}{(\mu\xi+1)^3}{\sigma^2}_x \tag{4.10}$$

with $\xi \in ]x, \bar{x}[$. Therefore the upper bound of the error (absolute value) on the mean of $X$ is

$$Err_{\bar{X}} = \frac{\mu|\alpha-\mu t|}{(\mu\xi+1)^3}{\sigma^2}_x \tag{4.11}$$

Thus we can define a measure, $Mb$, of the bias in the mean between first and second order approximation using the ratio of Lagrange upper bound of the mean $\bar{X}$ to the mean itself.

$$Mb = \frac{\mu|\alpha-\mu t|(\mu\bar{x}+1)}{(\alpha\bar{x}+t)(\mu\xi+1)^3}\sigma^2_x \tag{4.12}$$

**Error in variance.** We now define two measures which can be used to assess the error in the standard deviation (or variance) in truncating to first order.

The first, $Mv_1$, measures the ratio of the second order variance to first order one. Comparing equations (4.5) and (4.8) this ratio is

$$Mv_1 = \frac{2\mu^2}{(\mu\bar{x}+1)^2}\sigma_x^2 \tag{4.13}$$

The second measure, $Mv_2$, is obtained from the Lagrange remainder of the Taylor series [55], as already seen in the previous paragraph.

$$Mv_2 = 2\frac{\mu^2(\mu\bar{x}+1)^4}{(\mu\xi+1)^6}{\sigma^2}_x \tag{4.14}$$

where $\xi \in ]x, \bar{x}[$and we compute the worst case of $\xi = \bar{x} - \sigma_x$ for this bound in the range $\bar{x} \pm \sigma_x$. $Mv_2$ is then the ratio of this truncation error to the first order term.

**Typical results.** The significance of these measures is that they depend only on the homography matrix elements. Thus, once a matrix has been estimated the need for a second order approximation can be immediately assessed. All these measures are dimensionless and it make sense to compare their values to unity.

In typical imaging situations second order terms are not required. In figure 4.2 a graph plot of the measures defined is given in order to visualise their behaviour with respect to variations of the angle $w$. The three linearity measures are computed for the following values: $f = 8.5\,mm$, $d = 5\,m$, $t = 1\,m$, $x = 50\,pixels$, $\sigma_x = 1\,pixel$, and $w$ varies from $0^o$ to $85^o$ (see figure 3.2) Notice that for angle values close to $70^o - 80^o$ the non linearity of the projection function suddenly increases and the first error analysis is near the limit of its usefulness.

The same analysis is reported in the table below for discrete values of the angle $w$.

| w | $30^o$ | $40^o$ | $50^o$ | $60^o$ | $70^o$ | $80^o$ |
|---|---|---|---|---|---|---|
| $Mb$ | $5.9 \times 10^{-6}$ | $1.1 \times 10^{-5}$ | $1.9 \times 10^{-5}$ | $3.9 \times 10^{-5}$ | $9.5 \times 10^{-5}$ | $4.22 \times 10^{-4}$ |
| $Mv_1$ | $1.49 \times 10^{-6}$ | $3.3 \times 10^{-6}$ | $6.9 \times 10^{-6}$ | $1.6 \times 10^{-5}$ | $4.8 \times 10^{-5}$ | $3.7 \times 10^{-4}$ |
| $Mv_2$ | $1.49 \times 10^{-6}$ | $3.3 \times 10^{-6}$ | $6.9 \times 10^{-6}$ | $1.6 \times 10^{-5}$ | $4.6 \times 10^{-5}$ | $3.4 \times 10^{-4}$ |

Notice the very small values of the three measures also for large values of the angle $w$.

### 4.2.4 When is first order exact?

If $\mu = 0$ in equation (4.8) (for instance when $w = 0^o$) then the second order correction is zero and all the three above measures are null. With $\mu = 0$ the homography reduces to an affine transformation, i.e. it is linear. This illustrates the general result that if the homography is

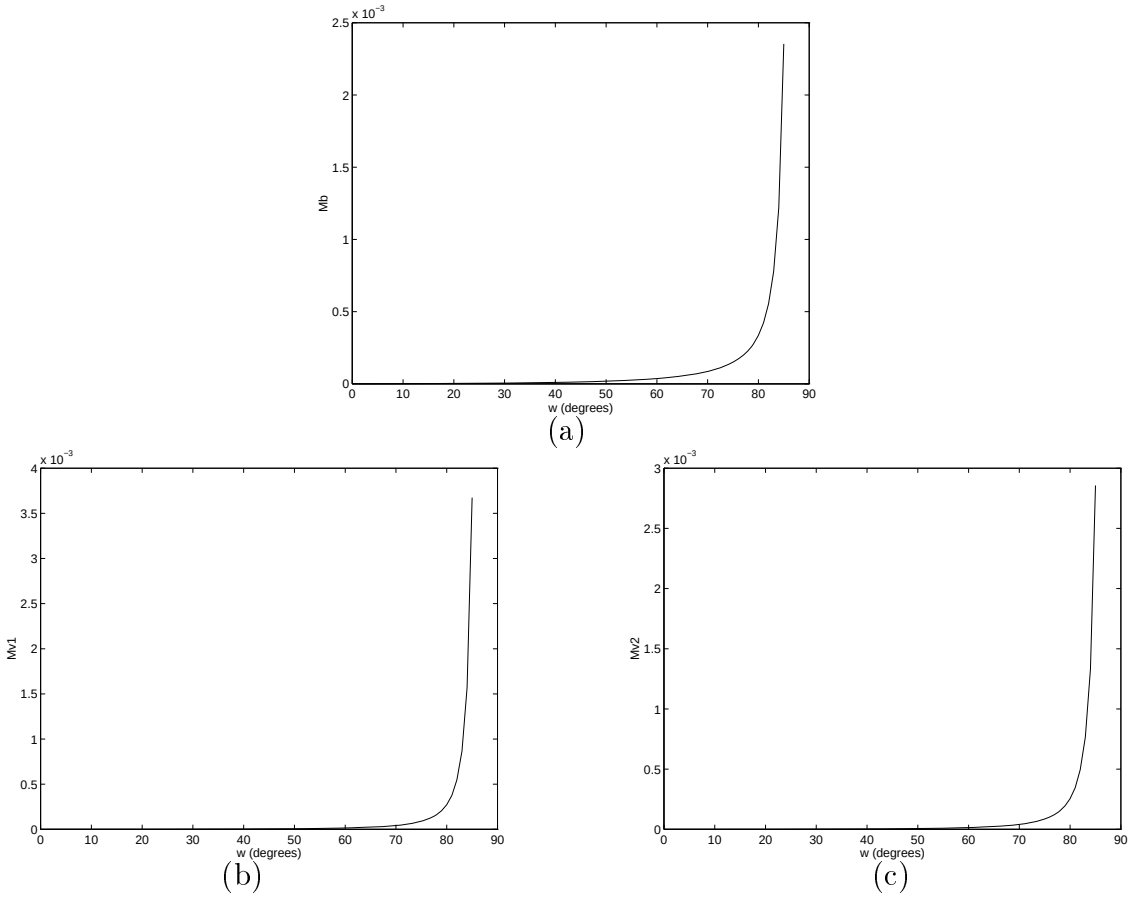



Figure 4.2: Comparison of measures of goodness of first order approximation vs. second order. (a) bias measure $Mb$; (b) first variance measure $Mv_1$; (c) second variance measure $Mv_2$.

affine the first order analysis is exact.

Generally, the H matrix can be decomposed into the product of matrices representing *linear* (affine) and *non linear* (projective) transformations on inhomogeneous coordinates as: $\mathtt{H} = \mathtt{AP}$ where

$$\mathtt{A} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & h_{13} \\ \alpha_{21} & \alpha_{22} & h_{23} \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathtt{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h_{31} & h_{32} & 1 \end{pmatrix}, \tag{4.15}$$

With $\alpha_{ij} = h_{ij} - h_{3j}h_{i3}$ for $i, j = 1, 2$.

If H is purely an affinity (linear on inhomogeneous coordinates) then P is the identity and the first order theory exact.

We have proven that the first order error propagation theory is good enough for typical values in the case of monodimensional camera model.

In the following we extend our first order analysis to the bidimensional camera model in the case of uncertainty affecting also the projective process. Then, in section 4.3.6 we prove its correctness via Monte-Carlo statistical simulations.

## 4.3 Computing uncertainties

There may be errors in the world and image points used to compute the homography, and there may be errors in the image points back-projected to make world measurements. All of these uncertainties must be taken into account in order to compute a cumulative uncertainty for the world measurement.

In this section we analyse the formulae used to compute the uncertainty for measurements under various error situations. The first order analysis is assumed to be sufficient.

We consider all the computation image and world points to be measured with error modelled as a homogeneous, isotropic Gaussian noise process. For the image computation points we define $\sigma_x = \sigma_y = \sigma$, and for the world ones $\Sigma_X = \Sigma_Y = \Sigma$.

We first consider the two sources of error (the uncertainty on the homography and the uncertainty in image point localisation) as operating separately and finally we merge them in order to compute an unique formula embracing both cases.

### 4.3.1 Uncertainty in the homography H, given uncertain computation points

In this section we compute the covariance of the homography $\mathtt{H}$ estimated from $n$ image-world point correspondences.

From section 3.3.2 we seek the eigenvector $\mathbf{h}$ with smallest eigenvalue $\lambda$ of the matrix $\mathtt{A}^\top\mathtt{A}$. But if the $n$ computation matches are affected by a localisation error also the $\mathtt{A}$ matrix will be uncertain and hence the 9-vector $\mathbf{h}$. This vector is then characterised by a $9 \times 9$ covariance matrix $\Lambda_{\mathbf{h}}$ computed as showed in Appendix A.

The above theory has a double advantage over other methods such as [11, 17] which require the inverse of $\mathtt{A}^\top\mathtt{A}$ in order to compute $\Lambda_{\mathbf{h}}$. These methods are poorly conditioned if only four correspondences are used, or if $n > 4$ and the correspondences are (almost) noise-free. In both cases the $\mathtt{A}^\top\mathtt{A}$ matrix is singular and thus not invertible. Because the derivation of expression (A.22) does not involve the inverse, it is well conditioned in both cases.

### 4.3.2 Uncertainty in X, given an uncertain H and exact x.

If $\mathtt{H}$ is a $3\times3$ noisy projection matrix with covariance matrix $\Lambda_{\mathbf{h}}$ computed as above where $\mathbf{h} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33})^\top$ and $\mathbf{x}$ is a non noisy image point, then equation (3.1) can be rewritten as

$$\mathbf{X} = \mathtt{B}\mathbf{h} \tag{4.16}$$

with $\mathbf{X}$ expressed in homogeneous coordinates and

$$\mathtt{B} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x & y & 1 \end{pmatrix} \tag{4.17}$$

For the first order theory we get

$$\Lambda_{\mathbf{X}} = \mathtt{B}\Lambda_{\mathbf{h}}\mathtt{B}^\top \tag{4.18}$$

### 4.3.3 Uncertainty in X, given an exact H and uncertain x.

If the homography is exact but the input image point $\mathbf{x}$ has a non null covariance matrix then the following equation is valid [11]:

$$\Lambda_{\mathbf{X}} = \mathtt{H}\Lambda_{\mathbf{x}}\mathtt{H}^\top \tag{4.19}$$

where $\Lambda_{\mathbf{x}}$ is the $3 \times 3$ covariance matrix of the homogeneous image point $\mathbf{x}$.

Notice that the conversion from the inhomogeneous $2 \times 2$ covariance matrix $\Lambda_{\mathbf{x}}^{2\times 2}$ in the form

$$\Lambda_{\mathbf{x}}^{2\times 2} = \begin{pmatrix} {\sigma^2}_x & \sigma_{xy} \\ \sigma_{xy} & {\sigma^2}_y \end{pmatrix} \tag{4.20}$$

to the homogeneous one is obtained as follows:

$$\Lambda_{\mathbf{x}} = \begin{pmatrix} \Lambda_{\mathbf{x}}^{2\times 2} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{pmatrix} \tag{4.21}$$

The problem now is to merge the two values obtained in the equations (4.18, 4.19) when both the homography and the input point are uncertain. This point is expanded in the next section.

### 4.3.4 Uncertainty in X, given uncertain H and x.

If both the H matrix and the input point $\mathbf{x}$ are noisy then we can just add the results obtained in the previous two sections and therefore:

$$\Lambda_{\mathbf{X}} = \begin{pmatrix} \mathtt{B} & \vdots & \mathtt{H} \end{pmatrix} \begin{pmatrix} \Lambda_{\mathbf{h}} & \vdots & \mathtt{0} \\ \cdots & \cdots & \cdots \\ \mathtt{0} & \vdots & \Lambda_{\mathbf{x}} \end{pmatrix} \begin{pmatrix} \mathtt{B}^\top \\ \cdots \\ \mathtt{H}^\top \end{pmatrix} \tag{4.22}$$

It is easy to convert the $3 \times 3$ homogeneous covariance matrix $\Lambda_{\mathbf{X}}$ into inhomogeneous coordinates. In fact if $\mathbf{X} = (X, Y, W)^\top$ then

$$\boldsymbol{\nabla} f = 1/W^2 \begin{pmatrix} W & 0 & -X \\ 0 & W & -Y \end{pmatrix} \tag{4.23}$$

and

$$\Lambda_{\mathbf{X}}^{2\times 2} = \boldsymbol{\nabla} f \Lambda_{\mathbf{X}} \boldsymbol{\nabla} f^\top \tag{4.24}$$

This matrix defines the uncertainty ellipse related to the localisation of the world-plane point $\mathbf{X}$.

### 4.3.5 Uncertainty in distance between points.

Suppose the two image points $\mathbf{x}_1$ and $\mathbf{x}_2$ are back-projected to $\mathbf{X}_1$ and $\mathbf{X}_2$ on the world plane with covariances $\Lambda_{\mathbf{X}_1}^{2\times 2}$ and $\Lambda_{\mathbf{X}_2}^{2\times 2}$, computed as above. Then, the uncertainty on the

Euclidean world distance $L$ between the two world points is

$$\sigma_L^2 = \boldsymbol{\nabla} g \begin{pmatrix} \Lambda_{\mathbf{X}_1}^{2\times 2} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \Lambda_{\mathbf{X}_2}^{2\times 2} \end{pmatrix} \boldsymbol{\nabla} g^\top \tag{4.25}$$

where $\boldsymbol{\nabla} g = \frac{1}{L}(X_1 - X_2, Y_1 - Y_2, X_2 - X_1, Y_2 - Y_1)$.

### 4.3.6 Monte-Carlo test

In the previous sections we have described a complete theory for computing the ellipse of uncertainty of a world plane point given uncertain homography computation points and uncertain image input points.

This theory relies on the first order truncation of the Taylor series of the non linear projection function.

In this section we want to prove that for typical situations this linear approximation is accurate enough. We have already proven it analytically in the mono-dimensional case and now we use the Monte-Carlo statistical test applied to the bi-dimensional camera model theory.

The *Monte-Carlo* test, in this case, consists in the following steps:

1. generating $n \geq 4$ image computation points (gaussian distributed).
2. generating $n \geq 4$ world computation points (gaussian distributed).
3. computing the related homography.
4. generating an image point $\mathbf{x}_i$ (gaussian distributed).
5. back-projecting the point $\mathbf{x}_i$ onto the world plane point $\mathbf{X}_i$.

The above process is iterated a large number $N$ of times and at the end the uncertainty ellipse of the statistic distribution of the world points $(X_i)$ is computed and compared with the analytically predicted one.

For instance in figure 4.3a a comparison between the covariance ellipse obtained by the first order analysis and the one obtained by a Monte-Carlo evaluation of the actual non-linear homography mapping is shown. Notice that the predicted ellipse and the simulated one are almost overlapping. These figures are obtained using parameters related to a real situation.

Several examples are reported showing the accuracy of the predicted ellipses compared to the simulated ones. A number $N = 100000$ of iterations is involved in each simulation.

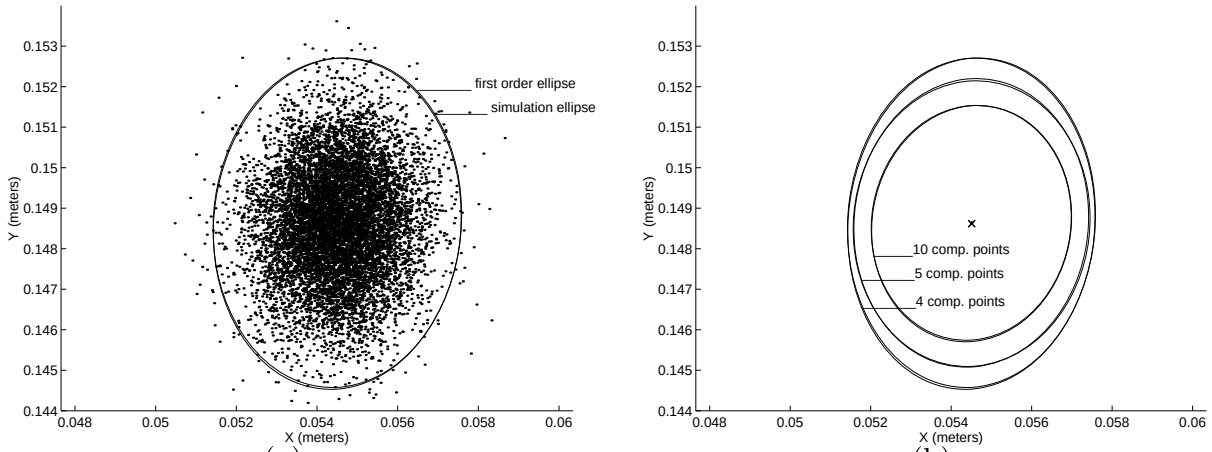



Figure 4.3: (a) `H` is computed from the theory of sections 3.3.2 and 4.3.1 using only 4 computation points. 10000 image points are randomly generated from a Gaussian distribution centered on an image test point $\bar{x}$ and then backprojected onto the world plane. The statistical covariance ellipse of the world points is computed and plotted together with the predicted one. 3 std. dev. are visualised for each uncertainty ellipse. (b) The first order and simulated uncertainty ellipses areas decrease as the number of computation points increase from 4 to 5 to 10 as expected from the theory. 3 standard deviation are visualised (99 percent of probability level).

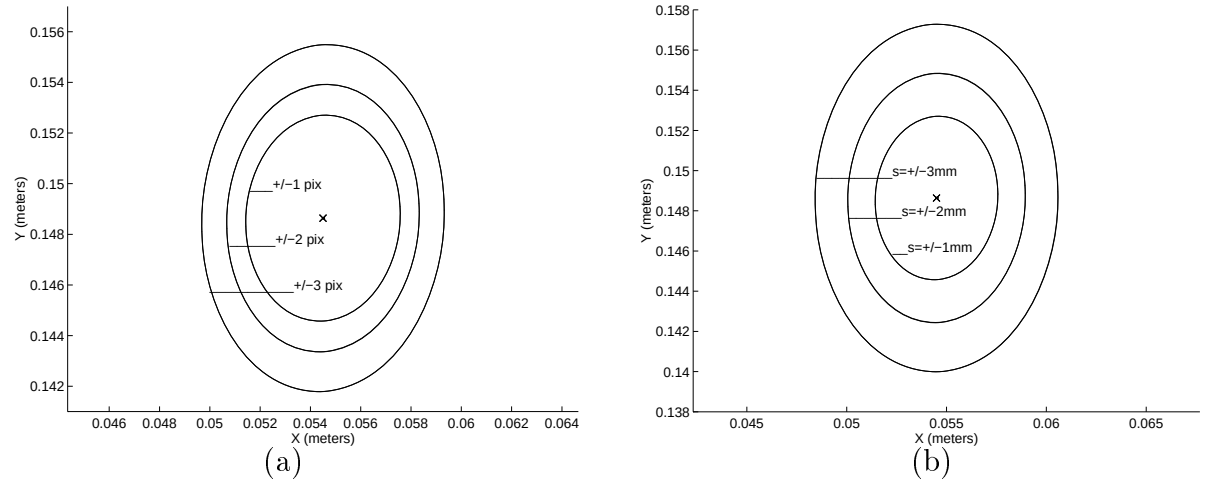



Figure 4.4: (a) Areas of the predicted ellipses and of the simulated ones increase as the uncertainty of the image computation points increase. 3 standard deviation are visualised. (b) Areas of the predicted ellipses and of the simulated ones increase as the uncertainty of the world computation points increase. 3 standard deviations are visualised.

**Increasing the number of computation points.** In figure 4.3b we can see what happens if we increase the number of computation points used to estimate the homography matrix. 6 ellipses are drawn, 3 have been predicted by our theory and other 3 have been obtained from the Monte-Carlo test. We notice that for each of these three cases it is very hard to distinguish between the simulated and the predicted ellipses.

Furthermore, as we expect, increasing the number of computation points increases the accuracy of the `H` matrix and then the accuracy in the final $\mathbf{X}$ point position (smaller ellipses).

**Increasing the image computation points standard deviations.** Figure 4.4a shows how the area of the uncertainty ellipses increases with the uncertainty of the image computation points.

Three predicted ellipses are drawn with the corresponding simulated ones. The standard deviation of the noise in the world points varies from $1\,pixel$ to $3\,pixels$.

**Increasing the world computation points standard deviations.** The increase of the uncertainty ellipses with the uncertainty in the localisation of the world computation points is shown in figure 4.4b.

The three predicted ellipses and the corresponding simulated ones are drawn also in this case. The standard deviation of the noise in the world points varies from $1\,mm$ to $3\,mm$.

In all the previous examples the simulated ellipses are almost exactly overlapping the analytically predicted ones, which justifies our use of the first order theory only in the next chapter.

# Chapter 5

# A plane measuring device

## 5.1 Description

In this section a typical application of the uncertainty theory we have described is shown.

The device shown here is meant to take distance measurements on real world planes such as floors, walls, ceilings, just using images acquired with conventional video cameras.

The entire system is very simple, cheap and does not involve any new or dedicated hardware. In fact the resources used are only:

- A CCD array video camera.
- A personal computer.
- A video acquisition card.

Once the images are taken they are loaded into the computer and ready to be processed by our software.

The entire measurement process can be split in two main stages: the calibration stage and the measuring stage.

### 5.1.1 Calibration stage

First of all, computing the transformation between the viewed plane and the image plane is necessary. This is achieved via the knowledge of a set of world-image point correspondences as already discussed.

Once we know how the $n$ chosen world plane computation points are projected onto the image it is possible to apply the theory of chapters 3.3 and 4 to compute the `H` matrix and its covariance.

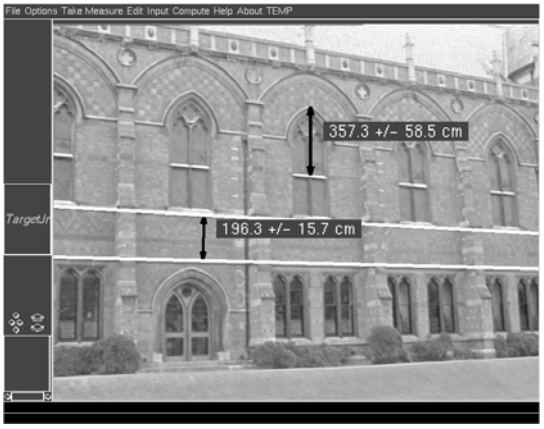



Figure 5.1: Example of the measurement stage which involves an easy-to-use graphical interface.

The choice of computation points is not completely arbitrary; in fact their number and position with respect to the camera have a significant effect on measurement uncertainties.

In section 5.2 some examples are provided showing how the uncertainties vary according to an increase of the number of computation points or to a change in their location.

### 5.1.2 Measurement stage

Once the homography has been computed the operator is allowed to choose two points on the image. They get back-projected onto the world plane and the distance between them computed and shown in the preferred unit (see fig. 5.1). Simultaneously the uncertainty in the distance between them is worked out using the formulae discussed.

Through the knowledge of the homography it is possible to retrieve other useful information such as parallelism of lines lying on a plane. The system allows the user to draw a line, draw the parallel line passing through a chosen point and retrieve the orthogonal distance between the two lines (see fig. 5.1).

The strength of this device is in its easy use; the operator interacts with the system just via a friendly graphic interface both during the calibration stage and the measuring one. The measurement queries are sent to the system via a pointing device like a mouse.

## 5.2 Examples

In this section a number of examples are presented showing the use of the device, and the correctness of the uncertainty analysis on real images.

It is shown that the ground truth measurements always lies within the estimated error bounds. Furthermore, the utility of the analysis is illustrated. The covariance expression

*predicts* uncertainty given the number of image-world computation points and their distribution. It is thus possible to decide where correspondences are required in order to achieve a particular desired measurement accuracy.

### 5.2.1 Accuracy in point localisation

**Varying the number of computation points.** Figure 5.2 shows an indoor scene. Figure 5.2a is the original image and in figures 5.2b-d an *uncertainty ellipses map* is superimposed.

These figures show the localisation uncertainty ellipses for some of the points in the scene and how they change according to a change in the number of the computation points. The `H` matrix is computed using four (fig. 5.2b), six (fig. 5.2c) and eight (fig. 5.2d) computation points. The standard deviations used are $\sigma = 1\, pixel$ for the image points and $\Sigma = 0.5\, cm$ for the world points.

In figure 5.2b the ellipses map is computed using only four computation points and it is easy to notice that the most distant ellipses are the largest. In figure 5.2c six computation points have been used and the ellipses are much smaller, as expected. In figure 5.2d, finally, eight computation points have been used and the uncertainty ellipses are even smaller.

**Varying the location of computations points.** Figure 5.3a is the indoor scene seen in the previous paragraph and in figures 5.3b-d an uncertainty ellipses map is superimposed showing how the ellipses change according to a change in the location of the computation points. The `H` matrix is computed using four computation points in all cases. The standard deviations used are again $\sigma = 1\, pixel$ and $\Sigma = 0.5\, cm$.

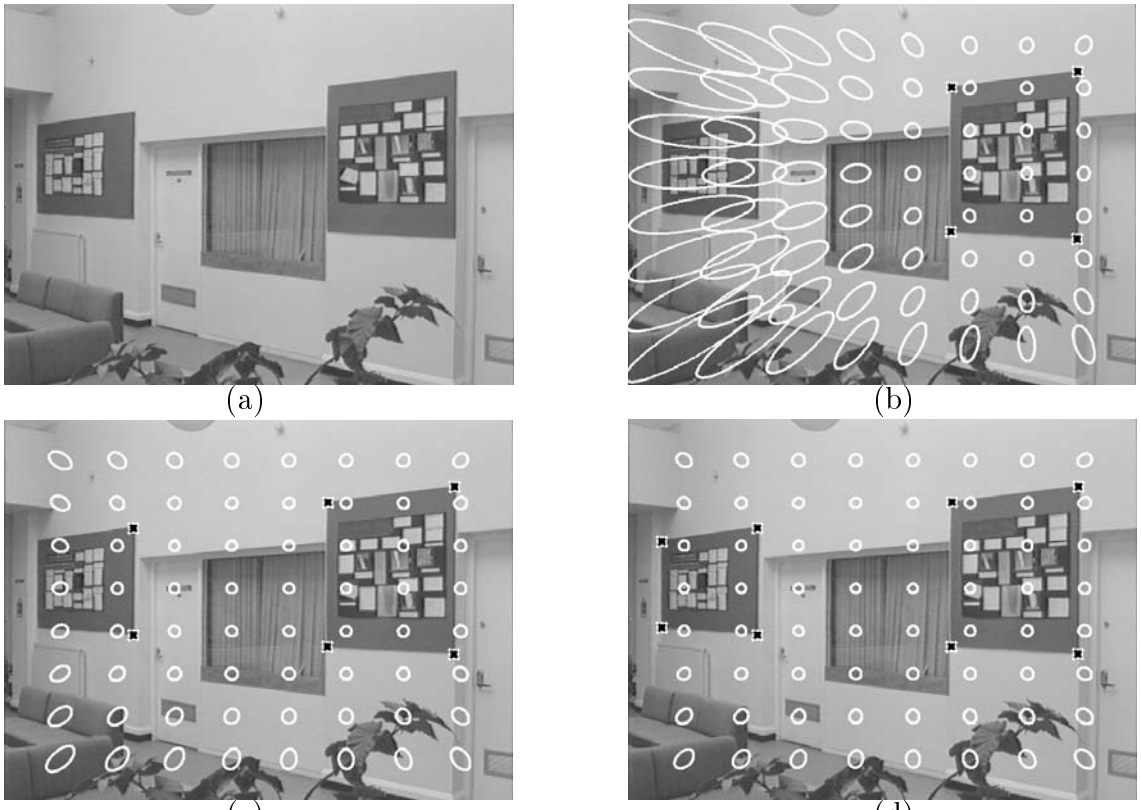
Notice that as the distance of the test points from the computation ones increases the uncertainty ellipse gets larger (less accuracy in localisation). More spatially homogeneous uncertainties are achieved by distributing the computation points across the scene.

From these images we can also deduct that if greater accuracy is required in a certain zone of the image then it's possible to move computation points closer to that zone.

In both images 5.2 and 5.3 the computation points used to estimate `H` are marked by black squares and the uncertainty ellipses are shown at the 6 standard deviation level for clarity.

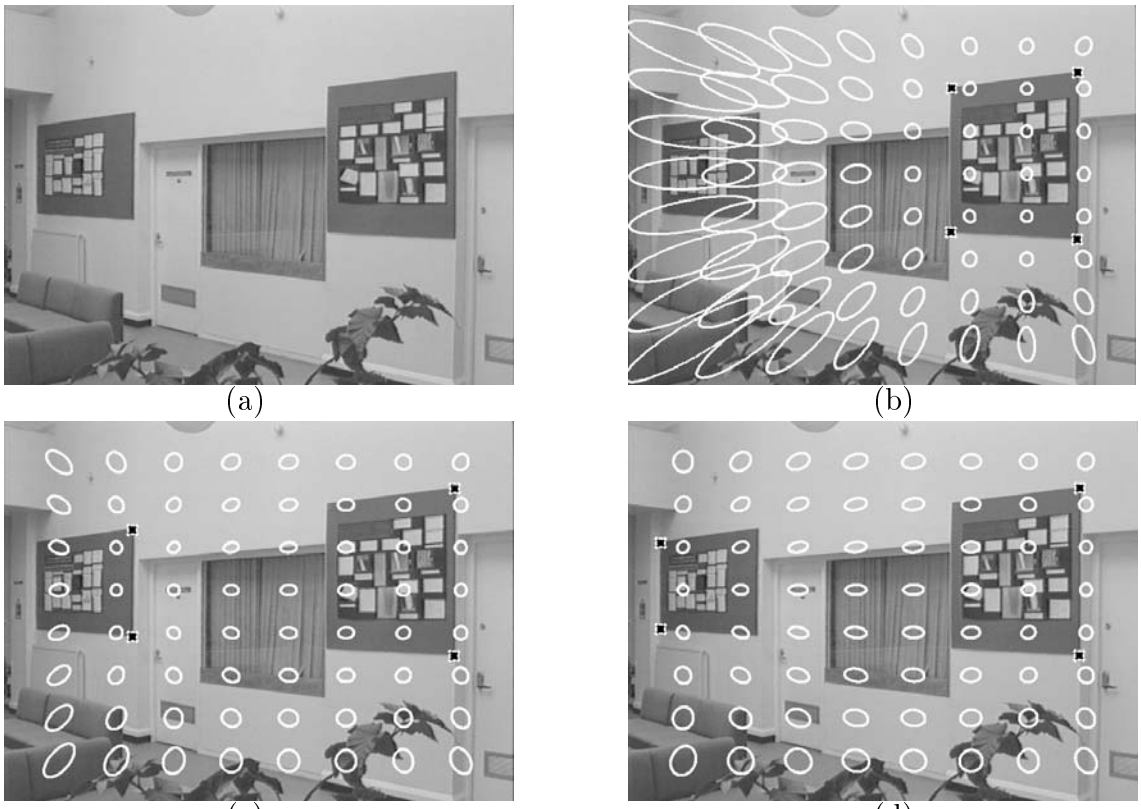
### 5.2.2 Accuracy of distances

A similar analysis is now conducted using images of an outside wall, computing distances instead of point locations.



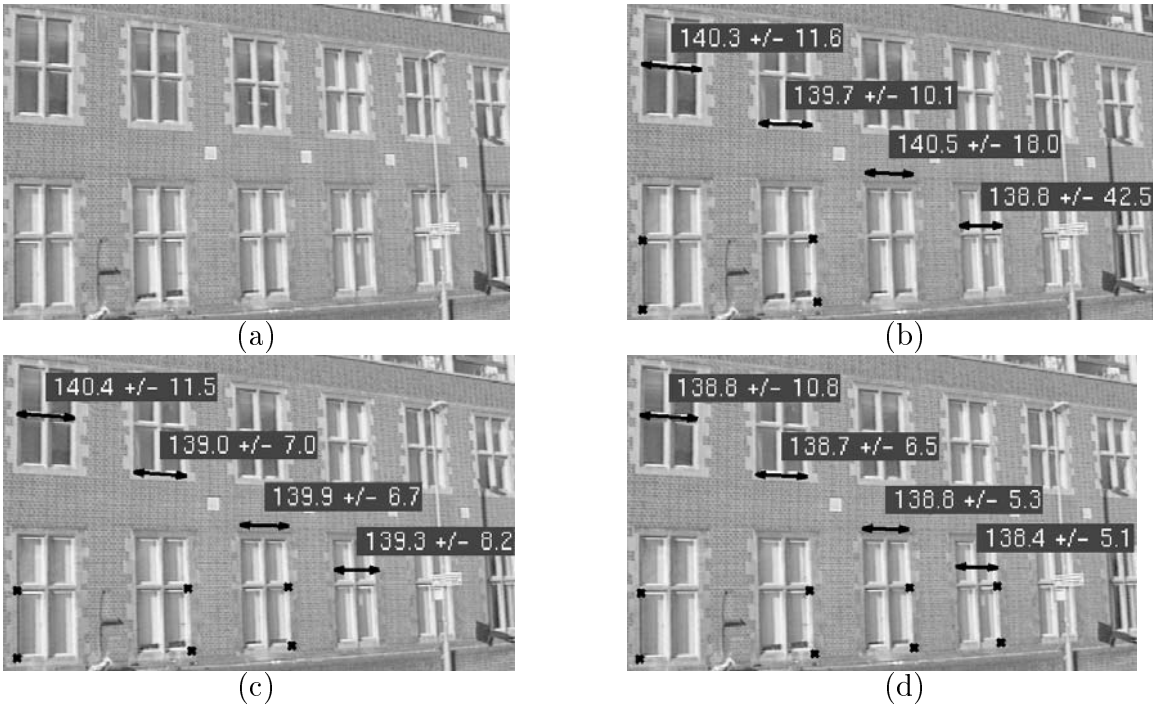
(a) (b) (c) (d)

Figure 5.2: (a) Original image. (b-d) images and ellipses map computed using different numbers of computation points.



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

Figure 5.3: (a) Original image. (b-d) images and ellipses map computed for four computation points in different positions.




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

Figure 5.4: (a) Original image. (b-d) length measurements based on a homography computed from the points marked by black asterisks. The uncertainty bound is $\pm 1$ std. dev. The actual width is 139 $cm$.

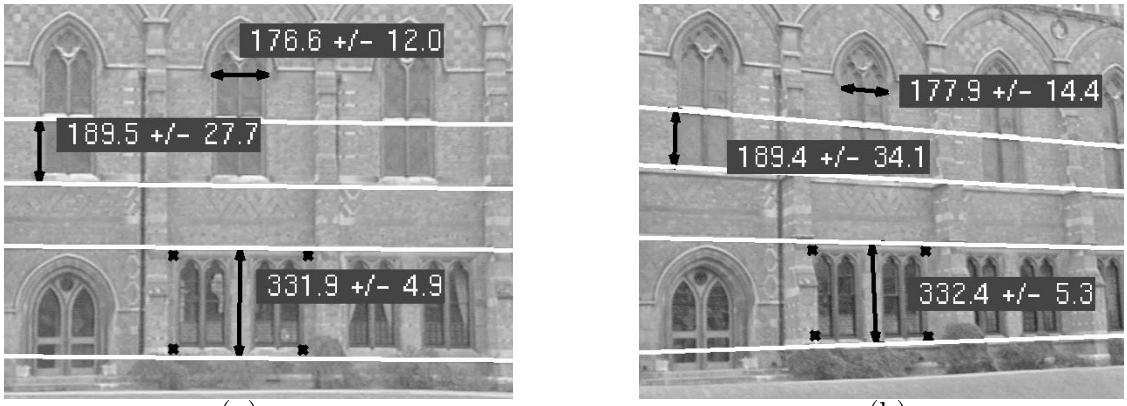
Figure 5.4b-d show length measurements for a homography computed from four, six and eight correspondences. The standard deviations used for the computation are $\sigma = 1\,pixel$ and $\Sigma = 1\,cm$. It is worth noticing that measurements further from the computation point present a larger uncertainty. Again increasing the number of computation points decreases all the uncertainties. Note that all the measurement ranges returned by the system include the actual window width ($139\,cm$).

### 5.2.3 Different views, same computation points.

Figure 5.5 illustrates that the uncertainties also depend on the observer viewpoint.

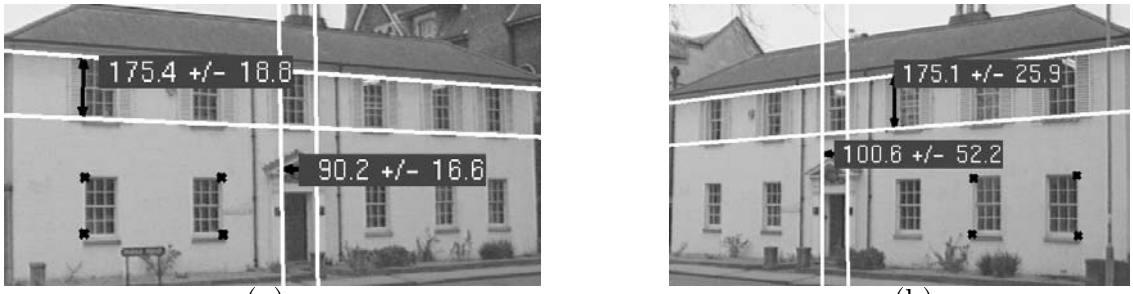
The computation points for both images are in the same zone of the image frame but the second image is affected by a more severe perspective distortion. In both cases the ground truth lies within the predicted measurement range, but this range is larger in the second view.

The figure also illustrates the computation of parallel world lines. Once a line is selected in the image, the one parameter family of lines parallel to it on the world plane is computed from the estimated $\mathbf{H}$, and since one of them is chosen the distance is computed and shown.




(a) (b)

Figure 5.5: Keble College Oxford. The computation points are the same, but the viewpoint distortion is more severe in (b). This is reflected in the larger (1 std. dev.) uncertainties. The actual width of the upper windows is 176 *cm*. Note the computed parallel lines.




(a) (b)

Figure 5.6: Crystallography labs Oxford - The actual height of all the windows is 174 *cm*, the door width is 100 *cm*.

### 5.2.4 Different views, different computation points.

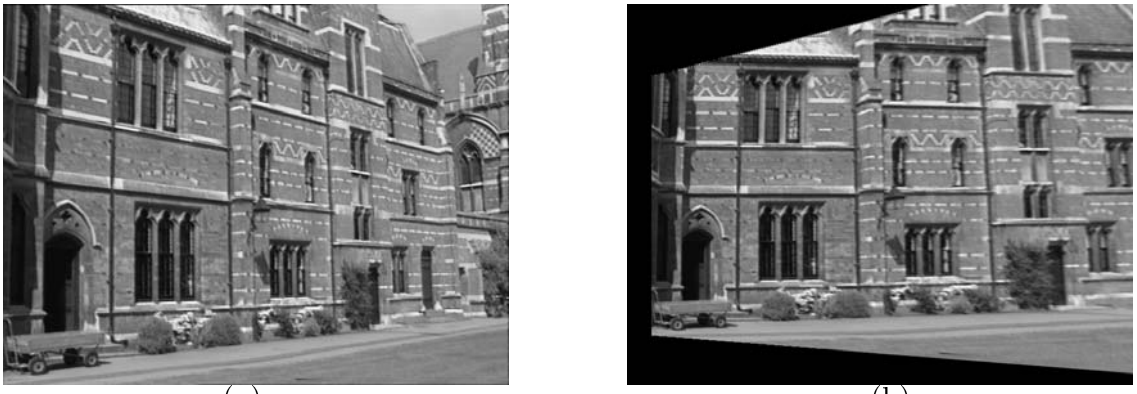
Figure 5.6 shows, again, two different views of a wall. Four different computation points are used in the two images. All the measures are taken between parallel lines and we can notice that notwithstanding the angle between camera and wall plane is large, the parallel lines are correctly computed.

Furthermore, also using different computation points we are able to obtain the same measures for the same objects, the ground truth is always in the uncertainty range returned by the system (see caption). Notice also in this case the computation of the parallel lines.

### 5.2.5 Front-to-parallel warping

Once we have the image of a plane and its homography mapping then it is easy to compute a *front-to-parallel* transformation of the image itself.

From a geometric point of view this operation means the synthesis of a new image, the one seen from a new camera with its acquisition plane parallel to the world plane.



(a) (b)

Figure 5.7: Keble College Oxford - (a) original image, (b) warped image

Figure 5.7 shows an image of a college wall (fig. a); this gets warped onto a front to parallel plane (fig. b). Notice the big window in the bottom left of the wall (near the door) which in the warped image is rectangular.

### 5.2.6 Warping between images

If we know two images of the same wall and the related homographies it is also possible to warp one image onto the other.

An example is shown in figure 5.8 where two pictures of the same wall have been taken from two different points of view, the two homographies image-world plane computed and from these the image1-image2 homography extracted. Finally the first image is warped onto the second one.

All the points which do not lie on the dominant plane of the scene (the building front wall) are mapped into unexpected positions (fig. 5.8c). The comparison between images 5.8b and 5.8c can give us useful information about the 3D structure of the scene.

(a) (b)

(c)

Figure 5.8: Crystallography labs Oxford - (a,b) original images from two different points of view. (c) first image warped onto the second.

# Chapter 6

# Conclusion and future work

## 6.1 Summary

In this report we have developed a first order model of uncertainty prediction which takes account of all the errors involved in estimating the homography. This is used for taking measurements.

This theory overcomes most of the problems normally encountered by others. In fact it works well either when the homography matrix is over determined or when the minimum number of correspondences is used. Furthermore it is not affected by instability problems when the computation points are perfectly non noisy because no matrix inversion is involved.

It is worth noting that although the theory has been developed here for plane to plane homographies, the same approach can be used for line to line and 3-space to 3-space homographies. In particular the matrix perturbation based derivation of covariance [26, 54] can be used where these relations are estimated from an equation of the form $\mathbf{Ah} = \mathbf{0}$. This includes estimating epipolar geometry.

The validity of this work has been demonstrated via several analytical and statistical tests and in a number of practical applications. A software system has been developed which applies the theory described for taking measurements on real images of walls and general flat surfaces. The approach can be applied to any world measurement, lengths and parallelism have been demonstrated, and area and angles can also be shown.

## 6.2 Overall project aims

The main aim of this project is achieving a complete metrical 3D reconstruction of indoor and outdoor scenes using only images. Attention is paid mainly to modeling the uncertainty of the reconstruction to predict and control it in order to make it as small as possible (best

accuracy).

Usually people live in structured environments; planar surfaces such as walls, floor, ceiling are present. This is the kind of environment that will be measured and that is why our work is oriented to a 3D reconstruction which uses a planar surfaces basis plus a proper model of their protrusions (columns, shelves, bay windows etc...).

We also propose to create an adequate 3D model of the reconstructed scene allowing an operator a minimal degree of interaction with the virtual environment.

In order to achieve this goal some new theory has to be developed and some technical aspects addressed.

## 6.3 Approach

### 6.3.1 Work in progress

The system we have described is able to take measurement and compute the related uncertainty only on planar, flat surfaces. But we are already working on some useful extensions of this planar case.

**Mosaicing.** Firstly, if we have a collection of images of the same plane, overlapping in several zones, then we want to be able to combine them in order to create a single, larger image. This is going to be achieved using automatic or semiautomatic feature-matching algorithms between the images.

**Plane plus perturbations.** As shown in this report, by now we are able to make a metric reconstruction of world planes and predict and control the accuracy in the distance measurements.

Now we need to develop a new theory for modelling the protrusions from those planes. In order to achieve this the most suitable way appears to be the use of the *plane-plus-parallax* approach. In fact, we expect that for environments where real planar surfaces dominate, the parallax theory is a robust and accurate way to compute depths in a 3D space.

It is then necessary to develop this theory in all its aspects and achieve the related error propagation analysis. Also in this case we need to estimate how good the first order error analysis is compared to the second order one and to statistical simulations.

A few results regarding the related geometry are shown in the following.

We describe one of the possible ways of using the plane plus parallax theory to compute 3D structure.

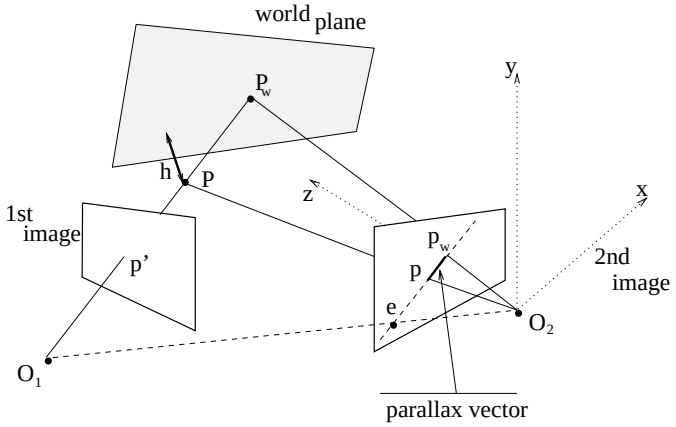



Figure 6.1: Parallax scheme

From [35, 48] we know that the parallax vectors field of an image is an epipolar field. In fact the following equation is valid (see fig. 6.1):

$$\mathbf{p}_w - \mathbf{p} = \frac{hT_z}{dZ}(\mathbf{p}_w - \mathbf{e}) \tag{6.1}$$

where $d$ is the distance of the first camera optical center ($O_1$) from the world plane; $T_z$ is the distance of $O_1$ from the second camera $x-y$ plane; $h$ is the height of the $P$ point with respect to the world plane; $Z$ is the distance of the $P$ point from the second camera $x-y$ plane. $\mathbf{p}_w$ is the 3D vector $\overline{O_2 p_w}$, $\mathbf{p}$ is the vector $\overline{O_2 p}$, and $\mathbf{e}$ is the vector $\overline{O_2 e}$.

Equation (6.1) tells us that all the parallax vectors intersect each other in the epipole of the image. Furthermore we have quantitative information too, which is, the parallax vector of a point in the 3D space is directly proportional to its depth from the reference plane and inversely to its distance from the camera.

Extending eq. (6.1) for two points $P_1$ and $P_2$ we get [31]:

$$\frac{\gamma_2}{\gamma_1} = \frac{\boldsymbol{\mu}_2 \cdot (\Delta \mathbf{p}_w)_\perp}{\boldsymbol{\mu}_1 \cdot (\Delta \mathbf{p}_w)_\perp} \tag{6.2}$$

where $\gamma_i = \frac{h_i}{Z_i}$ is a measure of the 3D shape of the point $P_i$; $\boldsymbol{\mu}_i$ is the parallax vector of the point $P_i$; $\Delta \mathbf{p}_w$ is the vector connecting the "warped" locations of the corresponding second frame points and $\mathbf{v}_\perp$ signifies a vector perpendicular to $\mathbf{v}$.

Notice that $(\Delta \mathbf{p}_w)_\perp$ is easily measurable on the image and it does not involve the knowledge of the epipolar point position. This increases the stability and robustness of the measurement process.

Usually for a stereo acquisition system $Z_i >> h_i$, and therefore we could approximate:

$$\frac{\frac{h_2}{Z_2}}{\frac{h_1}{Z_1}} \approx \frac{h_2}{h_1} \tag{6.3}$$

therefore

$$\frac{h_2}{h_1} \approx \frac{\boldsymbol{\mu}_2 \cdot (\Delta \mathbf{p}_w)_\perp}{\boldsymbol{\mu}_1 \cdot (\Delta \mathbf{p}_w)_\perp} \tag{6.4}$$

The equation (6.4) tells us that from the knowledge of only one point's height (say $h_1$), we can easily compute the heights of all the others (say $h_2$) once we know their parallax vectors $\boldsymbol{\mu}$.

The above description can be one of the ways of constructing a 3D structure based on the knowledge of plane-to-plane homographies.

It is also possible to think of other ways, like: from the knowledge of a minimal set of points whose the relative 3D structure is known, computing projective geometric invariants [42] and using them for retrieving the position of all the others whose parallax vectors are known.

The plane-plus-parallax theory needs to be more deeply investigated with an analysis of the uncertainty involved. This is going to be one of the next steps in the development of the present project.

This will give a number of new contributions to the video metrology community such as the complete modelling of uncertainties in the process of taking measurements of a 3D environment using only images.

**Computation points consistency.** The uncertainty estimation theory is being used also to check the consistency of new image-world correspondences on the basis of the computed homography.

Once we have a set of $n$ computation image-world correspondences we are able to compute the `H` matrix and its uncertainty. If we add a new pair of computation points, say $\hat{\mathbf{x}}$ - $\hat{\mathbf{X}}$, then we can check whether the world point $\hat{\mathbf{X}}$ falls into the uncertainty ellipse related to the image point $\hat{\mathbf{x}}$ (consistency) or not (unconsistency).

A more suitable way of performing this checking task appears to be the use of a RANSAC like [25] algorithm applied to all the correspondences introduced (not just to the latest ones).

### 6.3.2 Future work

**Radial distortion.** When a camera becomes part of a measuring device the accuracy and sensitivity of the camera itself has to be accurately investigated. We need, in fact, to use high

quality lenses and high resolution CCD arrays in order to get the most precise information from the viewed scene.

We will work on the analysis of possible distortions introduced by the camera during the image acquisition process. In all our theory, in fact, the camera has been modelled as a perfect perspective mapping (pinhole camera), but bad quality lenses produce degraded results.

The phenomenon of radial distortion and possible automatic ways for its correction will be investigated.

One way to avoid or at least minimise this problem is the use of software for automatic image correction by retrieving straight edges.

The possibility of using modern digital cameras will also be considered. But a digital choice creates some new problems, in fact the available technology of digital cameras is still not perfect and therefore good quality devices (big and non compressed images, quick loading into computer, zoom options) are still very expensive.

**Geometric minimisation.** The method used to estimate the `H` matrix from $n$ correspondences is the linear, homogeneous one (see sect. 3.3).

We will work out what happens if we use the non linear approach. Our aim is to minimise the Euclidean distances between the image points and the backprojections of the corresponding world ones (see eq. (3.2)).

The minimisation criterion in the case $n > 4$ is the Levenberg-Marquardt [44]. We expect a better accuracy in the estimation of the `H` matrix.

**Complete 3D reconstruction.** The possibility of constructing a real, working device for measuring distances in a 3D space shall be investigated. In fact, in order to build such a device it is necessary to think of how to solve some technical problems.

If we take several images of a scene we firstly need to collect the redundant information carried by each one with the one carried by the others. In particular it is necessary to combine the information related to a single plane to the one related to all other planes in the scene in order to reconstruct the space with its 3D structure.

Once the planar information related to each plane of the scene have been extracted and its protrusions modelled then we need to join all the planes in the 3D space.

The easiest way to do this appears to be the following one: given two planes, from the knowledge of the planar homographies related to each one the intersecting line and the angle between the planes are computed with respect to an absolute coordinates system. These geometric objects are represented in a proper 3D model (for instance using VRML or

Inventor languages).

We must address automatic and semiautomatic methods for choosing the planes and joining them.

The advantages of using a parallax-based plus a further plane-joining approach with respect to traditional three-dimensional reconstruction methods will be investigated. Mainly we want to work out the improvement in terms of robustness and accuracy.

**Further capabilities.** We would also like to investigate further extensions such as:

- Designing software where an operator can interact with the reconstructed 3D model of the scene placing new stylised objects, removing them and changing the point of view.
- Thinking of a way to make our system directly accessible via the Internet for useful networking applications.

## 6.4 Future work plan

We believe that in order to achieve the expected goals we should follow the time table described below.

| **Planned work** | **Predicted time** |
|---|---|
| Continuing the analysis of the plane plus parallax theory for the reconstruction of the perturbation to world planes. | 2 months |
| Analysis of the measurement uncertainties in a plane plus parallax approach. | 2 months |
| System integration allowing multiple overlapping images of a single plane which are finessed to create a single image through use of automatic and semiautomatic feature matching between the images. | 2 months |
| Combining the information gained regarding each wall, recognise and/or assume their relative position and reconstructing the 3D structure. | 6 months |
| Visualisation of objects of given dimensions to be inserted in the scene. | 6 months |
| The last period will be dedicated to writing the thesis and tying loose ends. | 6 months |

# Appendix A

# Computing homography uncertainty

From section 3.3.2 we seek the eigenvector $\mathbf{h}$ with smallest eigenvalue $\lambda$ of the matrix $\mathtt{A}^\top\mathtt{A}$. If the measured points are noise-free, or $n = 4$, then $\mathbf{h} = Null(\mathtt{A})$, and in general we can assume that for $\mathbf{h}$ the residual error $\mathbf{h}^\top\mathtt{A}^\top\mathtt{A}\mathbf{h} = \lambda \approx 0$.

We now use matrix perturbation theory [26, 54, 59] to compute the covariance $\Lambda_{\mathbf{h}}$ of $\mathbf{h}$ based on this zero approximation.

We define

$$\mathbf{z} = \begin{pmatrix} x_1 & y_1 & x_2 & y_2 & x_3 & y_3 & \cdots & x_n & y_n \end{pmatrix}^\top \tag{A.1}$$

the vector of the $2n$ components of the $n$ noisy image computation points, referred to an image coordinate system.

Because of the noise we have:

$$\mathbf{z} = \hat{\mathbf{z}} + \delta\mathbf{z} = \begin{pmatrix} \hat{x}_1 & \hat{y}_1 & \hat{x}_2 & \hat{y}_2 & \hat{x}_3 & \hat{y}_3 & \cdots & \hat{x}_n & \hat{y}_n \end{pmatrix}^\top + \tag{A.2}$$

$$+ \begin{pmatrix} \delta x_1 & \delta y_1 & \delta x_2 & \delta y_2 & \delta x_3 & \delta y_3 & \cdots & \delta x_n & \delta y_n \end{pmatrix}^\top \tag{A.3}$$

where the hat ($\hat{}$) indicates noiseless quantities.

Similarly for the world plane computation points we have that

$$\mathbf{Z} = \begin{pmatrix} X_1 & Y_1 & X_2 & Y_2 & X_3 & Y_3 & \cdots & X_n & Y_n \end{pmatrix}^\top \tag{A.4}$$

is the vector of the $2n$ components of the $n$ noisy world plane computation points referred to a world coordinates system and

$$\mathbf{Z} = \hat{\mathbf{Z}} + \delta\mathbf{Z} = \begin{pmatrix} \hat{X}_1 & \hat{Y}_1 & \hat{X}_2 & \hat{Y}_2 & \hat{X}_3 & \hat{Y}_3 & \cdots & \hat{X}_n & \hat{Y}_n \end{pmatrix}^\top + \tag{A.5}$$

$$+ \begin{pmatrix} \delta X_1 & \delta Y_1 & \delta X_2 & \delta Y_2 & \delta X_3 & \delta Y_3 & \cdots & \delta X_n & \delta Y_n \end{pmatrix}^\top \tag{A.6}$$

We assume once more that the noise is gaussian with null mean. We assume, also, that there is no correlation between the noise of different computation points.

That means: $E(\delta z_i \delta z_j) = \delta_{ij}\sigma_i^2$ and $E(\delta Z_i \delta Z_j) = \delta_{ij}\Sigma_i^2$ where $\delta_{ij}$ is the Kroneker delta

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}, i, j = 1 \cdots 2n \tag{A.7}$$

Furthermore we ipotesize an homogeneus and isotropic noise which means that it doesn't depend on the point position and it's the same for the $x$ and $y$ direction, thus $\sigma_i = \sigma_j, i, j = 1 \cdots 2n$ and $\Sigma_i = \Sigma_j, i, j = 1 \cdots 2n$.

The generic odd row $1 \times 9$ vector of the `A` matrix is

$$\mathbf{a}_{2i-1} = \begin{pmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x_i X_i & -y_i X_i & -X_i \end{pmatrix} \tag{A.8}$$

and the even row $1 \times 9$ vector is

$$\mathbf{a}_{2i} = \begin{pmatrix} 0 & 0 & 0 & x_i & y_i & 1 & -x_i Y_i & -y_i Y_i & -Y_i \end{pmatrix} \tag{A.9}$$

with $i = 1 \cdots n$.

But $\mathbf{a}_{2i-1} = \hat{\mathbf{a}}_{2i-1} + \delta\mathbf{a}_{2i-1}$ and $\mathbf{a}_{2i} = \hat{\mathbf{a}}_{2i} + \delta\mathbf{a}_{2i}$ and therefore

$$\delta\mathbf{a}_{2i-1} = \begin{pmatrix} \delta x_i & \delta y_i & 0 & 0 & 0 & 0 & -(\delta x_i X_i + x_i \delta X_i) & -(\delta y_i X_i + y_i \delta X_i) & -\delta X_i \end{pmatrix} \tag{A.10}$$

and

$$\delta\mathbf{a}_{2i} = \begin{pmatrix} 0 & 0 & 0 & \delta x_i & \delta y_i & 0 & -(\delta x_i Y_i + x_i \delta Y_i) & -(\delta y_i Y_i + y_i \delta Y_i) & -\delta Y_i \end{pmatrix} \tag{A.11}$$

.

It's easy to prove that for each $i, j = 1 \cdots n$:

- $E(\delta\mathbf{a}_{2i-1}^\top \delta\mathbf{a}_{2j-1}) = \delta_{ij}\mathtt{S}_i^o$
- $E(\delta\mathbf{a}_{2i}^\top \delta\mathbf{a}_{2j}) = \delta_{ij}\mathtt{S}_i^e$
- $E(\delta\mathbf{a}_{2i-1}^\top \delta\mathbf{a}_{2j}) = \delta_{ij}\mathtt{S}_i^{oe}$
- $E(\delta\mathbf{a}_{2i}^\top \delta\mathbf{a}_{2j-1}) = \delta_{ij}\mathtt{S}_i^{eo}$

where

$$\mathtt{S}_i^o = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 & -\sigma^2 X_i & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & -\sigma^2 X_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sigma^2 X_i & 0 & 0 & 0 & 0 & 0 & \sigma^2 X_i^2 + x_i^2 \Sigma^2 & x_i y_i \Sigma^2 & x_i \Sigma^2 \\ 0 & -\sigma^2 X_i & 0 & 0 & 0 & 0 & x_i y_i \Sigma^2 & \sigma^2 X_i^2 + y_i^2 \Sigma^2 & y_i \Sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_i \Sigma^2 & y_i \Sigma^2 & \Sigma^2 \end{pmatrix} \tag{A.12}$$

$$\mathtt{S}_i^e = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 & -\sigma^2 Y_i & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & -\sigma^2 Y_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sigma^2 Y_i & 0 & 0 & \sigma^2 Y_i^2 + x_i^2 \Sigma^2 & x_i y_i \Sigma^2 & x_i \Sigma^2 \\ 0 & 0 & 0 & 0 & -\sigma^2 Y_i & 0 & x_i y_i \Sigma^2 & \sigma^2 Y_i^2 + y_i^2 \Sigma^2 & y_i \Sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_i \Sigma^2 & y_i \Sigma^2 & \Sigma^2 \end{pmatrix} \tag{A.13}$$

$$\mathtt{S}_i^{oe} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & -Y_i & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -Y_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -X_i & 0 & 0 & X_i Y_i & 0 & 0 \\ 0 & 0 & 0 & 0 & -X_i & 0 & 0 & X_i Y_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{A.14}$$

$$\mathtt{S}_i^{eo} = \sigma^2 \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -X_i & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -X_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -Y_i & 0 & 0 & 0 & 0 & 0 & X_iY_i & 0 & 0 \\ 0 & -Y_i & 0 & 0 & 0 & 0 & 0 & X_iY_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{A.15}$$

Furthermore if we define the matrix $\mathtt{M} = \mathtt{A}^\top\mathtt{A}$ then $\mathtt{M} = (\hat{\mathtt{A}} + \delta\mathtt{A})^\top(\hat{\mathtt{A}} + \delta\mathtt{A}) = \hat{\mathtt{A}}^\top\hat{\mathtt{A}} + \delta\mathtt{A}^\top\hat{\mathtt{A}} + \hat{\mathtt{A}}^\top\delta\mathtt{A} + \delta\mathtt{A}^\top\delta\mathtt{A}$. Thus $\mathtt{M} = \hat{\mathtt{M}} + \delta\mathtt{M}$ and for the first order approximation we get $\delta\mathtt{M} = \delta\mathtt{A}^\top\hat{\mathtt{A}} + \hat{\mathtt{A}}^\top\delta\mathtt{A}$.

Now let's define $\mathbf{u}_1$ as the eigenvector corresponding to the null eigenvalue of the matrix $\hat{\mathtt{M}}$ (the solution vector $\mathbf{h}$). The others eigensolutions are: $\hat{\mathtt{M}}\hat{\mathbf{u}}_j = \hat{\lambda}_j\hat{\mathbf{u}}_j$ with $j = 2 \cdots 9$.

It has been proved [26, 50] that the variation of the solution is related to the noise of the matrix like in the following formula:

$$\delta\mathbf{u}_1 = -\sum_{k=2}^{9} \frac{\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^\top}{\hat{\lambda}_k}\delta\mathtt{M}\hat{\mathbf{u}}_1 \tag{A.16}$$

but $\delta\mathtt{M}\hat{\mathbf{u}}_1 = \delta\mathtt{A}^\top\hat{\mathtt{A}}\hat{\mathbf{u}}_1 + \hat{\mathtt{A}}^\top\delta\mathtt{A}\hat{\mathbf{u}}_1$ and we know that $\hat{\mathtt{A}}\hat{\mathbf{u}}_1 = 0$ therefore

$$\delta\mathtt{M}\hat{\mathbf{u}}_1 = \hat{\mathtt{A}}^\top\delta\mathtt{A}\hat{\mathbf{u}}_1 \tag{A.17}$$

therefore $\delta\mathbf{u}_1 = \hat{\mathtt{J}}\hat{\mathtt{A}}^\top\delta\mathtt{A}\hat{\mathbf{u}}_1$ where $\hat{\mathtt{J}} = -\sum_{k=2}^{9}\frac{\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^\top}{\hat{\lambda}_k}$.

Therefore:

$$\begin{aligned} \Lambda_{\mathbf{h}} = \Lambda_{\mathbf{u}_1} = E(\delta\mathbf{u}_1\delta\mathbf{u}_1^\top) &= \hat{\mathtt{J}}E\left(\hat{\mathtt{A}}^\top\delta\mathtt{A}\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1^\top\delta\mathtt{A}^\top\hat{\mathtt{A}}\right)\hat{\mathtt{J}}^\top = \\ &= \hat{\mathtt{J}}E\left(\sum_{i=1}^{2n}\hat{\mathbf{a}}_i^\top(\delta\hat{\mathbf{a}}_i\cdot\hat{\mathbf{u}}_1)(\sum_{j=1}^{2n}\hat{\mathbf{a}}_j(\delta\hat{\mathbf{a}}_j\cdot\hat{\mathbf{u}}_1)\right)\hat{\mathtt{J}}^\top = \\ &= \hat{\mathtt{J}}E\left(\sum_{i=1}^{2n}\hat{\mathbf{a}}_i^\top(\sum_{j=1}^{2n}\hat{\mathbf{a}}_j\hat{\mathbf{u}}_1^\top(\delta\hat{\mathbf{a}}_i^\top\delta\hat{\mathbf{a}}_j)\hat{\mathbf{u}}_1)\right)\hat{\mathtt{J}}^\top = \\ &= \hat{\mathtt{J}}\left(\sum_{i=1}^{2n}\hat{\mathbf{a}}_i^\top(\sum_{j=1}^{2n}\hat{\mathbf{a}}_j\hat{\mathbf{u}}_1^\top E(\delta\hat{\mathbf{a}}_i^\top\delta\hat{\mathbf{a}}_j)\hat{\mathbf{u}}_1)\right)\hat{\mathtt{J}}^\top \end{aligned} \tag{A.18}$$

having used that $(\delta\hat{\mathbf{a}}_i\cdot\hat{\mathbf{u}}_1)(\delta\hat{\mathbf{a}}_j\cdot\hat{\mathbf{u}}_1) = \hat{\mathbf{u}}_1^\top(\delta\hat{\mathbf{a}}_i^\top\delta\hat{\mathbf{a}}_j)\hat{\mathbf{u}}_1$.

Now considering that $\hat{\mathtt{J}}$ is a symmetric matrix ($\hat{\mathtt{J}}^\top = \hat{\mathtt{J}}$) the equation (A.18) can be written as

$$\Lambda_{\mathbf{h}} = \hat{\mathtt{J}}\mathtt{S}\hat{\mathtt{J}} \tag{A.19}$$

where $\mathtt{S}$ is a $9 \times 9$ matrix obtained as follows:

$$\begin{aligned}\mathtt{S} = \sum_{i=1}^{n} (&\hat{\mathbf{a}}_{2i-1}^\top \hat{\mathbf{a}}_{2i-1} \hat{\mathbf{u}}_1^\top E(\delta\hat{\mathbf{a}}_{2i-1}^\top \delta\hat{\mathbf{a}}_{2i-1})\hat{\mathbf{u}}_1 + \hat{\mathbf{a}}_{2i}^\top \hat{\mathbf{a}}_{2i} \hat{\mathbf{u}}_1^\top E(\delta\hat{\mathbf{a}}_{2i}^\top \delta\hat{\mathbf{a}}_{2i})\hat{\mathbf{u}}_1 + \\ &\hat{\mathbf{a}}_{2i-1}^\top \hat{\mathbf{a}}_{2i} \hat{\mathbf{u}}_1^\top E(\delta\hat{\mathbf{a}}_{2i-1}^\top \delta\hat{\mathbf{a}}_{2i})\hat{\mathbf{u}}_1 + \hat{\mathbf{a}}_{2i}^\top \hat{\mathbf{a}}_{2i-1} \hat{\mathbf{u}}_1^\top E(\delta\hat{\mathbf{a}}_{2i}^\top \delta\hat{\mathbf{a}}_{2i-1})\hat{\mathbf{u}}_1)\end{aligned} \tag{A.20}$$

Which means

$$\begin{aligned}\mathtt{S} = \sum_{i=1}^{n} (&\hat{\mathbf{a}}_{2i-1}^\top \hat{\mathbf{a}}_{2i-1} \hat{\mathbf{u}}_1^\top \mathtt{S}_i^o \hat{\mathbf{u}}_1 + \hat{\mathbf{a}}_{2i}^\top \hat{\mathbf{a}}_{2i} \hat{\mathbf{u}}_1^\top \mathtt{S}_i^e \hat{\mathbf{u}}_1 + \\ &\hat{\mathbf{a}}_{2i-1}^\top \hat{\mathbf{a}}_{2i} \hat{\mathbf{u}}_1^\top \mathtt{S}_i^{oe} \hat{\mathbf{u}}_1 + \hat{\mathbf{a}}_{2i}^\top \hat{\mathbf{a}}_{2i-1} \hat{\mathbf{u}}_1^\top \mathtt{S}_i^{eo} \hat{\mathbf{u}}_1)\end{aligned} \tag{A.21}$$

Notice that many of the above equations require the true noise-free quantities, which in general is not available. Weng et al. [58] pointed out that if one writes, for instance, $\hat{\mathtt{A}} = \mathtt{A} - \delta\mathtt{A}$ and substitutes this in the relevant equations, the term in $\delta\mathtt{A}$ disappears in the first order expression, allowing $\hat{\mathtt{A}}$ to be simply interchanged with $\mathtt{A}$, and so on.

Finally the $9 \times 9$ covariance matrix $\Lambda_{\mathbf{h}}$ is

$$\Lambda_{\mathbf{h}} = \mathtt{JSJ} \tag{A.22}$$

where $\mathtt{J} = -\sum_{k=2}^{9} \frac{\mathbf{u}_k \mathbf{u}_k^\top}{\lambda_k}$, with $\mathbf{u}_k$ the $k^{th}$ eigenvector of the $\mathtt{A}^\top\mathtt{A}$ matrix and $\lambda_k$ the corresponding eigenvalue. The $9 \times 9$ matrix $\mathtt{S}$ is:

$$\begin{aligned}\mathtt{S} = \sum_{i=1}^{n} (&\mathbf{a}_{2i-1}^\top \mathbf{a}_{2i-1} \mathbf{h}^\top \mathtt{S}_i^o \mathbf{h} + \mathbf{a}_{2i}^\top \mathbf{a}_{2i} \mathbf{h}^\top \mathtt{S}_i^o \mathbf{h} + \\ &\mathbf{a}_{2i-1}^\top \mathbf{a}_{2i} \mathbf{h}^\top \mathtt{S}_i^{oe} \mathbf{h} + \mathbf{a}_{2i}^\top \mathbf{a}_{2i-1} \mathbf{h}^\top \mathtt{S}_i^{eo} \mathbf{h})\end{aligned} \tag{A.23}$$

with $\mathbf{a}_i$ $i^{th}$ row vector of the $\mathtt{A}$ matrix, $n$ is the number of computation points.

# Appendix B

# Seminars and conferences attended

## B.1 Seminars attended

- Graduate Computer Vision Lectures and Exercise Classes organised by Dr. Andrew Zisserman.
- Computer vision reading group by various members of the RRG.
- Various RRG seminars on miscellaneous topics by speakers from inside and outside the group.

## B.2 Conferences attended

- The Fourth European Conference on Computer Vision — ECCV 96, Cambridge, UK

## B.3 Papers submitted

- British Machine Vision Conference, Essex, October 1997.

# Bibliography


[1] Armstrong M. Euclidean structure and camera calibration from image sequences. First year report, Dept. of Engineering Science, University of Oxford, 1994.

[2] Armstrong M., Zisserman A., and Beardsley P. Euclidean reconstruction from uncalibrated images. In *Proc. British Machine Vision Conference*, 1994.

[3] Armstrong M., Zisserman A., and Hartley R. Self-calibration from image triplets. In *Proc. European Conference on Computer Vision*, LNCS 1064/5, pages 3–16. Springer-Verlag, 1996.

[4] Avidan S. and Shashua A. Novel view synthesis in tensor space. In *Proc. Computer Vision and Pattern Recognition*, page Poster session 6, 1997.

[5] Beardsley P., Torr P., and Zisserman A. 3D model acquisition from extended image sequences. In *Proc. European Conference on Computer Vision*, LNCS 1064/1065, pages 683–695. Springer-Verlag, 1996.

[6] Blanc J. and Mohr R. From image sequence to virtual reality. In editor E.P. Baltsavias, editor, *ISPRS Workshop*, pages 144–149, Zurich, Switzerland, March 1995. ISPRS Working Groups.

[7] Callari F. and Ferrie F. Active recognition: Using uncertainty to reduce ambiguity. Technical Report 11, Centre for Intelligent Machine, McGill University, Montreal, Quebec, Canada, September 1995.

[8] Canny J.F. Finding edges and lines in images. Master's thesis, MIT, 1983.

[9] Cipolla R. *Active visual inference of surface shape*. PhD thesis, Dept. of Engineering Science, University of Oxford, 1991.

[10] Cipolla R., Okamoto Y., and Kuno Y. Robust structure from motion using motion parallax. In *Proc. 4th Int'l Conf. on Computer Vision, Berlin*, pages 374–382, Los Alamitos, CA, 1993. IEEE Computer Society Press.

[11] Clarke J.C. First order error propagation: A primer. O.U.E.L. Technical Report, 1996.

[12] Cramer H. *Mathematical Methods of Statistics*. Princeton Univ. Press., 1946.

[13] Csurka G., Zeller C., Zhang Z., and Faugeras O. Characterizing the uncertainty of the fundamental matrix. Technical Report 2560, I.N.R.I.A., France, 1995.

[14] Devernay F. and Faugeras O. From projective to euclidean reconstruction. In *Proc. Computer Vision and Pattern Recognition*, pages 264–269, 1996.

[15] Dubin S., Nissanov J., Zietz S., Schrope B., Morano R., and Hananiah R. Bioengineering approach to non-invasive measurement of body composition. In *Rocky Mountain Bioengineering Symposium*, pages 21–23, April 1994.

[16] Faugeras O. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. European Conference on Computer Vision*, LNCS 588, pages 563–578. Springer-Verlag, 1992.

[17] Faugeras O. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[18] Faugeras O., Luong Q., and Maybank S. Camera self-calibration: Theory and experiments. In *Proc. European Conference on Computer Vision*, LNCS 588, pages 321–334. Springer-Verlag, 1992.

[19] Faugeras O.D. Stratification of three-dimensional vision: projective, affine, and metric representation. *Journal of the Optical Society of America*, A12:465–484, 1995.

[20] Faugeras O.D. and Maybank S.J. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246, 1990.

[21] Faugeras O.D. and Robert L. What can two images tell us about a third one. In J. O. Eckland, editor, *Proc. 3rd European Conf. on Computer Vision, Stockholm*, pages 485–492. Springer-Verlag, 1994.

[22] Faugeras O.D. and Toscani G. The calibration problem for stereo. In *Proc. Computer Vision and Pattern Recognition*, pages 15–20, 1986.

[23] Figueroa F. and Mahajan A. A robust method to determine the coordinates of a wave source for 3-d position sensing. *ASME Journal of Dynamic Systems, Measurements and Control*, 116:505–511, September 1994.

[24] Figueroa F. and Mahajan A. A robust navigation system for autonomous vehicles using ultrasonics. *Control Engineering Practice*, 2(1):49–54, February 1994.

[25] Fischler M.A. and Bolles R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Assoc. Comp. Mach.*, 24(6):381–395, 1981.

[26] Golub G.H. and van Loan C.F. *Matrix Computations.* The John Hopkins University Press, Baltimore, MD, 1989.

[27] Harris C.J. and Stephens M. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conf., Manchester*, pages 147–151, 1988.

[28] Hartley R. Self-calibration from multiple views with a rotating camera. In *Proc. European Conference on Computer Vision*, LNCS 800/801. Springer-Verlag, 1994.

[29] Hartley R. A linear method for reconstruction from lines and points. In *Proc. International Conference on Computer Vision*, pages 882–887, 1995.

[30] Hartley R. and Sturm P. Triangulation. In *Proc. Conference Computer Analysis of Images and Patterns*, Prague, Czech Republic, 1995.

[31] Irani M. and Anandan P. Parallax geometry of pairs of points for 3d scene analysis. In Buxton B. and Cipolla R., editors, *Proc. 4th European Conf. on Computer Vision, LNCS 1064, Cambridge*, pages 17–30. Springer, 1996.

[32] Jarvis A.R. A perspective on range finding techniques for computer vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 5(2):122–139, 1983.

[33] Kanatani K. Statistical optimization for geometric computation: theory and practice. Technical report, AI Lab, Dept of Computer Science, Gunma University, 1995.

[34] Kruppa E. Zur ermittlung eines objektes aus zwei perspektiven mit innerer orientierung. *Sitz.-Ber. Akad. Wiss., Wien, math. naturw. Abt. IIa*, 122:1939–1948, 1913.

[35] Kumar R., Anandan P., and Hanna K. Shape recovery from multiple views: a parallax based approach. In *ARPA Image Understanding Workshop, Monterey, CA,*, 2929 Campus Drive, Suite 260, San Mateo, California 94403o, November 1994. ARPA, Image Understanding, Morgan Kauffmann Publishers.

[36] Kumar R., Anandan P., Irani M., Bergen J., and Hanna K. Representation of scenes from collections of images. In *ICCV Workshop on the Representation of Visual Scenes*, 1995.

[37] Kumar R. and Hanson A.R. Robust methods for estimating pose and a sensitivity analysis. *Computer Vision, Graphics and Image Processing*, vol. 60(3):313–342, 1994.

[38] Lawn J. and Cipolla R. Epipole estimation using affine motion-parallax. In *Proc. 4th British Machine Vision Conf., Guildford*, 1993.

[39] Lawn J.M. and Cipolla R. Robust egomotion estimation from affine motion parallax. In *Proc. 3rd European Conf. on Computer Vision, Stockholm*, pages 205–210. Springer-Verlag, 1994.

[40] Maas H.-G. Robust automatic surface reconstruction with structured light. In *International Archives of Photogrammetry and Remote Sensing*, volume XXIX of *Part B5*, pages 102–107. 1992.

[41] Mohr R., Boufama B., and Brand P. Accurate projective reconstruction. In *Applications of Invariance in Computer Vision*, pages 257–276,, Aores, Portugal, October 1993.

[42] Mundy J. and Zisserman A. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[43] Mundy J., Zisserman A., and Forsyth D. *Applications of Invariance in Computer Vision*. LNCS 825. Springer-Verlag, 1994.

[44] Press W., Flannery B., Teukolsky S., and Vetterling W. *Numerical Recipes in C*. Cambridge University Press, 1988.

[45] Reid I. and Zisserman A. Accurate metrology in uncalibrated video sequences. Technical report, Oxford University, Dept. of Engineering Science, 1996.

[46] Reid I. and Zisserman A. Goal-directed video metrology. In Cipolla R. and Buxton B., editors, *Proc. 4th European Conf. on Computer Vision, LNCS 1065, Cambridge*, volume II, pages 647–658. Springer, april 1996.

[47] Rice J.A. *Mathematical Statistics and Data Analysis.* Wadsworth and Brooks, California, 1988.

[48] Sawhney H.S. Simplifying motion and structure analysis using planar parallax and image warping. In *Proc. Computer Vision and Pattern Recognition*, 1994.

[49] Semple J. and Kneebone G. *Algebraic Projective Geometry.* Oxford University Press, 1979.

[50] Shapiro L.S. and Brady J.M. Rejecting outliers and estimating errors in an orthogonal regression framework. Technical report 1974/93, Department of Engineering Science, University of Oxford, 1993.

[51] Shashua A. Multiple-view geometry and photometry. In Springer Verlag, editor, *ACCV*, Singapore, December 1995.

[52] Shashua A. On photometric issues in 3D visual recognition from a single 2D image. *International Journal of Computer Vision — in press, 1995*, Hebrew University, 1995.

[53] Stern G. and Schindler A. Three-dimensional visualization of bone surfaces from ultrasound scanning. Technical report, A.I.DuPont Institute, 1994.

[54] Stewart G.W. and Sun J. *Matrix Perturbation Theory.* Academic Press Inc., USA, 1990.

[55] Stroud K.A. *Engineering Mathematics.* MacMillan Education, 3rd edition, 1987.

[56] Torr P. and Zisserman A. Robust parameterisation and computation of the trifocal tensor. In *Proc. British Machine Vision Conference*, 1996.

[57] Tsai R. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. Computer Vision and Pattern Recognition*, 1986.

[58] Weng J., Huang T.S., and Ahuja N. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. on Patt. Analysis and Machine Intelligence*, 11(5):451–476, 1989.

[59] Wilkinson J.H. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.

[60] Zhang Z. Determinig the epipolar geometry and its uncertainty: A review. *To appear in the International Journal of Computer Vision*, 1997.

[61] Zhang Z., Deriche R., Faugeras O., and Luong Q. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1995.

[62] Zhang Z. and Hanson A.R. 3d reconstruction based on homography mapping. In *ARPA Image Understanding workshop*, Palm Springs, CA, 1996.

[63] Zhang Z., Luong Q.T., and Faugeras O. Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. *IEEE Trans. Robotics and Automation*, 12(1):103–113, Feb. 1996.

[64] Zisserman A. Lecture series: The geometry of multiple views. Dept. of Engineering Science, University of Oxford, 1995.

[65] Zisserman A., Beardsley P., and Reid I. Metric calibration of a stereo rig. In *IEEE Workshop on Representation of Visual Scenes, Boston*, 1995.