NOTE TO USERS

Université d'Ottawa · University of Ottawa

Université d'Ottawa · University of Ottawa

FACULTÉ DE ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Étienne VINCENT

AUTEUR DE LA THÈSE - AUTHOR OF THESIS

Ph.D. (Computer Science)

GRADE - DEGREE

School of Information, Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT - FACULTY, SCHOOL, DEPARTMENT

TITRE DE LA THÈSE - TITLE OF THE THESIS

On Feature Point Matching, in the Calibrated and Uncalibrated Contexts, Between Widely and Narrowly Separated

R. Laganière

DIRECTEUR DE LA THÈSE - THESIS SUPERVISOR

CO-DIRECTEUR DE LA THÈSE - THESIS CO-SUPERVISOR

EXAMINATEURS DE LA THÈSE - THESIS EXAMINERS

P. Bose

É. Dubois

S. Roy

J. Zhao

J.-M. De Koninck, Ph.D.

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

DEAN OF THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

ON FEATURE POINT MATCHING, IN THE CALIBRATED AND UNCALIBRATED CONTEXTS, BETWEEN WIDELY AND NARROWLY SEPARATED IMAGES

Étienne Vincent

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science

School of Information Technology and Engineering
Ottawa-Carleton Institute for Computer Science
University of Ottawa

April 2004

Your file Votre référence
ISBN: 0-494-01775-9
Our file Notre référence
ISBN: 0-494-01775-9

Contents

List of Tables

List of Figures

Acknowledgments

Firstly, I would like to thank my supervisor, Dr Robert Laganière for his guidance and support.

I would also like to thank my Examiners, Dr Eric Dubois, Dr Prosenjit Bose, Dr Jiying Zhao, and Dr Sébastien Roy for their constructive comments in the evaluation of this work.

I am grateful to the professors of VIVA Lab for the use of their equipment, in particular Dr Eric Dubois and Dr Pierre Payeur.

I would like to thank Dr Gerhard Roth for his help in compiling the literature review of Chapter 5, and Mickaeal Biardeau for his help in gathering the experimental results of Section 3.5.

I am grateful to many students of VIVA Lab, such as Xiaoyong Sun ,Jia Li, Sébastien Gilbert, Samir Arbouche, Rimon Elias, and Daniel Wojtaszek for their occasional assistance and helpful discussion.

I would also like to thank my friends Jérôme Tétreault and Franck Binard for interesting discussions, my math professor Dr Michel Racine who has pushed me to pursue graduate studies, and finally, Man Nghi and my family for their encouragement.

Abstract

In this work, the correspondence problem for feature points between images is investigated. In this context, two important factors greatly influence the choice of a strategy: whether the camera system is calibrated or not, and how large is the separation between viewpoints. This work is divided into four parts, for the four important matching situations generated by these two factors.

In the case of *uncalibrated narrowly separated views*, a framework for empirically evaluating matching constraints is presented. Then, various new and existing constraints are compared.

In the case of *calibrated narrowly separated views*, a new type of feature is introduced, epipolar gradient features. These are then shown to be especially appropriate for matching in the context of quick reconstruction. The features are then matched with a new constraint based on trinocular line transfer.

In the case of *uncalibrated widely separated views*, it is shown how the shape of feature points can be used to recover local perspective deformation between two views, and improve matching results. To this end, a new corner detector that generates the required information is also introduced.

In the case of *calibrated widely separated views*, a more accurate estimate of local perspective deformation is obtained by incorporating the knowledge of the epipolar geometry. An application to fundamental matrix estimation is also introduced.

Chapter 1

Introduction

1.1 Matching

Finding corresponding points between different images of a given scene is a common and fundamental task in computer vision. Image point correspondences are needed in the estimation of depth or other measurements from images, in the construction of 3D scene models for telerobotic or virtual environment applications, in the autocalibration of a system of cameras, in obstacle detection for robot navigation, in indexing the content of an image database, or in detecting motion towards image compression.

There are two important conditions that must be satisfied when seeking corresponding points between images. The first is that several scene points must be visible from all the considered viewpoints. The second assumption is that corresponding image regions be similar in appearance, at least up to a transformation belonging to a restricted transformation space (such as the space of transformations due to changes in illumination, or viewpoint).

When selecting the matching process to be used in a given situation, three important choices must be made:

1. the type of features to be matched;

2. the similarity measure used to compare the features;

3. the general scheme used to apply the chosen similarity measure.

This last step often involves a robust estimator for the camera system's geometry. For example, a typical matching scheme between two uncalibrated images which uses image corner points as the features, and correlation of image intensities as the similarity measure is as follows [90, 107, 127]:

1. Harris feature points (corners) are detected in both images [39].
2. Correlation is applied between the image intensity values in small windows around feature points of the different images, then thresholded to obtain a set of candidate matches. This set typically contains several mismatches.
3. Some constraints are imposed on the candidate matches, in an effort to reject as many mismatches as possible.
4. A robust method, such as RANSAC [26], is used to estimate the epipolar geometry relating the two images from the set of candidate matches.
5. Candidate matches which are incompatible with the estimated epipolar geometry are rejected.
6. The estimated epipolar geometry may be used to guide the search for more correspondences.

Obtaining an estimate of the epipolar geometry in step 4 of the above scheme is often the goal in itself. In this case steps 5 and 6 would be omitted.

The features to be matched may belong to two major categories. They can be small image windows, in which case we speak of *area-based matching*, which is normally used to produce dense disparity maps between images. Alternatively we can select other features such as lines, edges, or feature points. This approach, called *feature-based matching*, only produces sets of sparse correspondences, as matching is limited to the selected features. Nevertheless, if these features are selected wisely, the complexity of

the process is greatly reduced. This work will be mainly concerned with feature-based matching, and specifically with *feature point matching.*

There are also two important categories of similarity measures between feature points. The first one consists of measures based on the correlation of intensity values in neighborhoods of the features. The second one involves describing the features with vectors of characteristics, and then defining a metric between such vectors. This second approach's advantage is that the chosen characteristics can be selected as invariants to some degree of possible deformations between the views. The difficulty is that the greater the class of transformations to which a description is invariant, the less discriminating that description becomes. Correlation-based similarity measures can also be used in ways which are robust to image deformation, if a tool exists to estimate this deformation. Then a feature's neighborhood can be warped prior to correlation, the advantage being that less information is lost than through the use of invariant characterizations.

The choices to be made within a given matching process will be guided by the particular circumstances of the problem. This work is concerned with the four general matching situations which are spanned by the possibilities that the camera system being used may be calibrated or not, and that the difference between the viewpoints may be small or large. In the remainder of this introductory chapter, some basic tools related to the correspondence problem are introduced, and the different matching situations to be investigated are briefly examined.

1.2 Epipolar and Trinocular Geometry

The simple *pinhole model* can be used to represent cameras [42] (see Figure 1.1). A point in space $\mathbf{X}$ is projected onto the image plane π, to a point $\mathbf{x}$, where the ray joining $\mathbf{X}$ and the camera's focal point $\mathbf{c}$ intersects π.

When two cameras look at the same scene, the projection $\mathbf{x}$, on one camera plane π, of an unknown point in space $\mathbf{X}$, can tell us something about where the point will

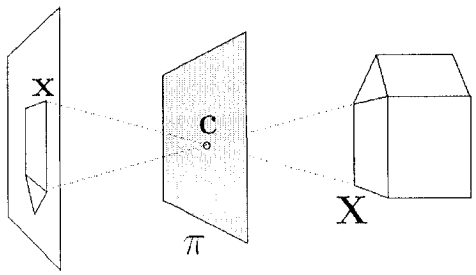


Figure 1.1: Pinhole camera model.

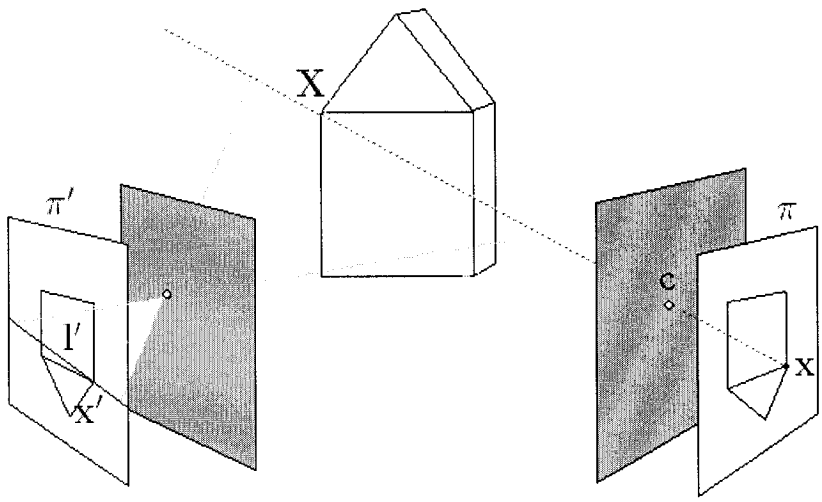


Figure 1.2: Two-view geometry.

land on the other camera plane (see Figure 1.2). More precisely, it is known that a matching point in the second image, must lie on the *epipolar line.*

Since $\mathbf{X}$ must be somewhere along the extension of the ray $\overline{\mathbf{cx}}$, its projection on π' must be on the projection of that ray. This projection, $\mathbf{l}'$ is the epipolar line of $\mathbf{x}$. Thus, points in one image are related to lines in the other through the image pair's *epipolar geometry.*

The relationship between points in one view, and their epipolar line in the other can be represented by the image pair's *fundamental matrix.* This is a 3×3 matrix $\mathbf{F}$, of rank 2, expressed in homogeneous coordinates, and thus having 7 degrees of

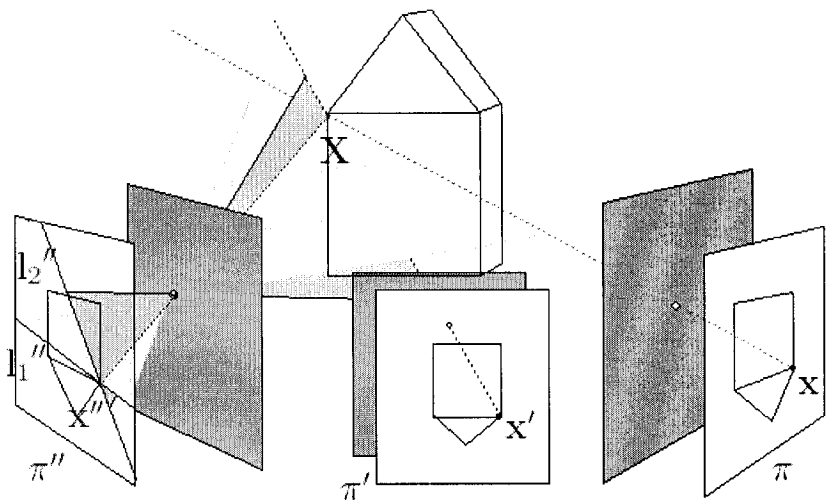


Figure 1.3: Three-view geometry.

freedom (or DOF). A Point $\mathbf{x}$ in the first image, is related to its epipolar lines $\mathbf{l}'$ in the second image, through:

$$\mathbf{F}\mathbf{x} = \mathbf{l}' \tag{1.1}$$

where the point $\mathbf{x}$ is represented with homogeneous coordinates as $(x, y, 1)^\top$, and the epipolar line is the set of points $\mathbf{x}'$, represented in homogeneous coordinates, such that:

$$\mathbf{l}'^\top \mathbf{x}' = 0 \tag{1.2}$$

Notice that Equations 1.1 and 1.2 can be combined to get:

$$\mathbf{x}'^\top \mathbf{F}\mathbf{x} = 0 \tag{1.3}$$

This relation will be useful, to impose a constraint on $\mathbf{F}$, when estimating a fundamental matrix from point correspondences.

When three cameras are used, and a match $(\mathbf{x}, \mathbf{x}')$ is already known between the first two images, the position of the matching point $\mathbf{x}''$ in the third image can be determined exactly. Indeed, the projection of $\mathbf{X}$ on π'' should be at the intersection of the epipolar lines $\mathbf{l}_1''$ and $\mathbf{l}_2''$, of $\mathbf{x}$ and $\mathbf{x}'$ respectively (see Figure 1.3).

This relationship between points in three images is called *trinocular geometry* [100]. In fact, trinocular geometry can determine the position of $\mathbf{x}''$ even when $\mathbf{l}_1''$ and $\mathbf{l}_2''$ are collinear, and thus have no single point of intersection. Trinocular geometry is represented by the *trifocal tensor*, a $3 \times 3 \times 3$ tensor $\mathbf{T}$ with 18 DOF, for which:

$$x_k'' = \sum_{i,j \in \{1,2,3\}} x_i l_j'^{\perp} T_{ijk} \tag{1.4}$$

where $\mathbf{l}'^{\perp}$ is the line going through $\mathbf{x}'$, and perpendicular to $\mathbf{l}'$, the epipolar line of $\mathbf{x}$. This formula can be used to compute the position of $\mathbf{x}''$, but usually, more stable results are obtained using:

$$x_l'' = x_i' \sum_{k=1}^{3} x_k T_{kjl} - x_j' \sum_{k=1}^{3} x_k T_{kil} \tag{1.5}$$

which defines nine trilinear constraints for $i, j \in \{1, 2, 3\}$, four of which are linearly independent. The coordinate of a third image point $\mathbf{x}''$ is then found by solving the over-constrained system of equations.

Epipolar and trinocular geometry do not depend on the structure of the scene, but when this structure is known, more can be said. Another important class of projective relations are *homographies* (projective linear transformations). Two views of a plane are related by a homography. Two images that are related by a pure camera rotation (that is when the focal point is the same between the views) are also related by a homography. These transformations can be described by 3×3 homogeneous and nonsingular matrices with 8 DOF. Thus, if two images are different views of a planar object, or are related by a pure rotation, the corresponding point in the second image for a point $\mathbf{x}$ in the first is given by:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \tag{1.6}$$

where $\mathbf{H}$ is the 3×3 homography matrix.

1.3 Types of Neighborhood Transformations

This section briefly reviews the different possible levels of approximation for the local image deformation induced by changes in viewpoint. These are important, as matching schemes must often be made tolerant to such distortion. Some degree of simplification is needed, as it is generally too complex to describe the effect of occlusions or the transformation of irregular surfaces.

When a scene point is located on a planar surface, the deformation between two views of the region surrounding that point may be described as a homography. Equation (1.6) can be rewritten so that for a point on the plane with homogeneous coordinates $[x \;\; y \;\; w]^\top$, the corresponding point in the other image is given by:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{00} & h_{10} & h_{20} \\ h_{01} & h_{11} & h_{21} \\ h_{02} & h_{12} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} \tag{1.7}$$

where the equality is up to a scale factor and thus the matrix has 8 DOF.

Unfortunately, because of this large number of DOF, it is challenging to estimate the proper homography between two views of a feature point's neighborhood, or to make similarity measures invariant to the set of such transformations. Thus, approximations of homographies are often used. One commonly used subset of all possible homographies is the set of affine transformations:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{00} & a_{10} \\ a_{01} & a_{11} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_0 \\ t_1 \end{bmatrix} \tag{1.8}$$

or if homogeneous coordinates are used:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} a_{00} & a_{10} & t_0 \\ a_{01} & a_{11} & t_1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} \tag{1.9}$$

with equality up to a scale factor. Affinities are described with 6 DOF, and will closely approximate homographic transformations between views of a planar area if the cameras are relatively far from the object being viewed.

A cruder simplification is obtained with a similarity, or homothetic transformations:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_0 \\ t_1 \end{bmatrix} \tag{1.10}$$

This is essentially an affine transformation without stretch and skew. It implies a rotation (of angle θ), translation (t_0, t_1), and a change in scale (of factor s), for a total of 4 DOF. The change in scale is often accounted for separately by matching algorithms. As for the translation, when two features are being compared, it is implied from their respective locations. Therefore, to obtain robustness to similarities, only rotations remain to be dealt with in practice.

A similarity, as an approximation of a general projective transformation, might appear over-simplistic. However, it was shown by Binford [12] that properties which are invariant to similarities are quasi-invariant to perspective deformation. Quasi-invariance, as opposed to true invariance, means that the property varies slowly, with changes in viewpoint. Formally, quasi-invariants are defined, with respect to some set of transformations, as measurements which are unchanged by at least one of the transformations, and that vary slowly when this transformation is perturbed (i.e. their Taylor expansion with respect to the transformation has a first order term which vanishes). According to Schmid [95], although quasi-invariance can be hard to demonstrate, quasi-invariants are often more stable than the true invariants.

1.4 The Calibrated Framework

Epipolar and Trinocular geometry are powerful tools, when used to guide the search for correspondence. Indeed, for a point in a first view, the search for correspondence can be limited to the area along its epipolar line in the second image. And once a point correspondence between two views is known, the matching point in a third view is determined by the trifocal tensor.

Calibration information, such as the knowledge of a fundamental matrix, trifocal tensor, or camera positions and orientations can also be used in rectification, the

process of warping a pair of images in such a way that their epipolar lines become collinear [110]. This makes the search for correspondence easier, as epipolar lines are now horizontal, and under some circumstances, will reduce the perspective deformation between corresponding image regions. Similarly, in Chapter 7, a local perspective deformation estimation method will be proposed which is constrained by the epipolar geometry. Calibration thus allows some image normalization, which may help in evaluating the similarity of possibly corresponding image regions.

Given these advantages, it is clear that when available, calibration information should be used in matching. Calibration may be available when the camera positions and orientations are fixed, at least relative to each other, and an offline calibration step took place prior to matching. However, calibrated matching techniques may also be useful in other circumstances, such as when an approximation of the camera system's geometry is available, or in a later step of a matching scheme, after some calibration information was estimated from an initial match set.

Unfortunately, calibrating a system of cameras is not an easy task. It can be performed directly from images through the identification of corresponding points, but will be very challenging under difficult matching situations, such as widely separated views. Alternatively, a calibration pattern can be used in an offline calibration step. However, this is often impractical, and sometimes impossible, if the cameras are not fixed, or the scene is too large or difficult to access.

For these reasons, uncalibrated matching techniques are also needed. Often, the goal of matching is actually calibration itself; this case obviously requires uncalibrated matching. In other cases, calibration information is simply unavailable. A common example of this would be an image sequence taken by a moving camcorder. Yet in other cases such as indexing in an image database, calibration would simply not make sense, as the images being compared are not necessarily related.

The term *weakly calibrated* is often used to describe the situation where the calibration information was obtained strictly from images, i.e. without physical measures such as the size of some objects or the focal length. In this work, the term *calibrated*

matching will refer to any case where some calibration information is available (either the fundamental matrix, trifocal tensor, or all intrinsic and extrinsic camera parameters). Otherwise, a solution to the correspondence problem will be deemed to belong to the realm of *uncalibrated matching.*

1.5 The Widely Separated Views Framework

Matching can be relatively easy between narrowly separated views. This is the situation where point correspondences are sought between images taken by cameras which were similarly oriented, and furthermore, located close to each other relative to their distance from the objects being viewed. In such a situation, corresponding points will undergo little change in image coordinates. Then, the search for correspondences can be restricted to small image windows around the feature locations. Additionally, there should be little difference in appearance between corresponding image regions, facilitating the comparison between views. There would also be relatively few points which are visible in one view but occluded in another.

However, it is often necessary to obtain point correspondences between more widely separated views. For instance, when the goal is 3D reconstruction from a pair of images, the result will be more accurate if the cameras are further apart. This is because reconstruction is done by finding the intersection of back-projected rays from each image. If the cameras are close, these rays are nearly parallel, and their intersection is not well defined. In other applications, the goal might be to use as few images as possible covering all aspects of an object. Then, widely separated cameras are again preferable. Finally, in other applications, only a limited number of images might be available, and it might again be necessary to match widely separated views. However, no completely satisfactory solution presently exists to the widely separated view matching problem, thus, correspondences have routinely been established by hand in some applications.

The situation where viewpoints are further apart is often referred to as *wide-baseline matching*, or in this work, as *widely separated view matching*. Other situations will be referred to as *narrowly separated view matching*

1.6 Matching Situations

This work consists of an investigation of matching under different circumstances. Firstly, as seen in Section 1.4, matching is divided in two categories according to the availability of calibration information. The first situation is that of calibrated matching, where epipolar or trifocal geometry can be used to guide the search for correspondence, and to develop similarity criteria between possible corresponding features. The other situation is that of uncalibrated matching, where no such information is available.

Secondly, the approach to matching will vary greatly according to the separation between the viewpoints, as described in Section 1.5. Correspondences can be found between narrowly separated views, where local perspective deformation between the views is small. Matching can also be attempted between widely separated views, where more robust constraints must be used.

When considering all possible matching situations, four major categories are found. Consequently, this work is divided into four parts, one for each of the situations, and structured as follow: Part I deals with narrowly separated uncalibrated matching, Part II deals with narrowly separated calibrated matching, Part III deals with widely separated uncalibrated matching, and Part IV deals with widely separated calibrated matching.

1.7 Feature Points

The robust detection of feature points constitutes a fundamental step in image characterization and matching. Feature points usually correspond to particular patterns

exhibiting significant intensity variations in more than one direction. Although they might belong to a large variety of high frequency textures, these special points are often designated as *corners*. One last introductory section is devoted to feature point selection, to put in perspective the different choices of feature point detectors made in later chapters.

Many feature point detectors can be found in the literature, and the results that they produce vary considerably. To evaluate their performance, the desirable properties that a good feature point detector should exhibit must first be established. Different criteria have been considered in past works. These are summarized as follow:

1. *Accuracy*, or the ability to consistently detect a given image pattern, at the exact same location, in spite of minor variability in intensity values, in orientation, or in scale. This property is significant when the detector is tuned to detect well-defined structures, such as specific types of junctions or corners. Accuracy can be assessed by measuring the alignment of extracted features located on a straight line, or, by measuring the distance between a detected junction and the point of intersection of the two lines defining it [55].

2. *Robustness*, or insensitivity to noise. Detection on noisy images can produce false positives, corresponding to noise patterns rather than true feature points. Also, in the presence of noise, some feature points can be lost or not localized properly. Robustness can be evaluated empirically, or theoretically, as done by Smith and Brady in [102], with a probabilistic analysis of a corner model.

3. *Sensitivity* of the detection, that is the ability to detect feature points in low contrast conditions. Most often, a tradeoff exists between sensitivity and robustness.

4. *Stability* of the detection. A detected feature point should continue to be detected after an image undergoes perspective deformation due to a change in viewpoint, a change in illumination conditions, or a zoom. This is essential in

the context of multi-view matching. A good measure of stability is the repeatability of detected features across several views (see Subsection 2.3.3).

5. *Controllability* of the detector. This is mainly determined by the number of parameters which control its behavior, and their relative sensitivity. It is certainly useful to be able to control the number of feature points which will be selected in an image, ideally using a single control parameter. Other parameters might also be used to filter out certain kinds of features, not of interest in a specific application. However, the effect of each parameter should be specific and predictable enough to allow an easy tuning.

6. *Richness* of the information provided about the detected feature points. When a detector returns various characteristics of feature points, and not just a strength measure, the additional information can be exploited in the task to follow. For example, it could be used to classify feature points into categories (e.g. type of junction [83]), or in matching, as a tool to normalize the patterns being matched (as in Section 6.4).

7. *Variability* in the characteristics of detected feature points. This property is referred to as *information content* by Schmid et al. [98]. A high variability ensures that several feature points are detected, regardless of the nature of the image under analysis. Good variability is also critical in matching, where the feature points must be easily distinguishable from each other.

8. *Complexity* of the detector, or the speed at which it identifies corners in an image. In many applications, feature point detection is a preprocessing operation that must be performed at frame rate. The comparison of detection speeds can however be difficult to achieve, since the efficiency of a given feature point detector depends on its implementation.

These are the main quality attributes that should be used to evaluate feature point detectors. Depending on the application, some of these might become more or

less important. However, it is expected that a good feature point detector will behave well with respect to all of these criteria.

In Chapter 2, Harris feature points [39] are used. Harris feature point detectors are the most commonly used ones, as they are fast, and yield good results in the narrowly separated, uncalibrated context. Their use will be further justified through a quick comparison with a SUSAN detector [102].

In Chapter 3, Harris feature points are used again, and some of their properties are studied. The problems associated with these features in their use for quick matching of images from relatively low quality cameras then leads to a new feature point detector described in Chapter 4. This *epipolar gradient* detector is compared to the Harris detector, in the context of calibrated matching.

In Chapter 6, more information is needed on the properties of selected feature points, in order to allow the recovery of local affine transformations between the regions around pairs of corresponding points. The *wedge-based* feature point detector is therefore introduced, which also identifies the shape of the detected corners. It is also seen that this detector compares favorably to the Harris and SUSAN detectors in some respects.

In Chapter 7, again, information of the shape of the detected feature points is needed. There the junction detector JUDOCA [56] was used, as it may achieve more precise results, and an implementation was then readily available at the time when this work was conducted. It should be expected, however, that the wedge-based detector of Chapter 6 would have produced similar results.

Part I

Matching Uncalibrated Narrowly Separated Views

Chapter 2

Uncalibrated Narrowly Separated View Matching

2.1 Introduction

Uncalibrated narrowly separated views is the context in which matching has been studied the most thoroughly. The problem can be considered solved, in the general case, although difficulties may arise for special image pairs with little texture, large illumination changes, or repeated patterns. The work that follows thus focuses on possible small improvements to the general algorithm described in Section 1.1.

In the case where the images to be matched are narrowly separated, two things can be assumed. Firstly, as the distance between the coordinates of corresponding points is expected to be small, the search for correspondence may be limited to a neighborhood of a point's image coordinates. Secondly, a point's image neighborhood should be only mildly deformed between views. This means that simple similarity measures, such as image intensity correlation, are appropriate for comparing points between the views.

The applications of uncalibrated matching include auto-calibration, where matches between views taken by a moving camera are needed to estimate the camera's internal parameters. Alternatively, only the epipolar or trinocular geometry might be sought.

The calibration information can then be used to perform further calibrated matching, or reconstruction. Uncalibrated matching also bears some resemblance to tracking, only then, the difference between images is due to movement of scene objects, rather than changes in viewpoint.

The important qualities of a matching scheme are its speed, its accuracy, and the number and distribution of matches that it produces. With uncalibrated matching in general, achieving high speeds of matching is especially challenging, since points must often be compared with several points in the other view, and since the robust estimation of the camera system's geometry normally involves an expensive iterative process. The choice of feature points, of similarity criteria between these points, and of a robust estimation method for the system's geometry will also influence the number and accuracy of resulting matches.

As will be seen in Chapter 3, the additional knowledge of the epipolar or trinocular geometry greatly simplifies the search for correspondence. Uncalibrated matching is nevertheless essential as a first step towards obtaining this calibration information.

2.1.1 Summary

In this chapter, a general framework for empirically evaluating feature point matching constraints will be described, and then used in comparing different possibilities in each step of a matching scheme.

This work has largely been published. The results of the experiments described in Sections 2.2 to 2.7 were published in:

Etienne Vincent and Robert Laganière,
An Empirical Study of Some Feature Matching Strategies,
in Proc. 15^{th} *International Conference on Vision Interface*, pp. 139-145, Calgary, Canada, May 2002.

A longer journal version appears as:

Etienne Vincent and Robert Laganière,
Matching Feature Points in Stereo Pairs: A Comparative Study of Some Matching Strategies,
in Machine Graphics & Vision, vol. 10, no. 3, pp. 237-259, 2001.

The work presented in Section 2.8 appears as:

Etienne Vincent and Robert Laganière,
Detecting Planar Homographies in an Image Pair,
in Proc. 2nd International Symposium on Image and Signal Processing and Analysis, pp. 182-187, Pula, Croatia, June 2001.

The main contribution of the work described in this chapter is the development and use of an empirical framework for evaluating matching constraints in the uncalibrated narrow baseline situation.

Section 2.2 describes this evaluation scheme for matching methods. Then, the following sections examine different components of matching algorithms. Section 2.3 studies the role of feature point detection. Section 2.4 looks at the way in which correlation is applied. Section 2.5 is concerned with matching constraints that require corresponding features to have similar properties. Section 2.6 examines matching constraints that require matches to have similar displacements as their neighbors. Section 2.7 looks at an application to fundamental matrix estimation, to justify the preceding work. Then, Section 2.8 goes on a tangent to study the special case when large planar regions are found in the images. Finally, Section 2.9 concludes with a more detailed list of the contributions of this chapter.

2.1.2 Previous Work

Many systems which produce accurate correspondences between uncalibrated and narrowly separated views can be found in the literature. The most commonly used

and successful schemes follow steps similar to those described in Section 1.1. These schemes typically attempt to match Harris feature points, and use correlation of image intensity neighborhoods as the comparison metric between them. Zhang et al. [127] describe a system which then uses a least median squares estimator to find the system's epipolar geometry, while Roth and Whitehead [90], as well as Torr and Zisserman [107] use a random sampling consensus scheme to find the system's trinocular geometry. Some surveys describing other matching algorithms can be found in the works of Brown [14], Jones [50], and Zhang [126].

In this chapter, the effectiveness of different matching strategies will be empirically evaluated. Other authors have already empirically compared different matching algorithms and constraints, both in the uncalibrated, and in the calibrated contexts. Scharstein et al. [93] have empirically compared several dense calibrated matching algorithms using ground truth data. Hsieh et al. [46] compare two matching approaches in the context of aerial images with defined evaluation metrics. Koschan [54] proposes a framework for evaluating stereo vision methods, and explicitly identified the constraints used in each of the studied methods. Hu and Ahuja [47] present an algorithm for calibrated matching that exploits many constraints, which are visually validated. Arbouche [4] presents a study of uncalibrated matching constraints that is similar to the one presented here. However, to evaluate constraints, Arbouche counts as valid matches any pair agreeing with the epipolar geometry, and thus misclassifies many mismatches as valid.

It should be pointed out, nevertheless, that many approaches to matching do not follow the general scheme described in Section 1.1. Bolles and Cain [13] describe a Local-Feature-Focus method, which is based on the identification of prominent rare primitive features which are likely to be unique in the images. Goshtasby et al. [32], and Chou and Chen [16] use a template matching approach. Haralick and Shapiro [37] use exhaustive depth-first searches through all possible sets of feature pairs to find possible consistent sets. Finally, Horn and Schunk [45] compute the optical flow using spatio-temporal derivatives. Optical flow computation is essentially dense

uncalibrated matching between images of a dynamic scene taken over time.

Many matching schemes also use iterative *relaxation* approaches. This starts from an initial set of all possible matches. Then, at each iteration, the probability that a pair is a valid match is estimated by evaluating its consistency with the neighboring pairs. Different ways of evaluating these probabilities are used by Faugeras and Berthod [23], Hummel and Zucker [49], and Kitchen and Rosenfeld [52]. Barnard and Thompson [5] also use a relaxation scheme, but first use some constraints on the set of possible matches to eliminate mismatches before using the costly iterative process. In this chapter, many matching constraints are evaluated, but iterative schemes were avoided, because of their higher complexity. Nevertheless, the consistency constraints used by relaxation schemes can be studied outside these processes, as done in Section 2.6.

2.2 Validating Point Correspondences

2.2.1 Using Ground Truth

The feature point matching problem consists of finding feature points, in different images, that correspond to the same scene elements. These matches are obtained from a large set of possible feature point pairs. To evaluate and compare various matching strategies, a *ground truth* set will be used, consisting of all possible good matches between fixed sets of previously selected feature points. It will then be possible to compare the result of the application of given matching constraints with this exact solution.

In the experiments to follow, the six image pairs shown in Figures 2.1 to 2.6 are used. The camera displacement in the image pairs *kitchen* and *lab* consist mainly of a translation (smaller in the case of *lab*), while in *church* it involves some rotation and in *building* the rotation is somewhat larger. In *house* the change in viewpoint is a zoom. Finally, in *objects* there is a significant change in lighting conditions between the views.

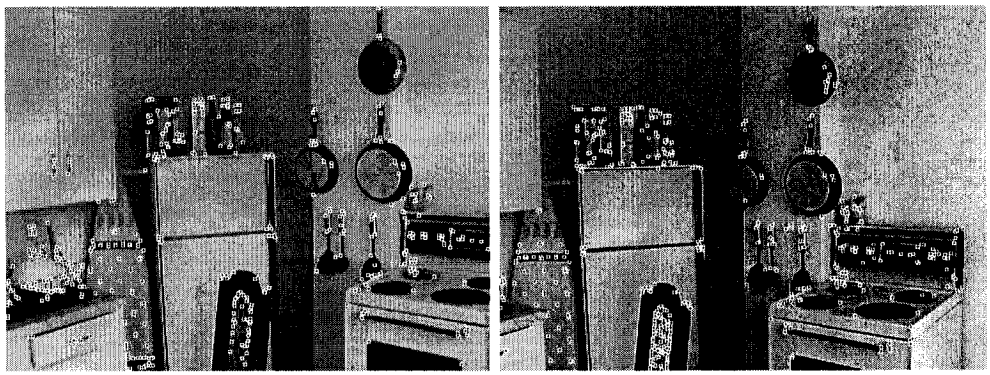


Figure 2.1: The *kitchen* image pair, with extracted feature points.

Figure 2.2: The *building* image pair, with extracted feature points.

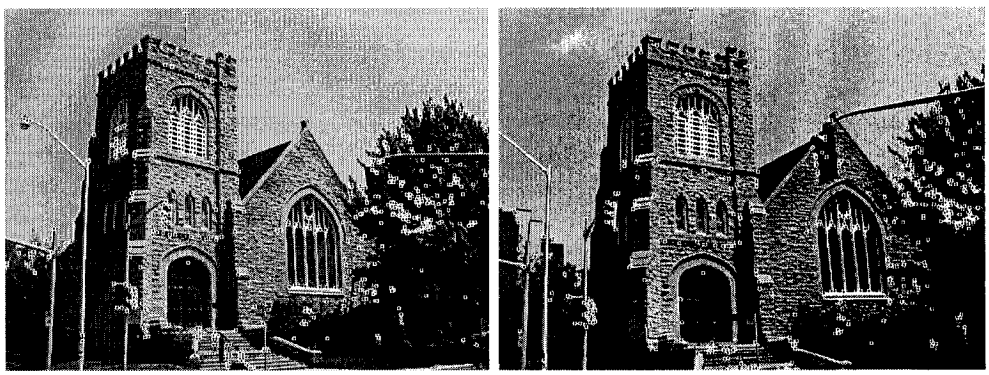


Figure 2.3: The *church* image pair, with extracted feature points.

Figure 2.4: The *lab* image pair, with extracted feature points, courtesy of the University of Tsukuba.

Figure 2.5: The *house* image pair, with extracted feature points.

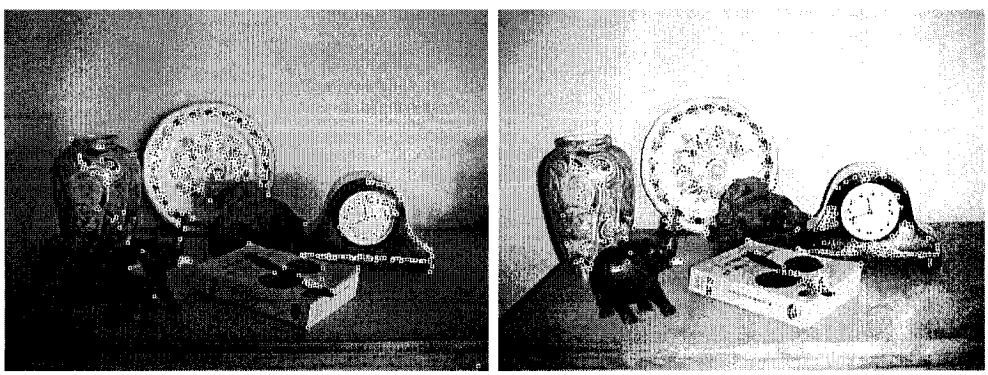


Figure 2.6: The *objects* image pair, with extracted feature points.

Approximately 500 Harris feature points were selected on each image (as described in Section 2.3). If N feature points are detected in the first image and N' in the second, then $N \times N'$ possible pairs should be considered. To produce the exact set of correct correspondences, these $N \times N'$ matches must be filtered to eliminate the mismatches, while preserving all good matches. However, the great size of this set would make an exhaustive visual examination too laborious.

Fortunately, to build the ground truth set, not all pairs need to be considered. First, the range of horizontal and vertical displacements of corresponding image locations can be easily determined by visual inspection of each image pair. All possible matches involving the displacement of a feature point outside this range can be rejected. Secondly, the matches that do not agree with the epipolar geometry (see Section 1.2) of the image pair can also be eliminated. To this end, the epipolar geometry of each image pair was estimated using the Projective Vision Toolkit [1] developed by Roth and Whitehead, and described in [90].

Here, uncalibrated matching is being investigated, but an epipolar geometry may still exist for the test image pairs. This epipolar geometry can be used to build the ground truth set, although it will be assumed to be unavailable at the time of matching. This unavailability of the epipolar geometry at the time of matching is nevertheless a reasonable assumption, especially for the common situation where the goal of uncalibrated matching is epipolar geometry estimation. As long as there is sufficient translation between the views, the epipolar geometry exists, and is used, here, to build the ground truth set.

After a pruning using feature point displacements and the epipolar geometry of the image pair, a smaller set of feature point pairs remains from which all the good matches can be extracted visually in reasonable time[2].

[1] Freely available from http://www.cv.iit.nrc.ca/~gerhard/

[2] These images, the detected feature points and the ground truth match sets are available at www.site.uottawa.ca/research/viva/projects/imagepairs/

2.2.2 The Evaluation Mechanism

Having identified the set of all possible good matches between selected feature points in a given image pair, it becomes possible to empirically evaluate the effectiveness of a matching strategy. A matching constraint is useful if it filters out many mismatches found in an input set of point pairs, while preserving most correct matches.

A given method will use different parameters or thresholds towards keeping or rejecting candidate matches, but there is usually a tradeoff involved in the setting of these parameters. In order to appreciate the effectiveness of a given approach, results will be shown on a graph showing the number of good matches in the resulting match set (on the y-axis) versus the proportion of good matches in that set (on the x-axis). A curve will be plotted to represent results obtained for experiments with different values of a control parameter for the method under study.

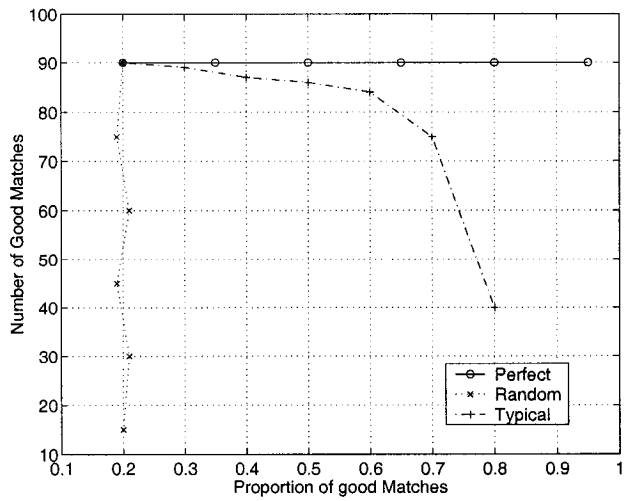


Figure 2.7: The evolution of the number of good matches in a given match set, versus their proportion in that set. The three curves correspond to cases of methods behaving perfectly, randomly, or typically, as discussed in the text. The common starting point corresponds to an initial set containing, in this fictitious illustrative example, 90 good matches representing 20% of the total set. Data points on each curve correspond to different control parameter values, filtering being more severe as we move to the bottom-right.

In such a graph, a perfect constraint produces a horizontal line. The constraint eliminates only mismatches, thus increasing the proportion of good matches as the threshold is tightened, but without reducing the number of correct matches. Conversely, a useless constraint produces a nearly vertical line, i.e. one that eliminates points randomly, thus without modifying the proportion of correct matches.

These ideal situations are illustrated in Figure 2.7. The common starting point in this graph corresponds to an initial set containing, for this fictitious illustrative example, 90 good matches representing 20% of the total set. Data points on each curve correspond to different control parameter values, filtering being more severe as we move toward the bottom-right. Note that, in practice, we might expect that an effective method would produce a nearly horizontal or slightly descending curve until some point where the curve starts dropping vertically, as more severe thresholding cannot further improve the quality of the match set.

2.3 Feature Point Selection

The choice of the type of feature points to be matched certainly has an impact on the results of a given matching scheme. Several feature point detectors exist; [55] and [102] offer good reviews on the subject.

2.3.1 Harris Feature Points

Among the most popular feature point detectors are those based on the *Plessey* operator described by Harris in [38]:

$$C(x,y) \;=\; S * (\nabla I(x,y) \cdot \nabla I^{\top}(x,y)) \;=\; S * \begin{bmatrix} I_x^2(x,y) & I_x I_y(x,y) \\ I_x I_y(x,y) & I_y^2(x,y) \end{bmatrix} \tag{2.1}$$

where S is a smoothing operator. At the point where it is computed, the greatest eigenvalue of this matrix is a measure of the image's rate of change in the direction of highest variation, while its smallest eigenvalue corresponds to its rate of change

in the perpendicular direction. If the smallest eigenvalue has a high magnitude, it means that, at the considered point, the image has a high rate of variation in two perpendicular directions, and thus, that the point is in a high curvature area of the image. Typically, Sobel filters are used to compute the derivatives, while a spatial averaging filter is used as the smoothing operator.

Noble [79], as well as Harris and Stephens [39] have defined operators which exploit this idea. For example, Harris and Stephens used the following strength measure:

$$cornerStrength(x, y) = Det(C(x, y)) - 0.04\ Trace^2(C(x, y)) \tag{2.2}$$

with $C(x, y)$ defined in equation (2.1).

However, a direct computation of the minimum eigenvalue of $C(x, y)$ is now preferred. The Intel Open Source Computer Vision library (OpenCV) [80] proposes an implementation[3] based on such an explicit eigen-decomposition.

Of course, with these feature point detectors, non-maxima suppression must be applied to eliminate clusters. Thus, only the points with a corner strength above some threshold, and without immediate neighbors having higher corner strengths are selected.

2.3.2 SUSAN Feature Points

The well-known SUSAN detector was proposed by Smith and Brady [102]. It is based on the computation of the area, inside a circular window, having pixel intensities similar to the circular window's central point. This area is called the univalue segment assimilating nucleus (USAN). A given image location is a feature point when the corresponding USAN covers an area below some given threshold.

The implementation of the SUSAN detector that was used in this work came with version 1.0 of the previously described Projective Vision Toolkit[4] described in [90].

[3] This function is called cvCornersEigenValsandVecs

[4] More recent versions use a different feature point detector

2.3.3 Comparing Harris and SUSAN Detectors

When feature points are used in matching, the most important property of the detector, among those listed in Section 1.7, is its stability. Stability can be measured as the *repeatability* rate of the detector: the proportion of feature points which are detected in two different views, despite the effects of perspective distortion. The Harris detector was shown by Schmid et al. [98] to be the most stable among a small selection of detectors, with regards to this property.

To confirm this result, and since Schmid et al. did not include the SUSAN feature point detector in their experiments, the Harris and SUSAN detectors will now be compared. Schmid et al. detected feature points in images of flat surfaces, which are thus related by a pure homography (see Section 1.2). This has the advantage of making it possible to very easily check if a given feature point was found in both images using Equation (1.6). The approach used here is more thorough, as it is not limited to such simple scenes.

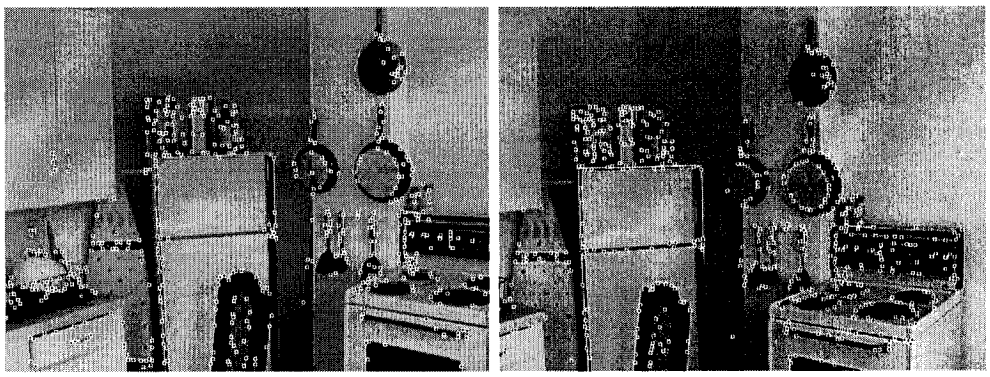


Figure 2.8: 500 feature points extracted with a SUSAN feature point detector on the pair from Figure 2.1.

The feature point detectors were used to extract approximately 500 points in each image pair from Figures 2.1 to 2.6. Then, all good matches among these feature points were extracted using the method described in Section 2.2. From this, the number of

scene feature points that have been correctly detected in both images (the repeatability of the feature point detector), could be obtained. Comparing these repeatability rates between Harris and SUSAN for an equal number of detected features resulted in a repeatability rate that was approximatively 3 times higher for the Harris detector. For example, in the case of the *Kitchen* image pair, using feature points detected by a Harris detector, 148 correct matches were found, while SUSAN only produced 48 (see Figure 2.8). The Harris detector is thus shown indeed to be more stable.

2.3.4 Tuning a Harris Detector

The Intel OpenCV library includes another function[5] for feature point detection. This one filters Harris feature points, to ensure that they are far enough from one another. This is done by iteratively finding the strongest feature point, and throwing out all other feature points that are closer than a threshold distance from it.

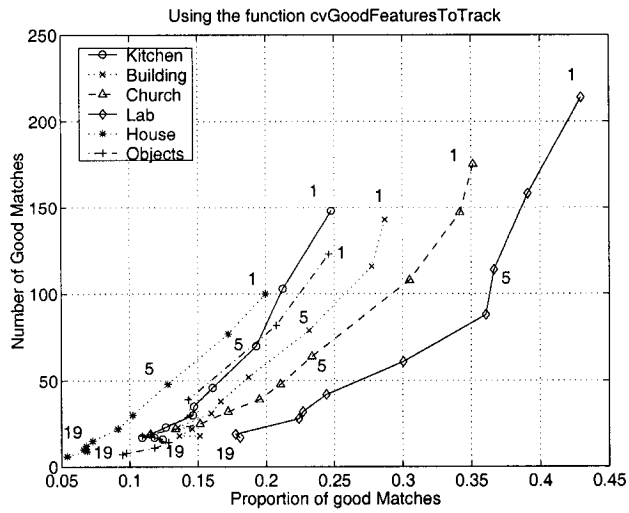


Figure 2.9: Eliminating feature points close to stronger ones, for different minimal distance between the feature points. The numbers shown represent the minimal acceptable distance between corners.

[5]called cvGoodFeaturesToTrack.

In order to determine if this method brings an increase in the quality of the candidate match set, other experiments were performed on the images shown in Figures 2.1 to 2.6. The feature points detected in both images were again counted, for different minimum distance thresholds. The resulting graph, shown in Figure 2.9, shows that any increase in the minimal distance constraint worsens the set of feature points by reducing the proportion of potentially matching feature points in that set. Thus, this added step was harmful, and should not be used, at least when feature points are detected for the purpose of matching.

Because of the results from the previous experiments, the Harris detector was selected to be used in the remainder of this chapter. However, in other parts of this work, different detectors will be introduced which prove to be superior, in some respects, for different matching situations.

It should nevertheless be noted that when using the Harris detector, the value of the threshold imposed on the least eigenvalue of Matrix (2.1) directly influences the number of feature points detected. The effect of this threshold on the quality of the resulting candidate sets, should then be examined. Feature points were detected using different thresholds on the image pairs from Figures 2.1 to 2.6, and the resulting repeatability rate was determined.

Results are shown in Figure 2.10. This graph shows that, within a reasonable range, the proportion of detected feature points remain relatively constant. It therefore follows that one can increase the number of good matches that can be found, just by accepting more feature points. However, this is done at the price of a proportional increase in the total number of feature points to analyze. The Harris detector's threshold should therefore be set so that the number of matches found is suitable, but not much greater than the amount needed to obtain enough matches in the considered application.

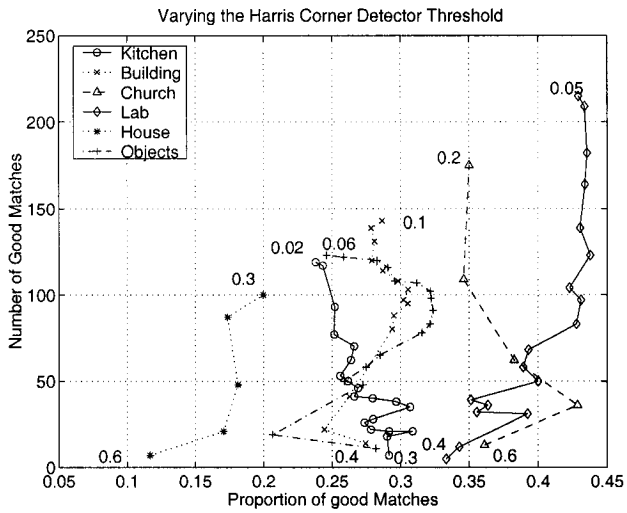


Figure 2.10: Modifying the corner strength threshold in the Harris feature point detector.

2.4 Similarity Measures

Correlation is the most common means by which feature points in different images are compared. It measures the similarity between image point neighborhoods. Several correlation-based measures of similarity are presented in works such as [3, 10, 18, 126, 127]. In this section, two different similarity methods will be examined.

2.4.1 Average Squared Difference

A simple method is the average squared difference (ASD) of the intensities of corresponding pixels. This would be applied between neighborhoods of the feature points. ASD is defined for a pair of points $(\mathbf{x}, \mathbf{x}')$ as:

$$\mathrm{ASD}(\mathbf{x}, \mathbf{x}') = \frac{1}{N} \sum_{\mathbf{p} \in \mathcal{N}} [I(\mathbf{x} + \mathbf{p}) - I'(\mathbf{x}' + \mathbf{p})]^2 \tag{2.3}$$

where $\mathcal{N}$ is a shape of area N that defines the regions to be compared around $\mathbf{x}$ and $\mathbf{x}'$.

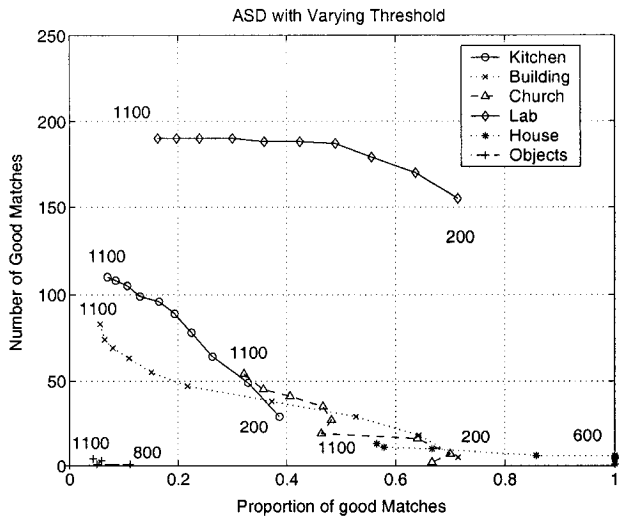


Figure 2.11: Comparing feature points of Figures 2.1 to 2.6 with ASD. Each data point represents a test with different correlation thresholds between 200 and 1100 on 11×11 windows.

Two parameters influence the performance of this similarity measure: the size N of the windows $\mathcal{N}(\mathbf{x})$ and $\mathcal{N}(\mathbf{x}')$, and the threshold used to decide when to accept or reject a match. The results shown in Figures 2.11 and 2.12, where ASD was applied to the image pairs of Figures 2.1 to 2.6, illustrate how these parameters affect the quality of the resulting match set. As expected, tightening the threshold increases the proportion of good matches, but at the same time, decreases the number of good matches quite rapidly in some cases.

The experiments also shows that increasing the size of the window is an effective way to identify and reject more false matches (an observation also made by Nishihara and Poggio in [78]), but this is only true up to a certain size (about 11×11 windows). It should also be noted that the ASD correlation was completely ineffective on the *objects* image pair, where there is an important change of illumination. A normalizing factor would therefore need to be introduced to cope with this problem.

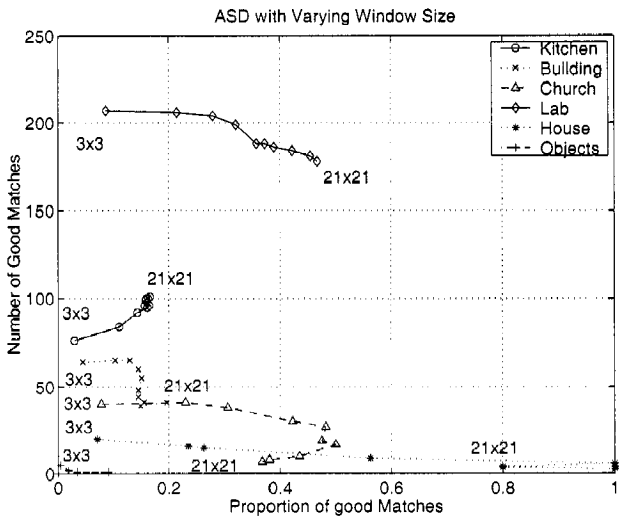


Figure 2.12: Comparing feature points of Figures 2.1 to 2.6 with ASD. Each data point represents a test with different window sizes between 3×3 and 21×21, and a threshold of 700.

2.4.2 Variance Normalized Correlation

Variance normalized correlation (VNC) is another commonly used correlation method. It offers the advantage of producing stable and reliable results over a wide range of viewing conditions. VNC is defined for a candidate match $(\mathbf{x}, \mathbf{x}')$ as:

$$\mathrm{VNC}(\mathbf{x}, \mathbf{x}') = \frac{1}{N\sqrt{\sigma_I^2(\mathbf{x})\sigma_{I'}^2(\mathbf{x}')}} \sum_{\mathbf{p}\in\mathcal{N}} [I(\mathbf{x}+\mathbf{p}) - \overline{I(\mathbf{x})}][I'(\mathbf{x}'+\mathbf{p}) - \overline{I'(\mathbf{x}')}] \qquad (2.4)$$

where $\overline{I(\mathbf{x})}$ and $\sigma_I^2(\mathbf{x})$ are respectively the sample mean and the sample variance of pixel intensities over a neighborhood of shape $\mathcal{N}$ around $\mathbf{x}$. The fact that VNC scores are normalized to a range $[-1, 1]$, is an advantage over other correlation functions, as it makes the choice of a threshold much easier.

Results of the application of VNC, shown in Figures 2.13 and 2.14, demonstrate a behavior similar to the ASD measure, except for the *Objects* pair where VNC did not fail when faced with a change in illumination, as did ASD. Nonetheless, for any choice of parameters, the results obtained by VNC are always superior to the ones of ASD. For example, by comparing Figures 2.11 and 2.13, and excluding the *Lab* and *Objects*

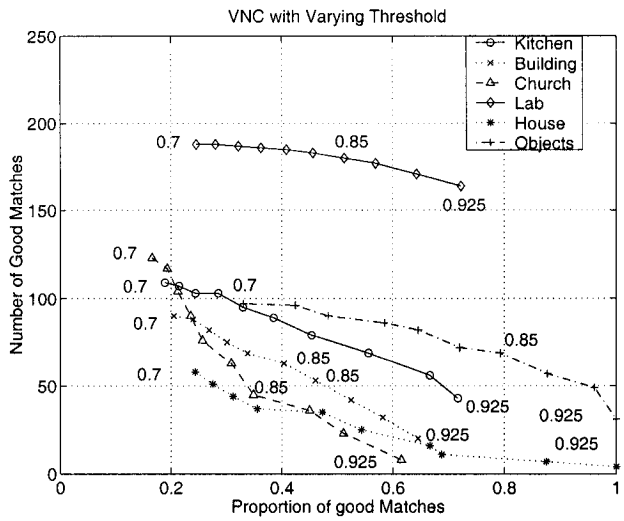


Figure 2.13: Correlating feature points of Figures 2.1 to 2.6 with VNC. Each data point represents a test with a particular threshold between 0.7 and 0.925, on 11×11 windows.

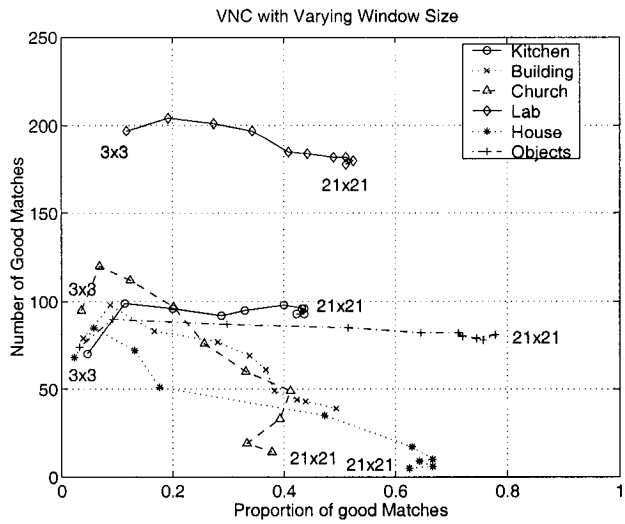


Figure 2.14: Correlating feature points of Figures 2.1 to 2.6 with VNC. Each data point represents tests with different window sizes between 3×3 and 21×21, and a threshold of 0.8.

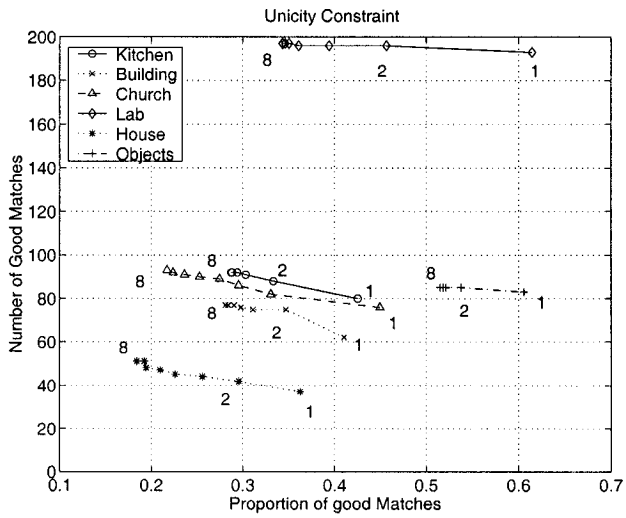


Figure 2.15: Imposing the unicity constraint with varying order, while applying VNC to the feature points of Figures 2.1 to 2.6, 9×9 windows and a threshold of 0.8 are used.

images, we see that for ASD, the average number of good matches is reduced by about 22.6 when the proportion of good matches is increased by 10%. For VNC, only 14.4 are lost, on average. These experiments simply confirm the well known fact that VNC, although more complex, is superior to ASD. Considering the computational power now available, VNC should always be the method of choice when correlating intensity values in image windows.

2.4.3 Unicity

So far, all pairs having a correlation score above some threshold value were accepted. Thus, a feature point could be matched with several others. Imposing unicity means that for each feature point, only the strongest match in the other image is accepted. A generalization of unicity was studied, where the n strongest matches are kept.

Figure 2.15 shows the results of applying VNC to the image pairs, while imposing unicity of different orders. This constraint proves beneficial, as it rejects many mismatches. This important improvement obtained in the proportion of good matches

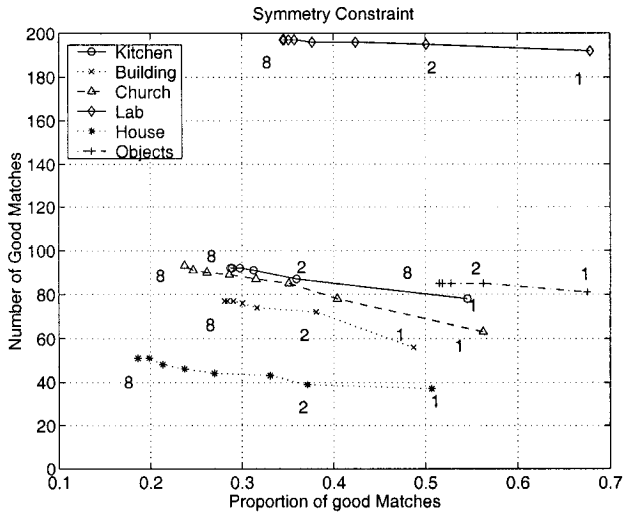


Figure 2.16: Imposing symmetry on the sets obtained in Figure 2.15.

is at the expense of a fairly small loss in the number of correct matches. The significant improvement in the quality of the match sets resulting from the use of a unicity constraint should justify its general use with order 2 or 1. This is especially true considering the very low additional computational cost that it brings.

2.4.4 Symmetry

When unicity is imposed, VNC becomes asymmetric. A point $\mathbf{x}'$ which is the strongest match for a point $\mathbf{x}$, could itself have, as its strongest match, a point $\mathbf{y}$ different from $\mathbf{x}$. However, this situation is physically impossible, and is thus a sign that at least one of the pairs $(\mathbf{x}, \mathbf{x}')$ or $(\mathbf{y}, \mathbf{x}')$ is a mismatch.

Imposing symmetry means keeping only pairs in which each point is the other's strongest match [29]. This increases the chances that matched points correspond to projections of the same physical scene point. Figure 2.16 shows the results of the same experiment as in Figure 2.15, but where the symmetry constraint has been applied in addition to unicity. It shows that imposing symmetry is clearly advantageous as it eliminates many mismatches while affecting few good ones.

Symmetry can also be implemented in a very efficient way. The idea is to simultaneously keep track of the best matches for points $\mathbf{x}$ in a first image, and points $\mathbf{x}'$ in a second, as different pairs of points are correlated. Then the additional cost for imposing first order unicity and symmetry is insignificant, when compared to the cost of computing the correlation scores.

Image pair	Proportion of good matches	Number of good matches
Kitchen	54.5%	78
Building	48.7%	56
Church	56.2%	63
Lab	67.8%	192
House	50.7%	37
Objects	67.5%	81

Table 2.1: Characteristics of the match sets obtained from VNC correlation using a 9×9 window and a threshold of 0.8 on which unicity of order 1 and symmetry have been applied. These are used in the experiments of the next sections.

For this reason, the experiments presented in the remainder of this work will use match sets obtained from VNC on 9×9 windows, and using a threshold of 0.8 with first order unicity and symmetry. Table 2.1 summarizes the characteristics of the resulting match sets for the images of Figures 2.1 to 2.6. The question is now to determine how the quality of these match sets can be further improved.

2.5 Using Feature Point Properties

Many properties have been used to describe image points. Some should remain relatively invariant with regards to small changes in viewpoints, and hence could be used to constrain matching.

2.5.1 Using Feature Point Characteristics

Curvature, gradient magnitude and direction, and orientation of incident edges are generally considered as good invariant feature characteristics. Deriche et al. [21] use the gradient direction and the curvature, among other constraints, in a correlation scheme. Similar constraints are described in by Arbouche [4]. More invariant properties are described in the widely separated view literature such as in [66, 77, 97] that will be described further in Chapter 5.

The principal curvatures of a point are given by the eigenvalues of Matrix (2.1), λ_1 and λ_2. Jung and Lacroix [51] use these eigenvalues to define the *cornerness* at a point (x, y) as:

$$c_{xy} = \lambda_1^2 + \lambda_2^2 \tag{2.5}$$

They then use the ratio of the cornerness of two feature points as a similarity measure between them.

A corner orientation constraint can also be defined using Matrix (2.1). As described in Subsection 2.3.1, its eigenvectors correspond to the two principal gradient directions. The corner orientation can then be defined as the direction corresponding to the highest eigenvalue, which is the direction of highest variation in image intensity at the considered point. Alternatively, the bisector of the angle defined by the two principal directions is a more robust characteristic, since after a small change of viewpoint, when both eigenvalues were initially similar, their order might be inverted. This orientation measure can be used to constrain matching by requiring that corresponding corners have a similar orientation.

Figure 2.17 shows the results of applying the constraint on cornerness, and Figure 2.18 shows the result of applying the constraint on corner orientation. It is seen that these constraints do not improve the candidate match set. This is not surprising since such properties are indirectly taken into account by correlation. If two image regions already exhibit high correlation, they will necessarily have similar properties. Thus, although they are sometimes used in matching constraints, such properties should not be used within a correlation-based scheme.

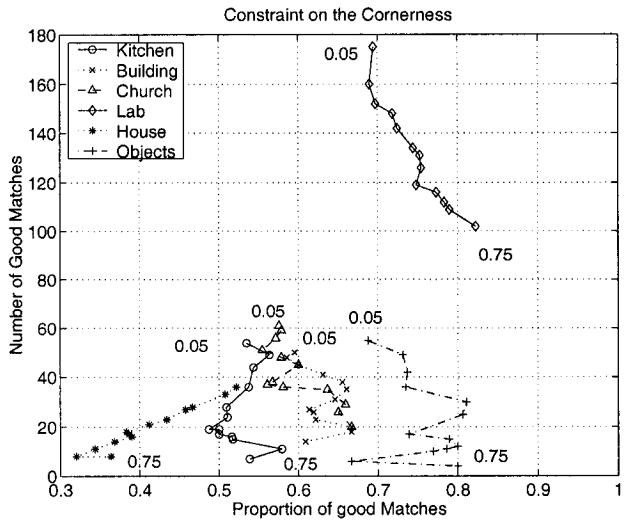


Figure 2.17: Applying a cornerness constraint to the match sets of Table 2.2 with varying thresholds.

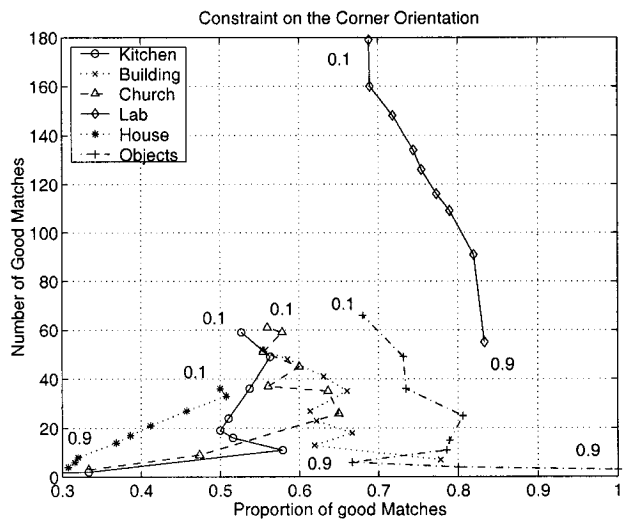


Figure 2.18: Applying a corner orientation constraint to the match sets of Table 2.2 with varying thresholds.

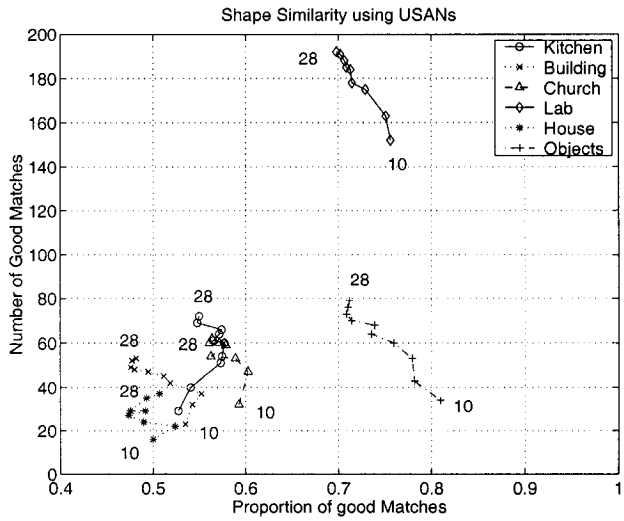


Figure 2.19: Applying the shape similarity constraint to the match sets of Table 2.2, on 9×9 windows, for different choices of threshold, and using USANs.

2.5.2 Corner Shape Similarity

Another possible strategy in using characteristics of feature points in matching constraints is to require that the corners in a pair have similar shapes. This is slightly more effective than the constraints described in the previous subsection. A corner shape is defined here as a small area around the feature point, belonging to the same scene object as this feature point. A method to extract the corner from its background is therefore required. Two such simple methods were investigated.

The first method uses univalue segment assimilating nuclei (USANs) [102], which were already described in Subsection 2.3.2. The idea is to extract the portion of feature point neighborhoods that is of similar intensity as the feature point itself.

The other method is inspired from rudimentary block truncation coding, which was introduced by Delp and Mitchell [19]. The correlation window is separated into two regions according to the window's average intensity value. The foreground consists of the area that contains the feature point.

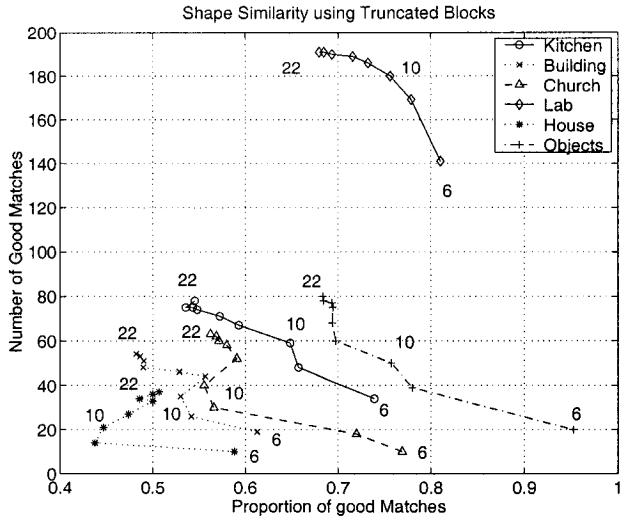


Figure 2.20: Applying the shape similarity constraint to the match sets of Table 2.2, on 9×9 windows, for different choices of threshold, and using truncated blocks.

Once corner shapes have been extracted, the Hamming distance between the obtained binary foreground/background maps are computed for feature point pairs. The pairs for which this distance is above some threshold are eliminated. Results are shown in Figure 2.19 and 2.20. While foreground extraction using USANs does not seem to be effective, results based on truncated blocks show some improvement in the proportion of good matches, although the corresponding reduction in the number of good matches might appear excessive for most applications.

2.5.3 Eliminating the Background

Some correlation functions, such as the one proposed by Darrell [18], attempt to consider only the scene objects on which feature points lie, in establishing correspondence. This is done to mitigate the effect of changes in a feature point's background that are brought by changes in viewpoint, when the background is at a significantly different depth. The foreground extraction methods of Subsection 2.5.2 can be used to determine the region to which correlation should be restricted.

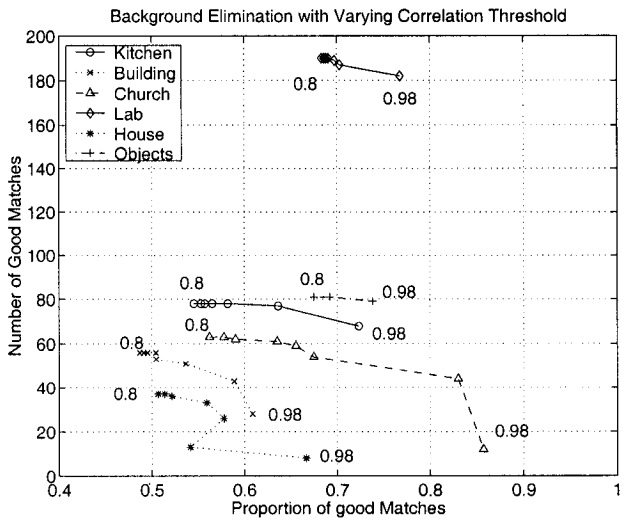


Figure 2.21: Applying the background elimination constraints to the match sets of Table 2.2, on 21×21 windows with a varying threshold, and a multiplier of 0 for the background.

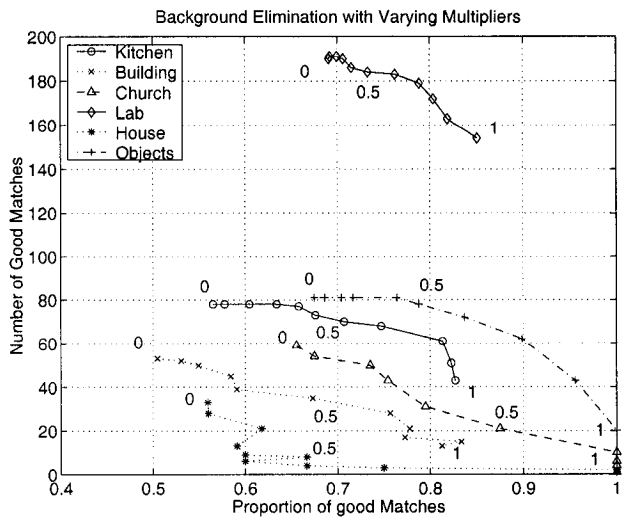


Figure 2.22: Applying the background elimination constraints to the match sets of Table 2.2, on 21×21 windows with a varying multiplier for the background, and a correlation threshold of 0.92.

Results obtained when using this kind of selective correlation are shown in Figure 2.21, where various VNC thresholds were used. In Figure 2.22, VNC was applied directly to the foregrounds of corners, but was weighted by a multiplying factor (in the range $[0, 1]$) when applied to the background. Clearly, these graphs show that performing a correlation on the foregrounds only gives better results than using the simple shape similarity criterion of the previous subsection. Higher proportions of good matches can be achieved with higher multiplying factors, although at the cost of reducing the number of these correct matches.

2.6 Enforcing Displacement Consistency

When the change in viewpoint between two images is limited, the displacements of neighboring matches should be similar. Hence, constraints could be established to ensure that neighbors have a similar behavior. Many authors use iterative processes to enforce such constraints. Relaxation is such an iterative process (see Subsection 2.1.2). In this case, an energy function, corresponding to some aggregate value of a constraint applied to the pairs in a candidate set, is iteratively minimized. Testing the same constraints outside of such an iterative scheme is a simpler, but still accurate way to establish their validity. This is why it was decided to limit the scope of this work to the study of the direct application of displacement consistency constraints.

2.6.1 Confidence Measure

Based on the principle that neighboring pairs of matches should behave in a similar way, a *confidence measure* is proposed by Zhang et al. [127]. It is defined for pairs of points $(\mathbf{x}, \mathbf{x}')$, using the feature points belonging to some neighborhoods $\mathcal{N}(\mathbf{x})$ and $\mathcal{N}(\mathbf{x}')$:

$$\mathrm{conf}(\mathbf{x}, \mathbf{x}') = \mathrm{VNC}(\mathbf{x}, \mathbf{x}') \sum_{\mathbf{p} \in \mathcal{N}(\mathbf{x})} \left(\max_{\mathbf{p}' \in \mathcal{N}(\mathbf{x}')} \frac{\mathrm{VNC}(\mathbf{p}, \mathbf{p}')\ \delta(\mathbf{x}, \mathbf{x}'; \mathbf{p}, \mathbf{p}')}{2 + \|\mathbf{x} - \mathbf{p}\| + \|\mathbf{x}' - \mathbf{p}'\|} \right) \tag{2.6}$$

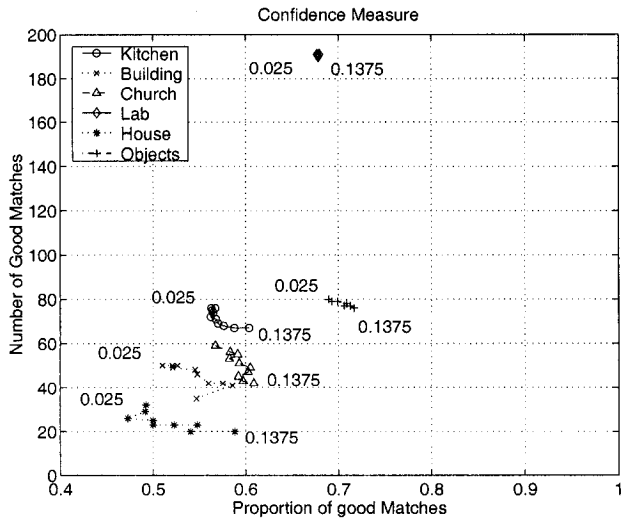


Figure 2.23: Constraints on the confidence measure applied to the match sets of Table 2.2, with varying thresholds.

where $\mathrm{VNC}(\mathbf{x}, \mathbf{x}')$ is the correlation score defined in Equation (2.4), $\|\mathbf{x} - \mathbf{p}\|$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{p}$, and:

$$\delta(\mathbf{x}, \mathbf{x}'; \mathbf{p}, \mathbf{p}') = \begin{cases} e^{-\frac{r(\mathbf{x},\mathbf{x}';\mathbf{p},\mathbf{p}')}{\varepsilon}} & \text{if } (\mathbf{p}, \mathbf{p}') \text{ is a candidate match and } r < \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

for some constant ε, where $r(\mathbf{x}, \mathbf{x}'; \mathbf{p}, \mathbf{p}')$ is a relative distance difference :

$$r(\mathbf{x}, \mathbf{x}'; \mathbf{p}, \mathbf{p}') = \frac{2|\,\|\mathbf{x} - \mathbf{p}\| - \|\mathbf{x}' - \mathbf{p}'\|\,|}{\|\mathbf{x} - \mathbf{p}\| + \|\mathbf{x}' - \mathbf{p}'\|} \tag{2.8}$$

Candidate matches, found in the neighborhood, having a relative position similar to the pair being considered, are essentially counted by this measure.

Figure 2.23 shows the results of constraining the confidence measure with $\varepsilon = 10$ and 61×61 neighborhoods $\mathcal{N}(\mathbf{x})$ and $\mathcal{N}(\mathbf{x}')$. A drawback of this measure is that it cannot be estimated if a point does not have close neighbors in the candidate set. Also, the method was found difficult to tune because of the number of parameters that must be adjusted. Although this constraint achieves positive results on some of the image pairs, it will be seen that it does not perform as well as the much simpler constraint of the following subsection.

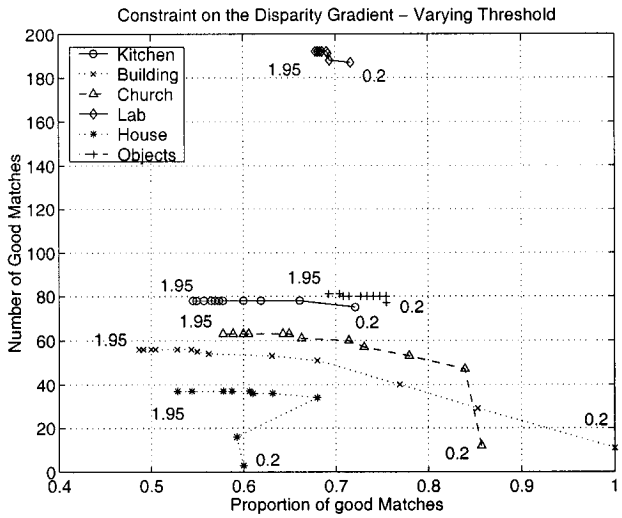


Figure 2.24: Constraint on the disparity gradient, applied to the match sets of Table 2.2, with varying thresholds.

2.6.2 Disparity Gradient

The disparity gradient was introduced by Klette et al. [53]. It is a measure of the compatibility of two pairs of points. For two pairs of feature points $(\mathbf{x}, \mathbf{x}')$ and $(\mathbf{y}, \mathbf{y}')$ with respective displacement vectors $\mathbf{x} - \mathbf{x}'$ and $\mathbf{y} - \mathbf{y}'$, the cyclopean separation $d_{cs}(\mathbf{x}, \mathbf{x}'; \mathbf{y}, \mathbf{y}')$, is defined as the vector joining the midpoints of the line segments $\overline{\mathbf{xx}'}$ and $\overline{\mathbf{yy}'}$, and the disparity gradient is defined as:

$$\Delta d(\mathbf{x}, \mathbf{x}'; \mathbf{y}, \mathbf{y}') = \frac{\|(\mathbf{x} - \mathbf{x}') - (\mathbf{y} - \mathbf{y}')\|}{\|d_{cs}(\mathbf{x}, \mathbf{x}'; \mathbf{y}, \mathbf{y}')\|} \tag{2.9}$$

Compatibility measures, such as the disparity gradient can be used in an iterative process, as done by Roth and Whitehead in [90], where incompatible matches are iteratively removed until all pairs have a similar disparity gradient. Here, the disparity gradient is used in a new way, in a local constraint that enforces that a match's displacement be similar to those of it's closest neighbors.

This constraint accepts pairs that share a disparity gradient below some threshold value with at least 2 of their 5 closest neighbors. Figure 2.24 shows the results of applying this constraint with varying thresholds, and demonstrates that it can

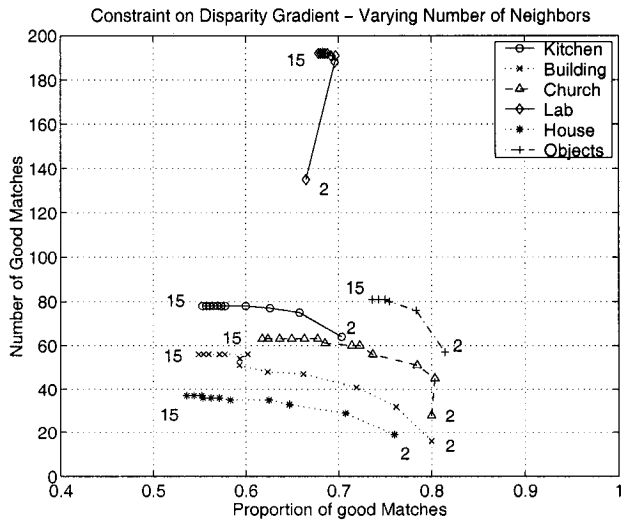


Figure 2.25: Constraint on the disparity gradient, applied to the match sets of Table 2.2, with varying numbers of neighboring pairs.

eliminate a significant number of outliers while eliminating few good matches. It must however start from a set with a significant proportion of correct matches.

A second experiment is presented which evaluates the effect of a change in the proportion of the neighbors that must be compatible with a match, in order for it to be considered valid. Figure 2.25 shows the result of applying the constraint with a threshold of 0.4, but where 2 out of n neighbors must be compatible, for different values of n. This experiment indicates that the disparity gradient constraint is most effective when 2 of between 3 and 5 closest neighbors have a low disparity gradient.

2.6.3 Relative Positions of the Neighbors

In a similar way to what may be done using the disparity gradient, the relative position of two pairs can be constrained. For correct matches, located close to each other, the vectors $\overrightarrow{\mathbf{xx'}}$ and $\overrightarrow{\mathbf{yy'}}$ should be similar when the separation between viewpoints is limited. Thus, a constraint was used, which required the angles between these pairs of vectors to be below a given threshold, with at least 2 of a pair's 5 closest

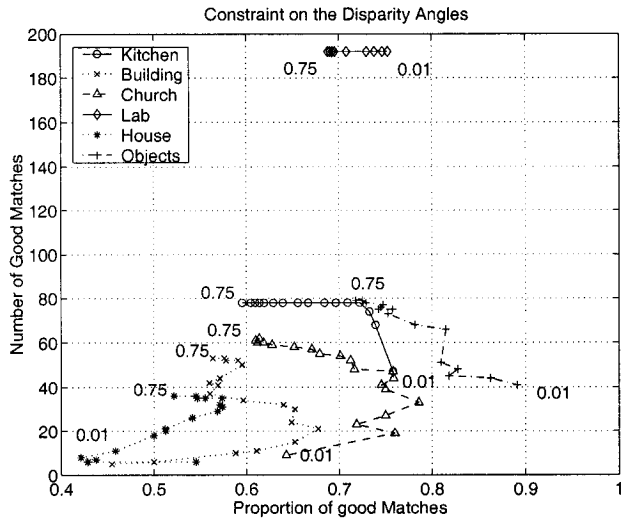


Figure 2.26: Constraint on the displacement angles, applied to the match sets of Table 2.2, with varying thresholds.

neighbors. Another constraint, that required the ratios of the vector magnitudes for 2 of the neighbors to be within certain bounds, was also tested.

Figure 2.26 shows the results of applying the constraint on angles between displacement vectors, and Figure 2.27 shows the constraint on differences between displacement vector magnitudes. As long as the separation between viewpoints is relatively small, these combined simple constraints give similar results as the constraint on the disparity gradient. However, it is more difficult to select two appropriate thresholds on the displacement angles and magnitudes, rather than a single one on the disparity gradient.

2.7 Application to Epipolar Geometry Estimation

The matching strategies that were studied in Sections 2.4 to 2.6 aim at improving match sets by filtering out mismatches. The legitimacy of this objective will now be demonstrated by showing how a better match set can greatly improve the efficiency of fundamental matrix estimation (step 4, of the general matching scheme presented

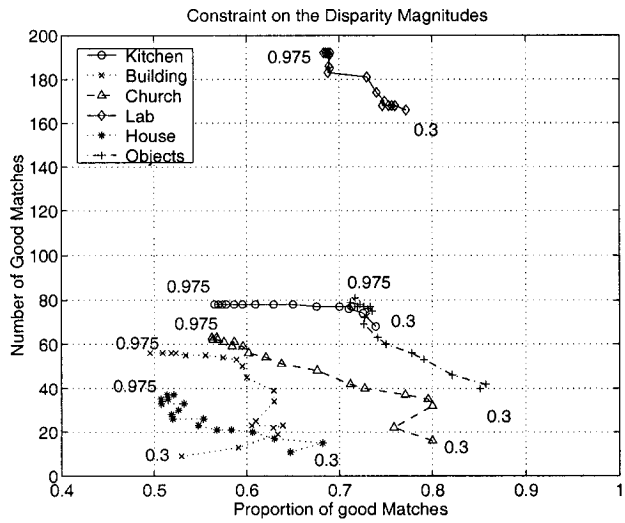


Figure 2.27: Constraint on the displacement magnitude, applied to the match sets of Table 2.2, with varying thresholds.

in Section 1.1).

It should however be noted that the epipolar geometry of a pair of images is not always well defined. For example, when there is no camera translation between the views, no unique relationship exists between points in one view and lines in the other. This is also true of images of a planar surface. In these degenerate cases, the images are related by a global homography. Global homographies can be estimated in similar ways as the epipolar geometry. This will be done in Chapter 6, for example, to construct the mosaic shown in Figure 6.15. In fact, there are other degenerate configurations, such as those defined by affine fundamental matrices that have 4 DOF, and result from a pure translation of the camera, parallel to the image plane.

In general, as the translation between two views diminishes, epipolar geometry estimation becomes less stable, as the relationship between the views approaches a homography. Torr et al. [108] overcome the difficulty of not knowing a priori whether an image pair is in a degenerate configuration by attempting to fit iteratively less degenerate camera geometry models to a set of feature point pairs: the cases of no motion, image translations, translation fundamental matrices, affine image transformations,

homographies, affine fundamental matrices, and general fundamental matrices.

2.7.1 Robust Estimation Schemes

Without knowing the structure of a scene, any correlation-based matching scheme involving only constraints such as those studied in Sections 2.4 to 2.6 will produce some mismatches among the selected pairs of feature points. This is why constraints based on the epipolar or trinocular geometry of the system of cameras are needed. The difficulty is that these camera system geometries are not always known, and must therefore be estimated simultaneously with the matches.

Some robust estimation schemes are required to extract a correct epipolar or trinocular geometry from match sets containing several mismatches. Here, the focus will be on epipolar geometry estimation, or the computation of fundamental matrices. The most common robust fundamental matrix estimation scheme is the random sampling consensus (RANSAC) introduced by Fischler and Bolles [26], and applied to fundamental matrix estimation in Beardsley et al. [7]. Other techniques have been used, but less often, such as the Least Median Squares Estimator (LMedS) used by Zhang et al. in [127].

In a RANSAC fundamental matrix estimation scheme, the minimum number of pairs needed to estimate a fundamental matrix are randomly selected at each iteration. In theory, this means 7 pairs (for the 7 DOF of $\mathbf{F}$), but 8 pairs may be preferred because they will allow a linear computation of the fundamental matrix using the constraints defined by Equation (1.3). (See [67] for more on fundamental matrix estimation). At each iteration, a matrix is computed from the selected pairs, and then tested against all available candidate matches. The cardinality of the subset of all candidate matches which supports this putative matrix estimate measures the accuracy of this estimate. It is expected that the putative fundamental matrix which agrees with the most candidate matches is the most accurate estimate. In practice, fundamental matrices are putatively computed in this way, until one is found which agrees with a given minimal number of pairs, or for a set number of iterations.

It can readily be seen that the probability p of finding a correct solution using a RANSAC scheme can be expressed as:

$$p = 1 - (1 - g^n)^i \tag{2.10}$$

where g is the proportion of inliers in the set from which candidates are putatively selected (correct matches in this case), i is the number of iterations performed, and n is the number of elements (here, candidate matches) that must be selected at each iteration (preferably 7 for $\mathbf{F}$ estimation, but 8 if a linear method is used).

2.7.2 Experiments

Match set	Proportion of good matches	Number of good matches	Expected number of iterations
Kitchen, VNC 0.8	20.0%	96	1 170 207
Kitchen, VNC 0.9	45.7%	59	1 564
Kitchen, VNC 0.8 + constraints	59.8%	70	181
Building, VNC 0.8	16.7%	83	4 951 418
Building, VNC 0.9	36.1%	48	10 407
Building, VNC 0.8 + constraints	65.3%	47	90
Church, VNC 0.8	12.3%	112	55 414 078
Church, VNC 0.9	19.1%	42	1 697 813
Church, VNC 0.8 + constraints	76.8%	43	24
Lab, VNC 0.8	27.4%	201	93 705
Lab, VNC 0.9	53.7%	195	431
Lab, VNC 0.8 + constraints	71.7%	195	42
House, VNC 0.8	13.2%	72	33 246 450
House, VNC 0.9	22.6%	36	433 770
House, VNC 0.8 + constraints	63.9%	23	107
Objects, VNC 0.8	29.8%	87	48 240
Objects, VNC 0.9	90.3%	56	6
Objects, VNC 0.8 + constraints	80.0%	80	17

Table 2.2: Characteristics of different match sets and the theoretical expectation of the number of iterations required to find the exact fundamental matrix using RANSAC.

Table 2.2 presents the characteristics of different match sets which could be used for fundamental matrix estimation, obtained between the image pairs of Figures 2.1 to 2.6. The first two lines of each row correspond to the sets obtained using only VNC, with thresholds of 0.8 and 0.9 respectively (see Subsection 2.4.2). The last line corresponds to the match set obtained with the 0.8 VNC threshold, on which the additional constraints of unicity and symmetry (see Subsections 2.4.3 and 2.4.4) were imposed, as well as the background elimination constraint using truncated blocks with a threshold of 0.9 and a background multiplier of 0.25 (see Subsection 2.5.3), and the disparity gradient constraint with a threshold of 0.4 (see Subsection 2.6.2).

The theoretical expectation of the number of iterations that should be needed, in order to yield an accurate estimate of fundamental matrix for these sets is also shown in Table 2.2. This expected number of iterations was obtained from Equation (2.10) which can be manipulated to give:

$$i = \frac{\log(1-p)}{\log(1-g^n)} \tag{2.11}$$

which now gives the number of iterations needed to find a solution when the number of pairs selected at each iteration is n (7 in this case), the probability of success is p (0.95 was used), and the proportion of good matches in the match set is g, comes from the first column of the table.

It is seen that simply using VNC with a loose threshold can yield poor results in terms of the proportion of correct matches found. Then, it is seen that the most effective way of improving the match set is to filter it, rather than to simply tighten the VNC threshold.

To further demonstrate the usefulness of the constraints presented in Sections 2.4 to 2.6 for filtering match sets, a RANSAC base scheme was applied to estimate fundamental matrices on these filtered sets. To this end, the Projective Vision Toolkit [90] already mentioned in Subsection 2.2.1 was used.

Figures 2.28 to 2.31 show some results which illustrate how the use of matching constraints makes the process of robust fundamental matrix estimation more efficient. Figure 2.28 shows the accurate epipolar geometry for the image pair shown in Figure

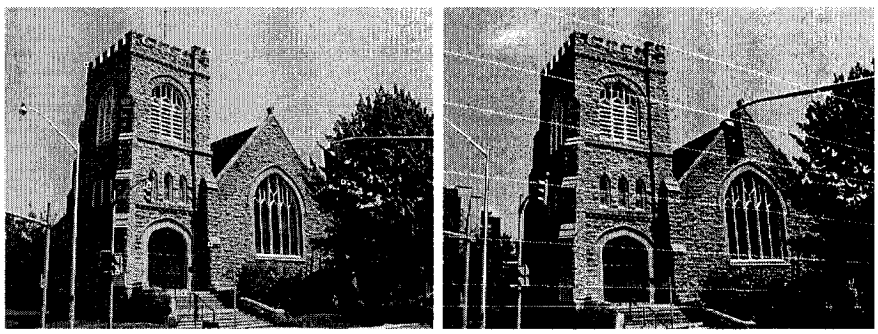


Figure 2.28: The epipolar geometry for the image shown in Figure 2.3. On the left, an image with selected points. On the right, the corresponding accurate epipolar lines in the other image, as calculated from the set of exact matches.

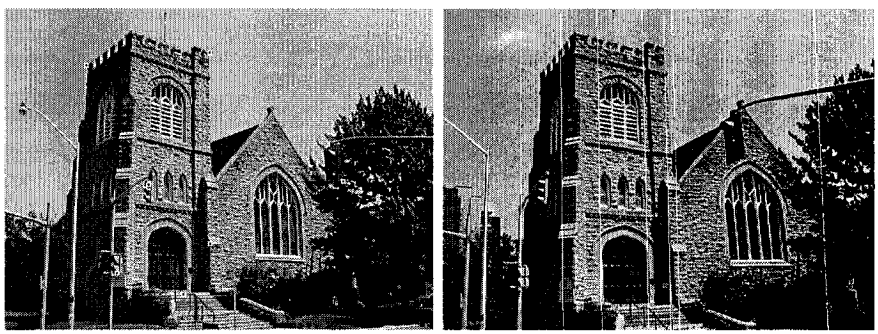


Figure 2.29: The same images as in Figure 2.28, but showing the epipolar lines obtained from the best fundamental matrix found after 500 RANSAC iterations on the set of non-filtered matches.

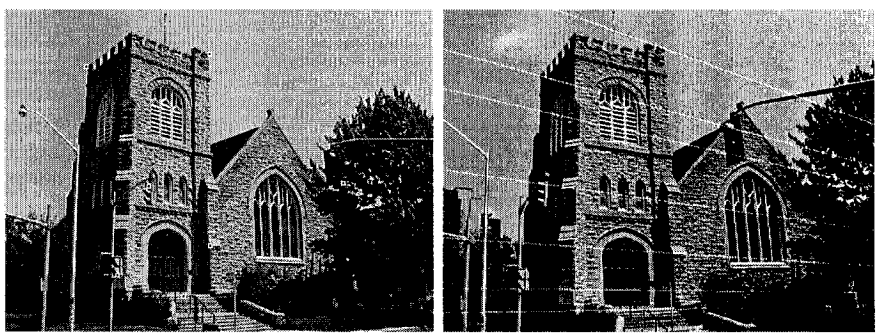


Figure 2.30: The same images as in Figure 2.28, but showing the epipolar lines obtained from the best fundamental matrix found after 500 RANSAC iterations on the set of filtered matches.

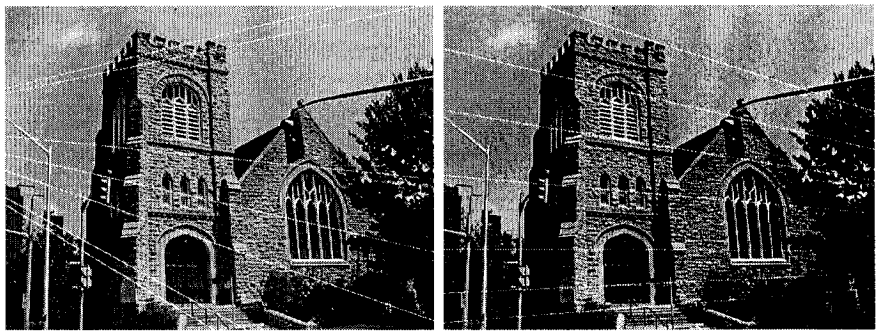


Figure 2.31: Epipolar lines for a fundamental matrix obtained from the same set of unfiltered matches as in Figure 2.29, but after 7500 RANSAC iterations (left image), and 10000 RANSAC iterations (right image).

2.3, which was computed from the ground truth set described in Subsection 2.2.1. In Figure 2.29, it can be seen that the best fundamental matrix found after 500 RANSAC iterations on the unfiltered match set found with a VNC threshold of 0.9 and described in Table 2.2, is an extremely poor approximation. However, after 500 RANSAC iterations on the filtered match set, excellent results are obtained which are shown in Figure 2.30. In Figure 2.31, the epipolar lines in the right image, for the same points that were selected in Figure 2.28 are shown after 7500 and 10000 RANSAC iterations. It is seen that an acceptable solution was found after 10000 iterations, although the one found previously using 500 iterations and the filtered match set was still more accurate. Thus, in this case, the use of the constraints was definitely advantageous over a simple increase of the VNC threshold, as 10000 RANSAC iterations can require a few seconds to run.

2.8 Planar Homography Detection

Obviously, the approach to epipolar geometry estimation presented in the previous section will not always be successful. One situation where this is the case is that of widely separated views, which will be investigated in Part III. Another would be the case of images containing large numbers of repeated structures, which would be a

source of many mismatches. As a RANSAC scheme can be used to estimate an image pair's epipolar geometry, it could also allow the estimation of a planar homography between two image regions. In this section, such a homography estimation scheme will be investigated as a tool to be used toward epipolar geometry recovery in difficult situations, such as when images contain numerous repeated structures.

Because of their abundance and simplicity, planes are used in several computer vision tasks such as auto-calibration [109], reconstruction [61], visual measurement [17], and obstacle detection [65]. As seen in Section 1.2, under perspective projection, the transformation between a world plane and its corresponding image plane is a *homography*.

To benefit from the presence of planes, these structures must be detected. Some strategies have been proposed, such as by Pritchett and Zisserman [85], who use line segment groups. The approach presented here is based on RANSAC. Similar approaches have been used by Lourakis and Orphanoudakis [65], for ground plane homography detection, or by Torr et al. [108], to detect a global homography, considered as a degenerate configuration of a fundamental matrix. Here, however, more than one planar structures will be detected in each image pair.

Fundamental matrix estimation for uncalibrated image pairs, is an important, but sometimes difficult step in many vision applications. Another contribution of the work in this section, will be to demonstrate that planar homography detection can be used as a first step towards fundamental matrix estimation.

2.8.1 Homography Estimation

Homographies were briefly introduced in Section 1.2. Pairs of points $(\mathbf{x}, \mathbf{x}')$ located on a planar area agreeing with a homography $\mathbf{H}$ are related by Equation (1.6). Since homogeneous coordinates are used, equality in this equation is only up to a scale factor.

When at least four pairs of image points lying on a same plane are available, $\mathbf{H}$ can be computed. For each pair of corresponding points, Equation (1.6) gives two

independent linear constraints on $\mathbf{H}$ when it is reformulated as:

$$\mathbf{x}' \times \mathbf{H}\mathbf{x} = \mathbf{0} \tag{2.12}$$

where $\mathbf{0}$ is the null 3-vector, and $\times$ represents the vector cross product. Four point correspondences are required to solve for $\mathbf{H}$, since this matrix has 8 DOF.

Obviously, with non-perfect data, more points should be used in a minimization scheme. N point correspondences will give $2N$ linear constraints on $\mathbf{H}$, using (2.12). This results in a homogeneous linear system of equations of the form $\mathbf{B}\mathbf{h} = 0$, for $\mathbf{h}$ a vector of the entries of $\mathbf{H}$. This kind of problem is usually solved as:

$$\min_{\mathbf{h}} \|\mathbf{B}\mathbf{h}\|^2 \quad \text{subject to} \quad \|\mathbf{h}\| = 1 \tag{2.13}$$

The solution is the eigenvector of $\mathbf{B}^\top\mathbf{B}$ corresponding to the smallest eigenvalue of this matrix. To obtain a stable system of equations, the point coordinates must be normalized, as will be further discussed in Subsection 3.2.2, where the goal is fundamental matrix estimation.

In the method just described, only algebraic quantities are minimized. However, it is preferable to minimize more meaningful geometric quantities, such as the retro-jection error for point pairs, as was done by Criminisi et al. [17]. This would be achieved using a non-linear minimization scheme. However, such a scheme is more costly and requires an initial approximation, usually provided by the linear method that was just described.

2.8.2 Detecting Planar Homographies with RANSAC

Thus, the constraints described in the previous subsection will be used to detect planar homographies in image pairs within a RANSAC scheme, similarly to what was done for fundamental matrix estimation in Subsection 2.7.1. There is however an important difference between homography and fundamental matrix estimation. In the case of $\mathbf{F}$, all matches obey Constraint (1.3), while for homographies, only image points on a common planar area satisfy Constraint (2.12). A different homography must be detected for each image plane.

The homography detection algorithm is as follows:

1. Harris feature points are detected in both images [39].
2. VNC is applied between feature point neighborhoods, to select a set of candidate matches.
3. Four points are selected from the set of candidate matches, and a homography is computed using (2.12).
4. Candidate matches agreeing with the homography are counted. $(\mathbf{x}, \mathbf{x}')$ is said to agree with $\mathbf{H}$ if:
$$\|\mathbf{H}\mathbf{x} - \mathbf{x}'\| < \epsilon \tag{2.14}$$
for some threshold ϵ.
5. Steps 3 and 4 are repeated until a sufficient number of candidate matches agree with a computed homography.
6. Using all consistent correspondences, $\mathbf{H}$ is recomputed by solving (2.13).

Figures 2.32 and 2.33 show an example of a result produced by applying this algorithm.

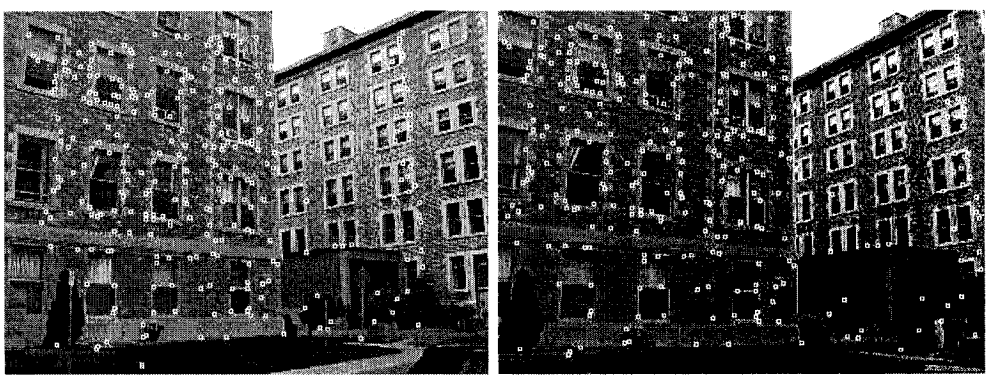


Figure 2.32: An image pair with points agreeing with a detected homography.

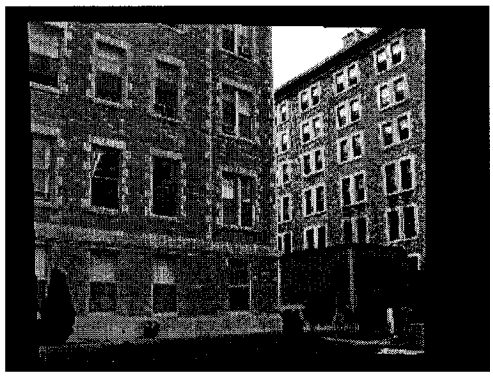


Figure 2.33: The right image from Figure 2.32, warped by the detected homography.

In some situations, images might be very hard to match. This is the case when they contain many repeated patterns, such as the pair shown in Figure 2.37. In such cases, the VNC threshold must be loosened, in order to accept more candidate matches, so that enough correct matches are to be included on each scene plane. However, this results in an even smaller proportion of valid matches in the candidate set, and a need for many iterations before detecting the homographies. The following heuristics were used to reduce the amount of computation needed by the algorithm:

1. If three of the four pairs selected in step 3 were collinear, they would be in a degenerate configuration and would not generate a proper homography, even if they were all correct matches. Thus, the area of the four triangles defined by these four points is computed and compared to a given threshold. This avoids the subsequent, more expensive steps, in the case of some degenerate configurations. Even if not perfectly collinear, three points for which the triangle has a small area usually produce a poor estimate of the homography. About 50% of the randomly selected four point configurations, used towards finding the homography shown Figure 2.32 could be discarded using this criterion.

2. A homography matrix should have rank 3. Therefore matrices with a determinant close to zero are degenerate. Matrices with a very large determinant are

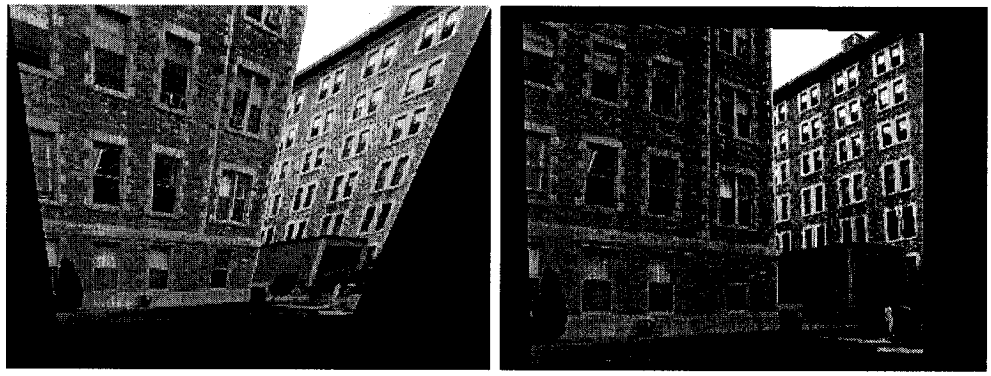


Figure 2.34: Relaxing the distance threshold. Left: The right image from Figure 2.32 warped by the coarse approximation of the homography found with a loose distance threshold. Right: The same image warped by the refined homography.

also degenerate, as the determinant of their inverse is small. Consequently, homographies with a determinant outside a range $[\frac{1}{n}, n]$, can be discarded before counting the number of compatible matches in step 4. With the pair shown in Fig. 2.32, and $n = 10$, 51% of the preliminary homographies found, that satisfied the minimum area requirement, were discarded by this criterion.

3. It was noticed that many of the homographies which are rejected are in fact close to being valid. Thus, a method for improving crude approximations of homographies could allow the acceptance of these close-to-valid estimates. In practice, a generous distance threshold ϵ (such as 3 pixels), and a relatively low threshold for the number of pairs that must be compatible with a homography are used. Then, to improve the homography, it is iteratively recomputed using only the matches compatible with it at that stage. This worked well when the distance threshold was first relaxed (up to 20 pixels), and then gradually tightened, allowing the homography to gravitate towards an accurate estimate. It was found that this technique could decrease by a factor of hundreds the number of iterations needed to detect a homography.

Figure 2.34 gives an example of a crude approximation to a homography which

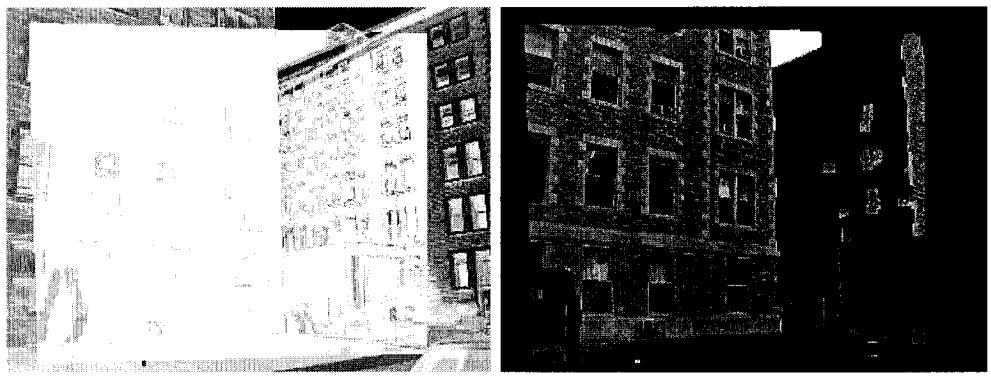


Figure 2.35: The Region Agreeing with a Homography from Figure 2.32. Left: The difference between the left image of Figure 2.32 and the warped image of Figure 2.33. Right: The region agreeing with the homography, found by setting a threshold on the difference image.

was found with few iterations (less then 1% of the iterations needed in Figure 2.32). After improving this crude homography through relaxation of the distance threshold, the resulting homography agreed with 38% more candidate matches then the one shown in Figure 2.32.

2.8.3 Finding Subsequent Homographies

Figure 2.36: Pairs of feature points agreeing with a second homography found in the images of Figure 2.32. It was found using the pairs outside the region agreeing with the homography shown in Figure 2.35.

Image pairs often contain more than one significant planar feature. When a plane is identified, the matches compatible with it, can be removed from the candidate set before a second plane is sought. Incorrect matches having at least one of their points lying in a part of an image where the detected plane is visible, can also be eliminated.

To decide which matches should be eliminated, the region of the image that agrees with the homography must be extracted. This was done by first warping one of the images using the detected homography. The warped image will correspond almost exactly with the other one, in the regions of the detected plane, while disagreeing elsewhere. The region of an image agreeing with a homography is then the one where the difference with its warped counterpart is low.

Figure 2.35 shows this difference. It also shows the region agreeing with the homography which was determined by thresholding the difference and cleaning it with morphological operators. This cleaning consisted in a closing using a small mask to eliminate isolated areas in the planar region, followed by an opening with a larger mask, to eliminate the isolated small areas falsely attached to the plane. All matches with a point in the identified region were discarded before detecting the next homography shown in Figure 2.36.

2.8.4 Fundamental Matrix Estimation from Homographies

One interesting application of homography detection would be for the special case of fundamental matrix estimation in scenes containing at least two planes. The fundamental matrix can be computed from the matches compatible with the homographies using Constraints (1.3), or directly from the homography matrices as done by Luong and Faugeras [68]. In this latter case, the fact that the matrix $\mathbf{H}^\top\mathbf{F}$ is skew-symmetric is used, that is:

$$\mathbf{H}^\top\mathbf{F} + \mathbf{F}^\top\mathbf{H} = \mathbf{0} \tag{2.15}$$

where $\mathbf{0}$ is the 3×3 null matrix. However, it was found that in practice, far better results are obtained when $\mathbf{F}$ is estimated from point correspondences, rather than directly from homographies.

The main advantage of first detecting homographies, and then using the correspondences that agree with them to estimate $\mathbf{F}$ comes in terms of efficiency, when the candidate match set contains many mismatches. It was seen in Subsection 2.7.2 that the theoretical expectation of the number of iterations needed to find a solution using a RANSAC scheme can be expressed as Equation (2.11). There, n would take the value 7 for fundamental matrix estimation, and 4 in homography detection, and the proportion of correct matches in a match set is g. Now, let us assume that at least half of these correct matches agree with a first homography. Let us further assume that after a first homography was detected, and pairs belonging to the corresponding planar region were eliminated, half of the remaining correct matches agree with a second homography. Then, the ratio of the total number of iterations required to detect the two homographies, to the number of iterations required to estimate the fundamental matrix directly can be expressed as:

$$\frac{i_H}{i_F} = \frac{2\frac{\log(1-p)}{\log(1-(\frac{g}{2})^4)}}{\frac{\log(1-p)}{\log(1-g^7)}} = \frac{2\log(1-g^7)}{\log(1-(\frac{g}{2})^4)} \tag{2.16}$$

Then, solving numerically for $\frac{i_H}{i_F} = 1$, it is found that if $g \leq 0.315$, it is advantageous to first detect the two homographies, and then use the points compatible with them to estimate the fundamental matrix. Furthermore, in the situation where $g = 0.2$, it would take four times more iterations to estimate the fundamental matrix directly from the candidate correspondences.

Figures 2.37 to 2.39 show the points agreeing with three homographies that were identified in an image pair. The candidate match set was constructed using a Harris feature points [39], and VNC. This set contained only about 15% of valid matches, mainly because of the repeated window patterns. The three homographies were obtained after 1071, 2605, and 677 iterations respectively. The fundamental matrix was also computed directly from this candidate set with a RANSAC scheme, using Roth and Whitehead's implementation of the direct RANSAC scheme [90], and required 24500 iterations. Thus, as expected, it was greatly advantageous, to first detect the homographies. The fundamental matrix could easily be approximated using the

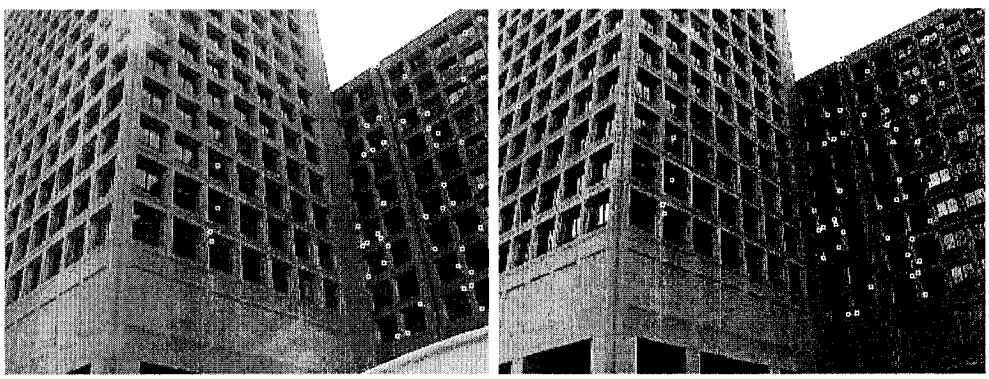


Figure 2.37: Pairs of points agreeing with a first Detected Homography.

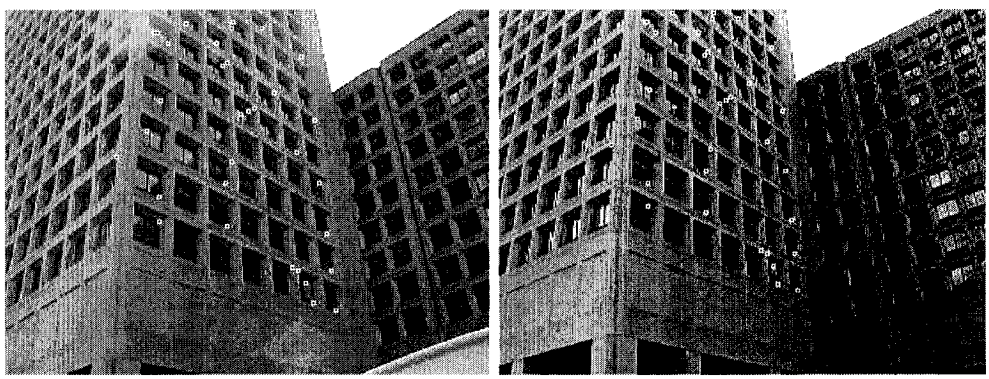


Figure 2.38: Pairs of points agreeing with a second Detected Homography.

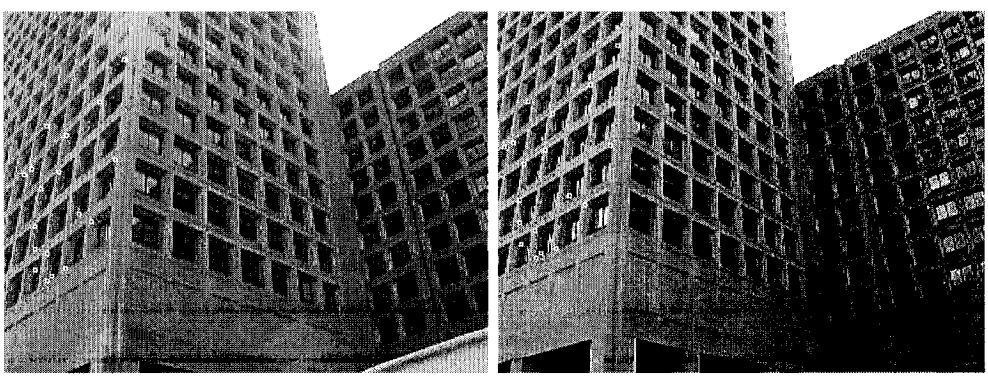


Figure 2.39: Pairs of points agreeing with a third Detected Homography.

candidate pairs that agreed with the homographies.

2.9 Conclusion

In this chapter, a framework for empirically evaluating the effectiveness of matching constraints was developed, and applied to several such constraints. The subset of these constraints that produced the best results was then identified and shown to allow the determination of accurate sets of matches toward epipolar geometry estimation. The main contributions of this work are:

- The proposal of an empirical testing framework for matching constraints.
- The introduction of a graphical representation for the results of experiments with matching constraints, which allows easy evaluation of their usefulness.
- The gathering, and online publication of test data for evaluating matching constraints, including image pairs, sets of feature points, and sets of all accurate matches between the feature points.
- The empirical evaluation of many common matching constraints.
- The introduction of background elimination constraints which are useful for eliminating false matches in situations where scenes contain several objects at different depths.
- The introduction of a simple non-iterative way of enforcing displacement consistency constraints.
- The experimental demonstration that it is preferable to enforce unicity, symmetry, background elimination and a disparity gradient constraints, rather than to increase the correlation threshold to eliminate false matches.
- The experimental demonstration that fewer RANSAC iterations will be needed when the chosen matching constraints are enforced.

- The introduction of a RANSAC-based scheme for planar homography detection in image pairs, incorporating several heuristics to improve its efficiency.
- The demonstration that when significant planar features are present, homography detection can speed up RANSAC schemes for fundamental matrix estimation.

Part II

Matching Calibrated Narrowly Separated Views

Chapter 3

Calibrated Narrowly Separated View Matching Using Harris Features

3.1 Introduction

Calibrated matching has been widely used between two images, especially in dense stereo matching. Here a solution for sparse feature matching is proposed. Although sparse matching is less common in the calibrated case, it can be useful when the information from a dense disparity map becomes overwhelming, especially in real-time applications. Sparse feature point matches could be used towards some limited reconstruction, the generation of new views, or the introduction of virtual objects, especially in dynamic scenes.

In all calibrated stereo matching schemes, even when an accurate estimate of the fundamental matrix is available, many mismatches cannot be avoided. In this Chapter, calibrated matching between three views will be studied. The third view is introduced mostly as a tool to verify matches between the first two, and allows the almost complete elimination of mismatches.

Given three images, and the trinocular geometry relating them, there is little ambiguity in finding matching points. For a point in the first view, the corresponding point in the second view must lie on its epipolar line, and the corresponding point in the third view may be determined by trilinear transfer, as described in Section 1.2. Additionally, if narrow separation is assumed, a point's neighborhood will undergo little distortion between the three views. Thus, if a point triplet agrees with the trinocular geometry and exhibits high correlation between its point neighborhoods, it is very unlikely to be a false match. For these reasons, calibrated narrowly separated view matching can be considered the easiest matching situation.

Here, a possible application to fast reconstruction for evolving scenes will be emphasized. If an operator must remotely manipulate machinery, while having access to views of the work environment taken by a limited number of fixed cameras, he would like to have real-time access to arbitrary viewpoints of the work environment, or to have access to a dynamic model of the environment on which measurements could be taken. If the position, orientation, and internal parameters of the fixed cameras are known, and if point correspondences are established between images taken simultaneously by the cameras, the position of those points in space can be computed by triangulation. Such points would form the basis for the needed model of the environment. In this context, the main challenge in matching will therefore be speed, as well as accuracy and appropriate distribution of the feature point matches, which would allow a quick generation of the best possible model.

Most of the work described in this chapter was done as part of the Sensori-Motor Augmented Reality for Telerobotics Project (SMART), funded by the Institute for Robotics and Intelligent Systems (IRIS). This project aimed to develop a system for the teleoperation of machinery in hostile environments.

3.1.1 Summary

Besides the camera system's trinocular geometry, the tools used in this chapter are generally the same as those introduced for uncalibrated matching in Chapter 2. The

main goal here, is to develop a framework for calibrated matching, which integrates guiding with trinocular geometry to the previously described scheme. The study of this system will allow the identification of its weaknesses, which will bring about the solutions of Chapter 4.

With the exception of the experiments of Section 3.5, the work presented in this chapter was published in:

> Etienne Vincent and Robert Laganière,
> **Matching Feature Points for Telerobotics**,
> *in Proc.* 1^{st} *International Workshop on Haptic Audio Video Environments and their Applications*, pp. 13-18, Ottawa, Canada, November 2002.

The main contribution of the work described in this chapter is the description of a basic system for calibrated narrowly separated view triplet matching. Section 3.2 describes how the camera system's trinocular geometry is estimated. Then, Section 3.3 discusses how feature points are chosen. Section 3.4 describes how feature points are matched. Section 3.5 studies different approaches to making Harris feature point selection more stable. Finally, Section 3.6 concludes with a more detailed list of the contributions of this chapter.

3.1.2 Previous Work

Although the idea of guiding matching between images with their epipolar geometry has been used extensively towards dense matching, or to refine match sets after the robust estimation of a fundamental matrix, little work was done on calibrated feature point matching as it is presented here. The most similar systems found in the literature seek to obtain dense disparity maps from images, under the assumption of a known epipolar geometry, but cannot be fairly compared to a sparse approach. These works include, for example, the systems of Agrawal and Davis [1], and Robert et al. [86], or the system of Lhuillier and Quan [59], where the epipolar geometry is

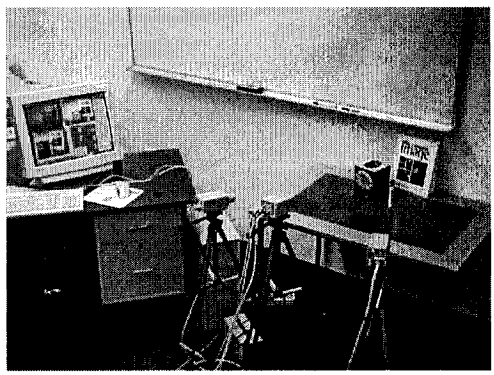


Figure 3.1: The experimental setup.

estimated simultaneously with the disparity map.

Others, such as Ohta and Kanade [81] use an approach based on dynamic programming, a common tool in non-linear optimization, to compare image scanlines. This method assumes rectified images, and thus a known epipolar geometry.

3.2 Weak Calibration of the Camera System

For the experiments described in this work, the setup shown in Figure 3.1 was used. It consists of three fixed cameras pointing towards the same general area. Before feature point matches can be found, the fundamental matrix relating the first two cameras, and a trifocal tensor relating the three cameras must be determined. That is, the system needs to be weakly calibrated.

3.2.1 The Calibration Pattern

One way of performing weak calibration is to do it from point correspondences between the views. Here, these point correspondences will be determined automatically, in an offline process which precedes matching. The calibration pattern shown in Figure 3.2 is used.

Intel's Open Source Computer Vision Library (OpenCV) [80] includes a function

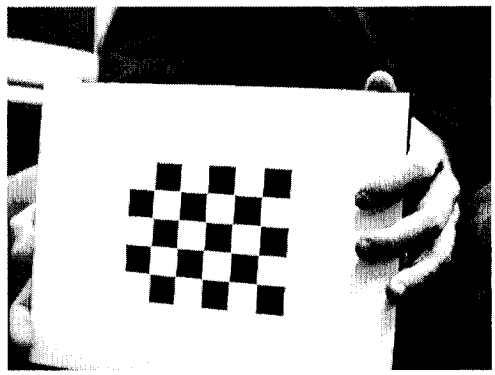


Figure 3.2: The calibration pattern.

which detects this pattern. The function returns a list of the corner locations detected on the chessboard pattern. When these lists of corners are aligned, for three images taken simultaneously, point correspondences are obtained.

However, a single triplet of views of the pattern is insufficient for the estimation of the fundamental matrix and trinocular tensor. This is because the resulting detected point matches are then coplanar, a degenerate configuration. Thus, at least two shots are needed, each consisting of three views of the pattern at different locations. More shots should however be used to obtain better estimates. Over the different shots, the pattern should be positioned to scan as much of the scene viewing volume as possible. This will result in an estimate of the camera system's geometry which is accurate over all possible regions where matches will later be sought.

Figure 3.3 shows the pattern having been detected in all three views. A grid of white squares is overlaid on the pattern, as it is being detected, so that a user of the system can visually verify that it was detected accurately. Usually, five shots of the pattern will produce an accurate weak calibration, if they are well distributed over the whole viewing volume of the scene.

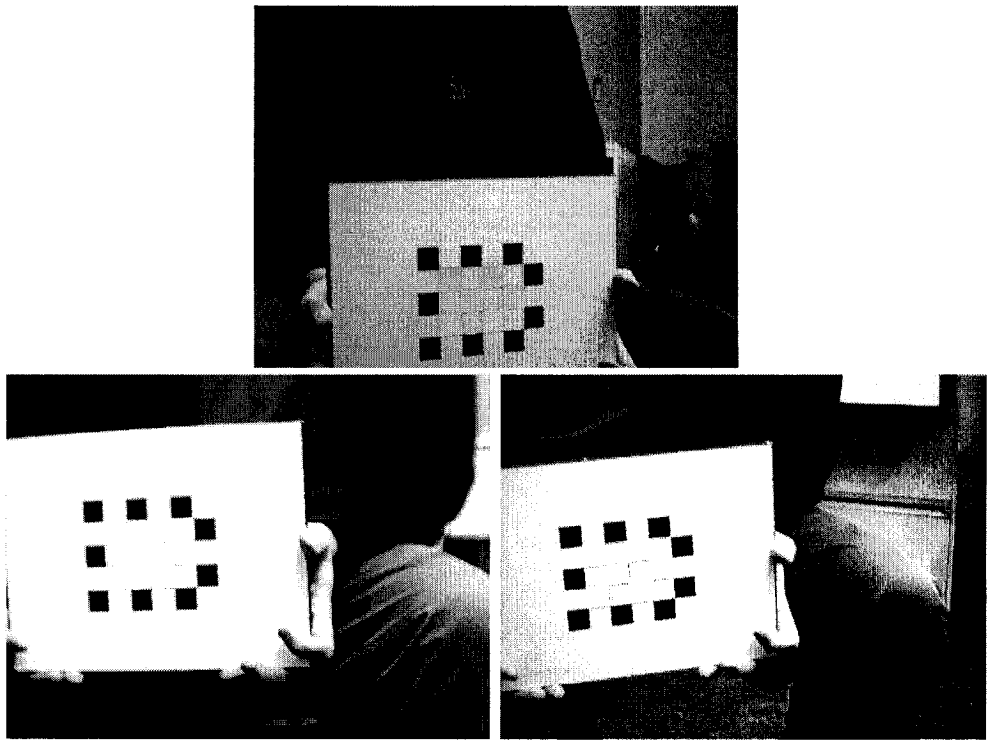


Figure 3.3: Three views of the detected calibration pattern.

3.2.2 Fundamental Matrix Estimation

From the point correspondences obtained between the corners of the chessboard pattern, the fundamental matrix and trifocal tensor can be estimated. The camera positions, orientations and internal parameters could be directly estimated, and from this information, the matrix and tensor could be obtained. However, this approach was found to yield unstable results. Better results were obtained through a direct estimation of the fundamental matrix and trifocal tensor from the point correspondences.

Many different processes, using different parameterizations and minimization methods, have been proposed for fundamental matrix and trifocal tensor estimation, notably those of Papadopoulo and Faugeras [82], and Torr and Zisserman [107]. These

can be rather complicated, as they may impose non-linear constraints on the fundamental matrix and trifocal tensor. Such constraints can then only be enforced through iterative minimization schemes. However, it was found that such accuracy in the fundamental matrix and trifocal tensor are not necessarily needed, when they are used only to guide matching. It was found to be sufficient to perform a direct linear estimation of the elements of the fundamental matrix and trifocal tensor using Equations (1.3) and (1.5).

In the case of the fundamental matrix, each point correspondence $(\mathbf{x}, \mathbf{x}')$ puts one linear constraint on the elements of $\mathbf{F}$ through Equation (1.3). This results in an overconstrained system of homogeneous linear equations of the form $\mathbf{Bf} = \mathbf{0}$. This system can be solved by finding the $\mathbf{f}$ such that:

$$\min_{\mathbf{f}} \|\mathbf{Bf}\|^2 \quad \text{subject to} \quad \|\mathbf{f}\| = 1 \tag{3.1}$$

which is the eigenvector of $\mathbf{B}^\top\mathbf{B}$ corresponding to the minimal eigenvalue.

However, it should be noticed that this solution allows for 8 DOF, and does not enforce that $\mathbf{F}$ be of rank 2. This can be enforced afterwards by using the singular value decomposition $\mathbf{UDV}^\top$ of $\mathbf{F}$, where $\mathbf{D}$ is a diagonal matrix. If $\mathbf{F}$ truly had rank 2, one of the diagonal entries of $\mathbf{D}$ would be 0. If $\mathbf{D}'$ is defined as the matrix $\mathbf{D}$ where the least diagonal element was replaced by a 0, then $\mathbf{F}' = \mathbf{UD}'\mathbf{V}^\top$ is the rank 2 matrix which is closest to the original $\mathbf{F}$, using the Frobenius norm [42]. Of course, it would have been preferable to enforce the rank constraint on $\mathbf{F}$ directly in the minimization, but this would have yielded a non-linear constraint, whereas the solution described here provides sufficiently accurate fundamental matrices, when it is to be used only in guiding the search for correspondence.

It should be noted that, as it was described, the linear solution to the estimation of $\mathbf{F}$ is highly unstable. This is because different point correspondences will weight differently on the system described in Equation (3.1). The solution to this problem, as explained by Hartley [40], is to normalize the coordinate system before solving. The goal is to transform the points so that they weight more evenly on Equation (3.1). The most common solution is to first translate the points so that their average

coordinate becomes the origin, and then to apply the isotropic scaling that will put them at an average distance of 1 from the origin. Of course, once a fundamental matrix is found in this new system of coordinates, it must be denormalized back to image coordinates before being used.

Better results could be obtained by minimizing some more geometrically significant measures than the algebraic constraints of Equation (1.3). Ideally, there would be a simultaneous search for the entries of $\mathbf{F}$ and some point correspondences $(\hat{\mathbf{x}}, \hat{\mathbf{x}}')$ which agree with $\mathbf{F}$, and minimize:

$$\sum \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \|\mathbf{x}' - \hat{\mathbf{x}}'\|^2 \tag{3.2}$$

where $\|\mathbf{x}-\hat{\mathbf{x}}\|$ is the distance between $\mathbf{x}$ and $\hat{\mathbf{x}}$. However, such a minimization becomes complicated, expensive, and is again not needed when $\mathbf{F}$ will only be used to guide matching.

3.2.3 Linear Trifocal Tensor Estimation

The trifocal tensor is estimated with a system of homogeneous linear equations similar to Equation (3.1), but based on the constraints of Equation (1.5), and using the same normalization of point correspondences. However, such a system allows for 26 DOF in $\mathbf{T}$, whereas the true $\mathbf{T}$ is only determined by 18 DOF. Additional constraints should be imposed to get an accurate estimate of the tensor; however, such constraints would again yield a non-linear system, more complicated to implement. The resulting approximation of the tensor is thus relatively poor, but turns out to be sufficient for the purpose of matching. Since it is estimated from point triplets, if these are chosen to be relatively well distributed over the viewing volume, the estimated tensor should produce relatively accurate transfers across this volume.

Thus, it was chosen to estimate the fundamental matrix and trifocal tensor independently from point correspondences established with the calibration pattern. Another possibility would have been to first compute the tensor from the correspondences, and then to extract the fundamental matrix from the tensor. However, it

was found that this approach was much less stable, as a small error in the tensor can translate into a much larger error in the fundamental matrix. For the reasons stated above, the estimated tensor is not accurate when used for purposes other than point transfers, for points that are near correspondences that were used in its estimation.

Once the fundamental matrix and trifocal tensor have been estimated, it was found that it can be useful to visually check their accuracy. This was done using the chessboard pattern. The pattern is detected in two views, then, the epipolar lines of corners in the first image can be drawn in the second one, and the transferred position of corners in the third image can be used to draw the expected appearance of the pattern in that image. It can easily be visually checked that the epipolar lines go through the corners in the second image, and that the computed chessboard corresponds to the actual pattern in the third image. With this system, the computed position of corners in the third image are generally closer than one or two pixels to their real location, for any position and orientation of the pattern.

3.3 Feature Point Selection

The goal, here, is fast matching, that is obtaining point correspondences between three views in close to real time. Attempting to match every point would be costly. In fact, comparing points in the first image with every possible matching point along their epipolar line in the second image is still very costly . Thus, the matching process is limited to a small number of selected feature points. Feature points in the first image are only compared to feature points along their epipolar line in the second image. The detection of feature points in the third image is not needed, however, as the search for correspondence there is limited to a single point's neighborhood.

Only points for which corresponding points are likely to be unambiguously distinguishable should be used as feature points. These points should also, as much as possible, represent significant scene features, to result in a good model of the scene. Points of high curvature on image edges are good candidates for this. These points'

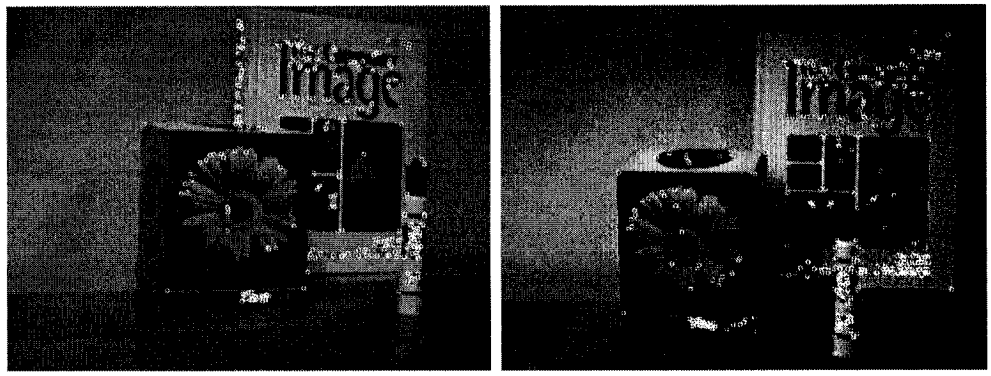


Figure 3.4: Detected feature points in the first two images.

neighborhoods should have a high information content and often correspond to scene corners. For these reasons, Harris feature points are used here, as in Subsection 2.3.1. Figure 3.4 shows some Harris feature points that were detected on images produced by the first two cameras of the setup.

3.4 Matching the Feature Points

The search for correspondences in the second image is thus limited to feature points on the appropriate epipolar line. In practice, to allow for inaccuracies in $\mathbf{F}$, a narrow band along the epipolar line is considered. Variance normalized correlation (VNC) is applied between the points, as described in Section 2.4, with Unicity and Symmetry. Only similar pairs, with a VNC score above some threshold are kept.

The matching process could stop here. Pairs of corresponding points would be sufficient for most applications. However, at this point, the process tends to include too many mismatches. There are too many non-corresponding points which exhibit high correlation and agree with the epipolar geometry. This is why a third image is used, as a way to verify correspondences between the first two.

Given a pair of points in the first two views, the trifocal tensor determines the position of a corresponding point in the third image using Equation (1.5). To allow for inaccuracies in $\mathbf{T}$, a small window in the third image is considered in the search

for the corresponding point. It was found to be preferable to consider each point in such a small window (3×3, or 5×5), rather than to detect feature points in the third image and only consider those. This is because the detection of feature points is more expensive, as long as a limited number of feature point matches are sought, and because many possible correspondences are lost when no feature point is found in the third image.

Now, VNC can be applied between the points in the second image, and all points in the small windows of the third image. Only the point with the highest correlation score in the third image is kept, and this score must be above some threshold as well. Thus, only point triplets agreeing with the camera system's trinocular geometry, and exhibiting high correlation are selected. These are unlikely to be mismatches.

Nevertheless, some mismatches might still be present. The disparity gradient constraint discussed in Subsection 2.6.2 is thus used to filter out these mismatches. Since there should be far fewer mismatches than in the context of Chapter 2, this constraint can be somewhat tightened. Thus, it is required that a match be consistent with 2 out of 3 of its neighbors, by having a disparity gradient below some threshold, rather than using 5 neighbors, as in Subsection 2.6.2.

Figure 3.5 shows the result of applying the above matching scheme to one frame taken by the three cameras of the experimental setup. In the bottom-left image, the lines join the coordinates of feature points there, to their coordinates in the top image, and thus represent the displacement between these two views. Similarly, the lines in the top image indicate the displacement between points in that image, and the bottom-right one. It can be seen that in this frame, there were no mismatches. This is usually the case, but in general, a few mismatches can not be completely avoided.

On a system with a Pentium II 333MHz processor, a speed of 2.55 frames per second was achieved. Of the 392 milliseconds required to process each frame on average, 33 ms were needed to capture the images, 346 ms to detect Harris feature points, 13 ms for the correlation-based matching of feature points, and finally, applying the

Figure 3.5: The matched feature points. Displacement vectors between the top image and the bottom-left image are shown on the bottom-left. Displacement vectors between the bottom-right and top images are shown on the top image.

disparity gradient constraint required less than 1 millisecond.

3.4.1 Algorithm

The following is a description of the calibrated matching algorithm described in this chapter.

1. Harris feature points are detected in images 1 and 2.
2. For each feature point in image 1, the epipolar line in image 2 is computed.
3. All feature points in image 2 that lie on the epipolar line are compared to the feature point in image 1 through VNC (see Subsection 2.4.2). Only the best

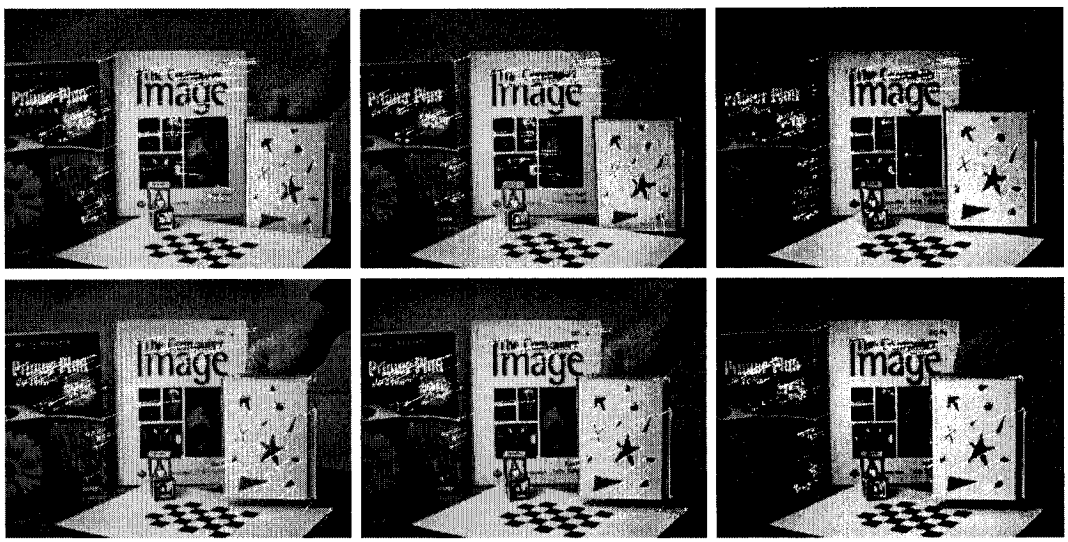


Figure 3.6: The displacement of matched feature points in the first view of 6 frames captured by the system.

correspondence is kept, it must have a VNC score above a threshold value, and satisfy the symmetry constraint 2.4.4.

4. For each selected pair of feature points, the coordinate of a point in image 3 is computed through trinocular transfer.

5. All points in a small area around the point in image 3 are compared to the point in image 2 through VNC. Only the best match is kept, and the VNC score must be above the threshold value.

6. The set of detected point triplets is filtered with the disparity gradient constraint.

3.5 Stability of the Harris Feature Point Detector

It can be noticed, when using the system described in the previous section, that its performance is not as stable as would be hoped. When the system runs on a

	Average Proportion of Detected Points in the Preceding Frame	Average Number of Matches Found
3×3 neighborhoods local max on image 1 only	83.12%	50.61
3×3 neighborhoods local max on both image	77.65%	38.35
5×5 neighborhoods local max on image 1 only	82.26%	34.47
5×5 neighborhoods local max on both images	84.45%	29.61

Table 3.1: Results of some experiments on different ways to apply non-maxima suppression in Harris feature point detectors. The average proportion of feature points which are also present in the previous frame are shown in the second column. The average number of matches found in an image triplet are shown in the third column.

static scene, there are often matches which appear in one frame but are not found in the next, as can be seen in Figure 3.6. This happens as the set of feature points detected in each frame change significantly. In fact, when counted, only 77.65% of feature points in a frame were also found in the previous frame on average. This is probably due to the relative low quality of the cameras that were used, where the pixel intensities that are returned can vary significantly from one frame to the next. Nevertheless, some experiments were conducted to study the problem.

Changes in the way that non-maxima suppression was applied in the Harris feature point detector were investigated in an attempt to increase stability. One idea was that the stability would be increased if non-maxima suppression was applied by enforcing that a feature point have a higher corner strength measure than points in a larger area, rather than only considering its immediate neighbors. Thus local maxima on 5×5 neighborhoods were considered versus the usual 3×3 neighborhoods. Another idea was to only apply non-maxima suppression when detecting feature points in the first image, rather than in both images.

Some results are shown in Table 3.1[1]. It is clear that both modifications have the

[1]Thanks to Mickaeal Biardeau for his help in compiling these experimental results.

intended effect of increasing the stability of the Harris detector. However, it is also clear that using larger neighborhoods for non-maxima suppression ends up reducing significantly the number of valid matches that are found. This strategy should therefore be discarded. That is not surprising, in light of the experiments shown in Figure 2.9 where it is seen that keeping only maxima in larger neighborhoods reduces the number and proportion of potential matches, as many of them are relatively close to each other.

When it comes to the strategy of only taking local maxima on the first image, the result is an increase in the stability of the feature point detector, and in the number of matches found. However, it should be noticed that such a strategy will significantly increase the number of points which are extracted in the second image, by not removing the large clusters found around strong corners. This can in turn greatly increase the time needed by the matching algorithm, and thus, despite its advantages, this strategy was not part of the final implementation.

Another possible modification to the way in which non-maxima suppression is applied would be to only keep feature points which are strong local maxima. These are feature points having a difference in corner strength measure with their neighbors above some threshold value. Conversely, the non-maxima criteria could be relaxed, so as to keep points around a local maximum which themselves have a corner strength no less than some threshold value below the corner strength of the local maximum. These two situations can be described with a single parameter, the minimum difference in corner strengths between a local maximum and a neighboring point.

For a positive threshold on the minimum difference in corner strengths, only points having a corner strength above their neighbors by at least a given margin are kept. A negative threshold on the minimum difference in corner strengths means that the points which are kept are not necessarily local maxima, but are surrounded by points with corner strengths that are no more than the threshold value above theirs. Figure 3.7 shows the results of experiments demonstrating that a negative threshold on the minimum difference in corner strengths produces slightly more stable feature points.

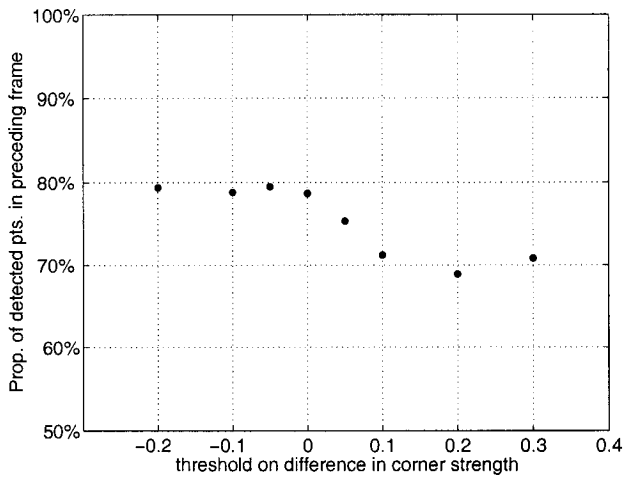


Figure 3.7: The stability of Harris feature points, measured as the average proportion of feature points which were also present in the previous frame, for different thresholds on the minimum difference in corner strengths used in the non-maxima suppression.

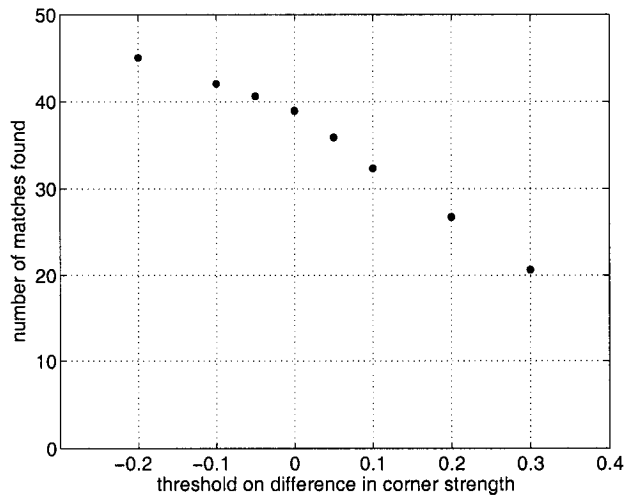


Figure 3.8: The number of point matches found, for different thresholds on the minimum difference in corner strengths used in the non-maxima suppression.

Figure 3.8 shows that a negative threshold on the minimum difference in corner strengths also tends to produce more matches between the three images. This seems to indicate that such a negative threshold should always be used. However, the question of added computation time should also be considered. Lowering this threshold also increases the number of feature points which are kept, which increases the time needed to search for matches. In fact, the increase in the resulting number of feature points is almost enough to explain the increased number of matches that are found. In practice, it was chosen not to use such a strategy, although there might be benefits to a small negative threshold on the difference in corner strengths.

3.6 Conclusion

In this chapter, a method for finding correspondences between calibrated narrowly separated image triplets was described. The main contributions of this work are:

- The development of a fast and continuous calibrated matching application, and its highly efficient implementation.
- The demonstration that a tensor obtained through direct linear minimization of algebraic error from point triplets is sufficient to guide matching between three views, although such a tensor estimate is often far from the true physical tensor.
- The formulation of the idea of using a third image to verify the matches found between two.
- The implementation of a convenient application for detecting a calibration pattern, performing weak calibration, and verifying a trifocal tensor estimate.
- The empirical study of the best way to select local maxima for Harris feature point detectors.

Chapter 4

Calibrated Narrowly Separated View Matching Using Epipolar Gradients

4.1 Introduction

The solution to the calibrated narrowly separated view feature point matching problem described in the previous Chapter is not completely satisfactory. The main problem, when the ultimate goal of matching is quick reconstruction, is that the feature point matches are often poorly distributed in the images, and do not include the most useful areas on the borders of the main scene objects. Furthermore, it is noticed that when scenes with few well defined corner points are used, very few matches result. This is nevertheless a common situation, especially when relatively low resolution cameras are used. Figure 4.9, for example, shows a case where Harris feature points prove inadequate.

These problems are addressed in this chapter. An alternative feature detector is presented which is still fast, but results in more and better distributed matches in a calibrated system of cameras. This detector relies on the concept of *epipolar*

gradients, introduced in this work. A simple way to compare feature points is also presented which exploits the additional calibration information that was not available in the uncalibrated context of Chapter 2.

The work described in this chapter has largely been published. Some of the results shown in Sections 4.2 to 4.7 were published in:

> Etienne Vincent and Robert Laganière,
> **Matching with Epipolar Gradient Features and Edge Transfer**,
> *in Proc.* 10^{th} *International Conference on Image Processing*, vol. 1, pp. 277-280, Barcelona, Spain, September 2003.

Then, a second version including more experiments and the results of Section 4.6 appears as:

> Etienne Vincent and Robert Laganière,
> **Models From Image Triplets using Epipolar Gradient Features**,
> *in Proc. Vision, Video and Graphics*, pp. 143-150, Bath, United Kingdom, July 2003.

This publication was selected to be included in a special journal issue. In this publication, the results of Section 4.7 were also added:

> Etienne Vincent and Robert Laganière,
> **Models From Image Triplets using Epipolar Gradient Features**,
> to be published in *Image and Vision Computing Journal.*

The main contribution of the work described in this chapter is the introduction of a new matching scheme that is especially suited for sparse calibrated narrowly separated view matching. Section 4.3 introduces epipolar gradient feature points. Then, Section 4.4 describes a matching constraint based on edge transfer. Section 4.5 shows some experimental results of the proposed matching scheme. Section 4.6 presents models

constructed from the matched feature points. Finally, Section 4.7 suggests a way of further improving the distribution of computed matches by carefully selecting the camera system's configuration. Finally, Section 4.8 concludes with a more detailed list of the contributions of this chapter.

4.2 Trinocular Line Transfer

In Section 1.2 it was mentioned how the trifocal tensor can be used to transfer points between three images. When the position of a point is known in two views, its position in a third view can be computed with Equation (1.5). This was used in Section 3.4 to guide the search for matching points among three views.

Trifocal tensors can also relate lines between three images. When the equation of an image line is known in two views, it can be transferred to another third one using:

$$l_i = \sum_{j,k \in \{1,2,3\}} l'_j l''_k T_{ijk} \tag{4.1}$$

where $\mathbf{l}$, $\mathbf{l}'$ and $\mathbf{l}''$ are the 3-vectors representing the line in the three images.

This result will be used in Section 4.4.1 to compute the expected slope of the tangent to an edge in the third image.

4.3 Feature Point Selection

In the context of matching, a good feature point detector will select points that are likely to be easily distinguishable from each other, and will select the same points consistently, despite changes in viewpoint. When the further goal is reconstruction, the selected points should also represent significant scene features, such as points on the borders of scene objects, as these are more relevant to the determination of a relevant model.

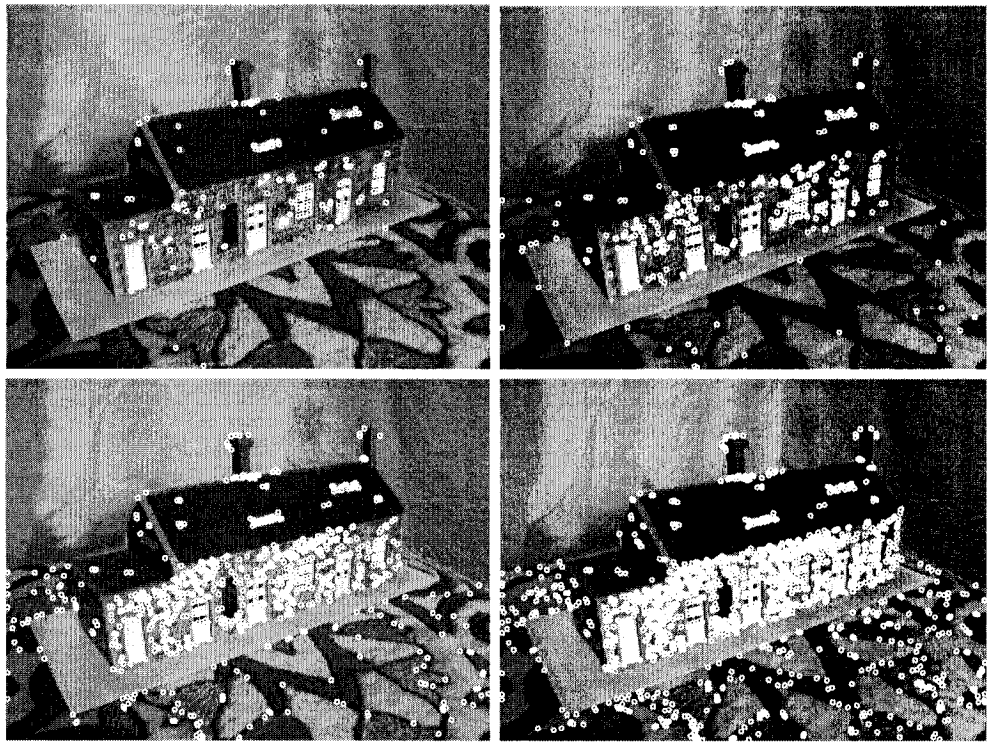


Figure 4.1: Detected Harris feature points using different thresholds. The images contain respectively 226, 433, 865 and 1532 feature points.

4.3.1 Harris Feature Points

The most commonly used feature points are Harris corners [39], which were described in Subsection 2.3.1. A sample set of feature points detected on an image (obtained from [76]) is shown in Figure 4.1, to allow later comparison with epipolar gradient feature points seen in the next subsection.

The Harris feature point detector has been shown by Schmid et al. to be relatively stable [98]. However, in practice, it depends on the presence of appropriate textural content. When textures are not present, few corresponding points might be extracted in both images. Furthermore, Harris feature points are often not distributed well enough to permit an adequate reconstruction of the complete scene using only those points. In Figure 4.1, for example, most points are found on the front wall of the

house, a situation that was not solved by loosening the detection threshold. Also, no matter what threshold was used, few feature points are located on the borders of the main scene features, such as the front wall or roof. The epipolar gradient features, introduced in the next subsection, are meant to overcome these weaknesses.

4.3.2 Epipolar Gradient Feature Points

The idea behind epipolar gradient features is to select points where the image intensity gradient is locally collinear with epipolar lines. These are points where epipolar lines cross perpendicular image edges. If the cameras are positioned in a reasonable configuration, the corresponding points in the other views would also lie on edges that are nearly perpendicular to epipolar lines. Thus, epipolar gradient features in one image should have corresponding points in other images which are also epipolar gradient features. This stability makes them good candidates for matching. Additionally, being points on epipolar lines which cross strong edges, these features can be accurately localized and should be easily discernable from other points on the same epipolar line, allowing matching to be unambiguous. Finally, since they lie on important image edges, such points are often found on the border of significant scene objects and are thus important for scene reconstruction.

More formally, let $\mathbf{I}$ and $\mathbf{I}'$ be two images, with $\mathbf{x}$ a point on $\mathbf{I}$, and $\mathbf{x}'$ its corresponding point on $\mathbf{I}'$. Then, it is clear from Figure 4.2 that $\mathbf{x}'$ is on $\mathbf{l}'$, the epipolar line of $\mathbf{x}$ in $\mathbf{I}'$. Similarly, $\mathbf{x}$ will lie on $\mathbf{l}$, the epipolar line of $\mathbf{x}'$ in $\mathbf{I}$. Now $\mathbf{l}$ and $\mathbf{l}'$ should also correspond, so all points on $\mathbf{l}$ will have their corresponding point lying on $\mathbf{l}'$. Thus if $\mathbf{X}$, the world point projected onto $\mathbf{x}$ and $\mathbf{x}'$, is centered on a locally continuous surface, the points on $\mathbf{l}$ that are immediately next to $\mathbf{x}$ should correspond to points in $\mathbf{I}'$ that lie on $\mathbf{l}'$ and are immediately next to $\mathbf{x}'$. Thus, the intensity gradient of $\mathbf{I}$, at $\mathbf{x}$, in the direction of $\mathbf{l}$, should be similar to the intensity gradient of $\mathbf{I}'$, at $\mathbf{x}'$, in the direction of $\mathbf{l}'$.

The intensity gradient in the direction of the epipolar line will be referred to as the *epipolar gradient*. It can be computed by projecting $\bigtriangledown I(\mathbf{x})$ onto the epipolar line

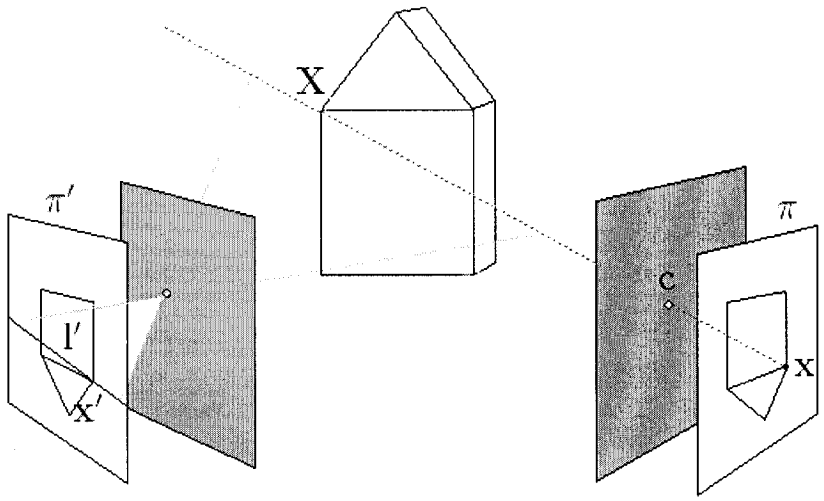


Figure 4.2: Two-view geometry.

$\mathbf{l} = (l_1, l_2, l_3)$, giving the explicit formula:

$$\nabla_{ep}(\mathbf{x}) = \frac{\nabla I(\mathbf{x}) \cdot (\frac{-l_3}{l_1}, \frac{l_3}{l_2})}{\|(\frac{-l_3}{l_1}, \frac{l_3}{l_2})\|} \tag{4.2}$$

where $\mathbf{l}$ can be obtained from $\mathbf{x}$ and $\mathbf{F}$ using an arbitrary line $\mathbf{k}'$ not going through the second image's epipole as:

$$\mathbf{l} = \mathbf{F}^T\mathbf{k}' \times \mathbf{F}\mathbf{x} \tag{4.3}$$

Thus, in a pair of images for which the epipolar geometry is known, a point having a high epipolar gradient in one image should have a high epipolar gradient in the other image as well. Of course, the stability can be limited by the importance of the change in angles between epipolar lines and image edges caused by changes in viewpoint. Nevertheless, a moderate change in this angle will not significantly reduce epipolar gradients. Imposing a threshold on the magnitude of these epipolar gradients will lead to a set of features that are reasonably robust to change in viewpoint. Furthermore, these points will be found on strong image edges with orientations perpendicular to epipolar lines. This is a desirable property, as there is an ambiguity in attempting to match other points, such as those that are in low contrast areas, or on contours

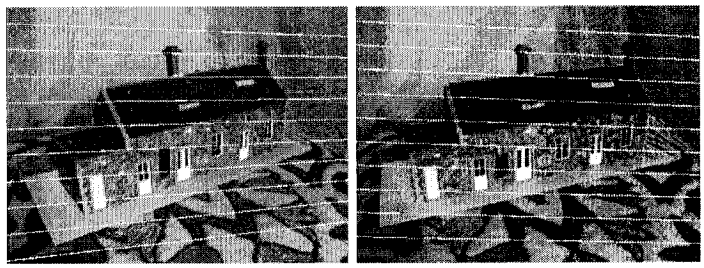


Figure 4.3: An image pair's epipolar geometry.

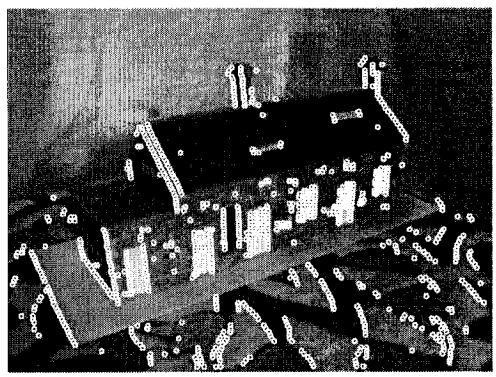


Figure 4.4: The detected epipolar gradient feature points.

which are oriented along epipolar lines.

It should be noted that points with high epipolar gradients can form thick clusters along image edges. To clean up the sets of epipolar feature points, only local maxima in the direction of the epipolar lines should be kept.

Figure 4.4 shows the detected epipolar gradient feature points on one image of the test pair whose epipolar geometry is depicted in Figure 4.3. Note that the feature points that were selected are indeed found on strong images edges that are approximately close to perpendicular to epipolar lines. In one image, points may be detected only on every few lines, to limit their number.

It can be empirically noticed, when epipolar gradient feature points are compared to the same number of Harris feature points, that the former are usually more evenly distributed among the different scene surfaces. This is due to the fact that edges

are often more evenly distributed than high curvature points in typical images. In Figure 4.1, for example, Harris features are mostly concentrated on the house's front wall, while epipolar gradient features in Figure 4.4 are often found on the boundaries between different surfaces which should allow them to quickly yield a more complete reconstruction of scene objects.

4.4 Matching the Feature Points

Now that feature points suitable for matching have been selected, these points must be matched. The common VNC-based approach gives good results when the difference between viewpoints is limited, and was used in Subsection 2.4.2. In the context of this chapter, however, only point triplets agreeing with the trinocular geometry are being compared. Furthermore, the first two points are always epipolar gradient feature points. Thus, a less discriminating similarity measure could be used, if it allows a gain in performance.

4.4.1 Edge Transfer

The similarity measure presented here is based on the consistency of edge orientations between views. Some authors have proposed imposing a bound on edge orientations between views to constrain matching, as described in Subsection 2.5.1, such as Deriche et al. [21]. However, their constraints are only satisfied in the case of very small changes in viewpoint. Horaud and Monga [44] have presented an orientation constraint which measures the consistency of the change in angle with the change in viewpoint, but requires the camera projection matrices. For the constraint presented here, only the system's trinocular geometry is needed.

Two descriptors will be used, together with a simple similarity metric. The most important descriptor is based on the transfer of lines perpendicular to intensity gradients from the first two images to the third one. Other similarity measures which use gradient directions have been proposed by Lowe [66] or Mikolajczyk and Schmid [73].

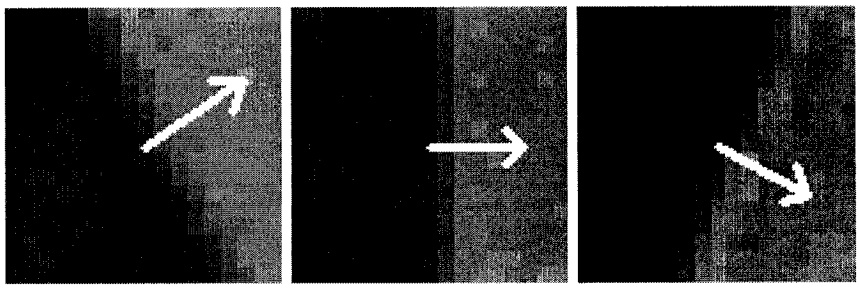


Figure 4.5: Three corresponding image regions, with the image intensity gradient at the center point. The edge transfer similarity measure checks that the lines perpendicular to these gradient vectors are consistent with the trinocular geometry.

These methods rotate one view to align gradients, before further comparing feature point neighborhoods, thus achieving invariance to image rotation. Here, by using trifocal geometry together with the intensity gradients, a higher degree of invariance will be achieved.

The *edge transfer* similarity measure is based on the fact that, using the edge orientation computed at corresponding points in two images, the orientation of the edge at the corresponding point in a third image can be computed using Equation (4.1). This is because the lines going through the points and tangent to the image edges (the lines perpendicular to the intensity gradients at the points) should correspond (see Figure 4.5). Thus, a measure of similarity between three points would be the difference in orientation between the tangent to the edge of one of the points, and the line obtained by transferring the tangent to the edges of the two other points.

The other descriptor that is used is based on the image intensity values in the area around points. The average intensities on each side of the tangent to the edge going through the point are considered. These values should be preserved in different views of the same point taken simultaneously. First, it is determined which side of the edge corresponds to which in the other image. Then, the measure of similarity is taken as the difference between the two average intensities of the most different corresponding sides.

Let $\Delta I(\mathbf{x}, \mathbf{x}', \mathbf{x}'')$ be the maximum difference between the intensities on corresponding sides of edges going through $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{x}''$, and $\Delta\theta(\mathbf{x}, \mathbf{x}', \mathbf{x}'')$ be the difference between the angle in the direction perpendicular to the intensity gradient orientation at $\mathbf{x}''$ measured in $\mathbf{I}''$, and the angle computed from the gradient orientation at $\mathbf{x}$ and $\mathbf{x}'$. Then, the chosen similarity measure between $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{x}''$ will be:

$$s(\mathbf{x}, \mathbf{x}', \mathbf{x}'') = \max\left(\frac{\Delta I(\mathbf{x}, \mathbf{x}', \mathbf{x}'')}{\sigma_{\Delta I}}, \frac{\Delta\theta(\mathbf{x}, \mathbf{x}', \mathbf{x}'')}{\sigma_{\Delta\theta}}\right) \tag{4.4}$$

where $\sigma_{\Delta I}$ and $\sigma_{\Delta\theta}$, the standard deviations of the descriptors, are used to normalize the descriptors to a similar range. This measure is related to the Mahalanobis distance that will be described in Subsection 5.8.1, where the two descriptors are assumed to be uncorrelated, and based on the l_∞ norm instead of the usual vector norm.

4.4.2 Displacement Consistency Constraint

Sometimes, the similarity measure presented in the previous section is not very discriminating. Consequently, even when the search for matches is guided by the trinocular geometry, mismatches should still be expected. However, mismatches can be very undesirable in several applications. Fortunately, when many matches are identified throughout the images, and mismatches are relatively few, they can generally be eliminated by enforcing the disparity gradient constraint introduced in Subsection 2.6.2, as it was used in Subsection 3.4.

4.4.3 Algorithm

Here is a description of the calibrated matching algorithm described in this chapter.

1. Epipolar gradient features are detected in images 1 and 2.
2. For each feature point in image 1, the epipolar line in image 2 is computed.
3. All feature points in image 2 that lie on the epipolar line are considered. For each of these considered pairs, the position of a point in image 3 is computed by trinocular transfer.

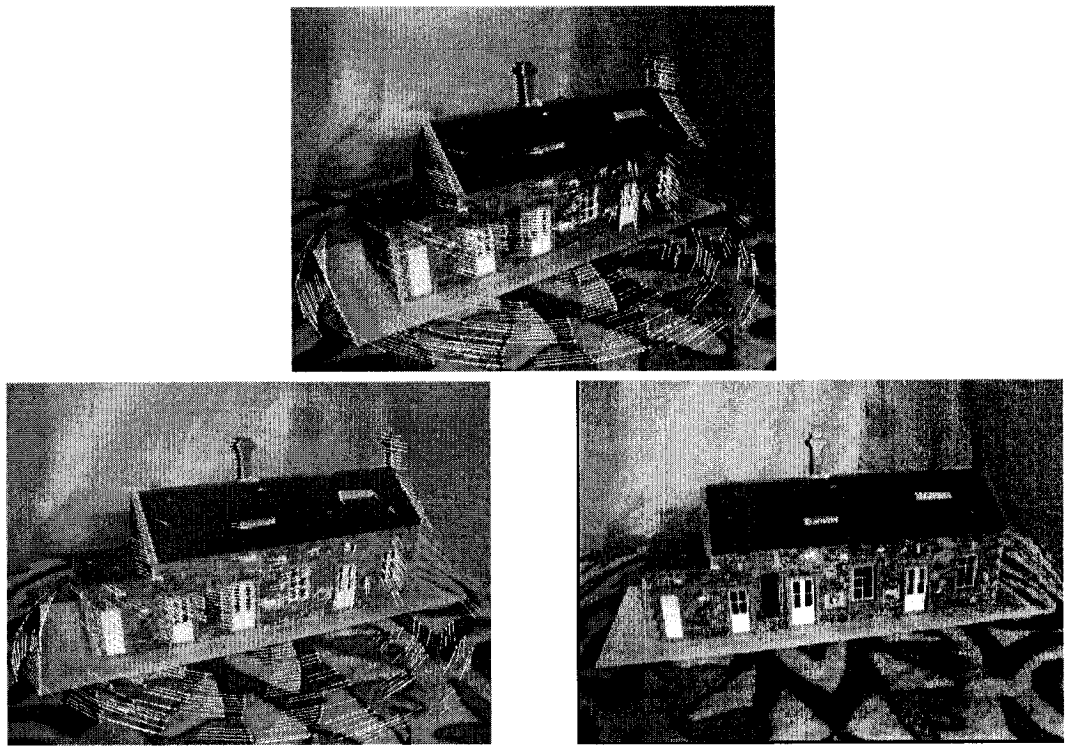


Figure 4.6: Matches found between epipolar gradient feature points. The displacement vectors for matches between images 1-2 are shown on image 1 (bottom-left), and the ones between images 2-3 are shown on image 2 (top).

4. Of the possible point triplets for a point in image 1, only the one giving the highest measure of similarity according to Equation (4.4) is kept, if the measure is above a given threshold.

5. The set of detected point triplets is filtered with the disparity gradient constraint.

4.5 Experimental Results

Figure 4.6 shows the result of applying the proposed matching scheme to an image triplet. The displacement vectors for matched feature points between the bottom-left

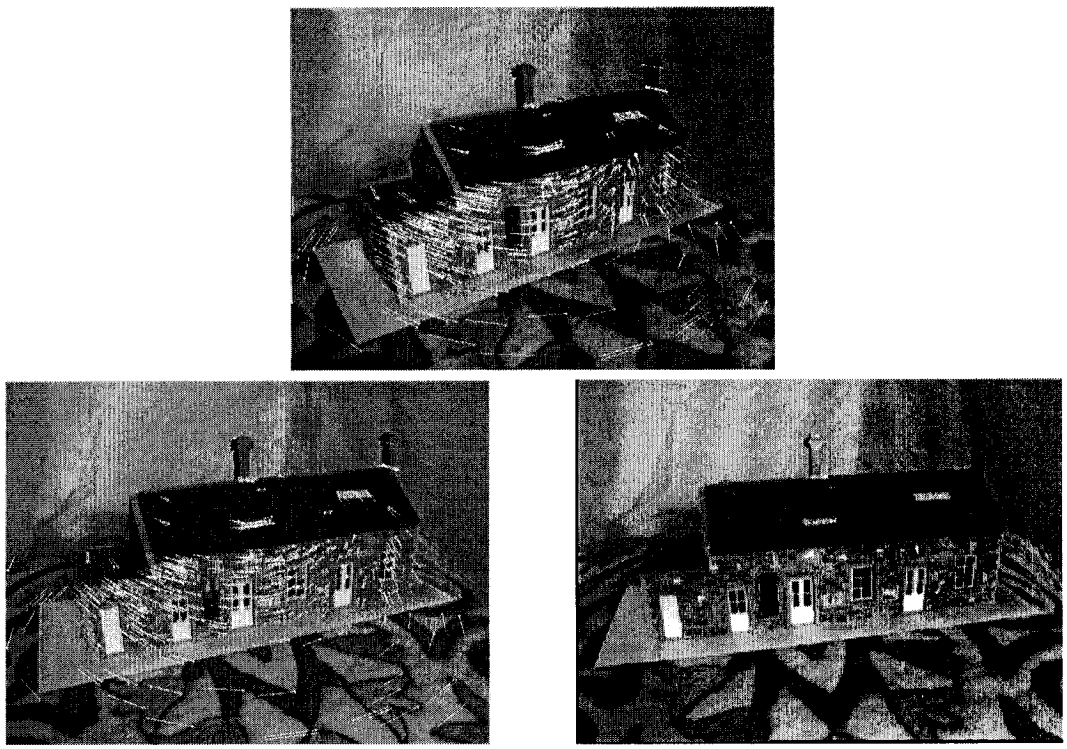


Figure 4.7: Matches found between Harris feature points. The displacement vectors for matches between images 1-2 are shown on image 1 (bottom-left), and the ones between images 2-3 are shown on image 2 (top).

and top images are drawn on the bottom-left image, and the ones between the top and bottom-right are drawn on the top image.

Figure 4.7 shows the displacement vectors obtained when a Harris detector and correlation are used instead, as in Chapter 3. The same number of feature points were used in both experiments, and the thresholds relevant to the matching process were chosen to maximize the resulting number of matches. It can be seen that the method based on epipolar gradients resulted in more matches (601 versus 423), and provides scene features which are more relevant to scene reconstruction. The matches obtained through the Harris detector are mostly located on the front wall of the house, while the matches obtained using the epipolar gradient features are distributed more evenly

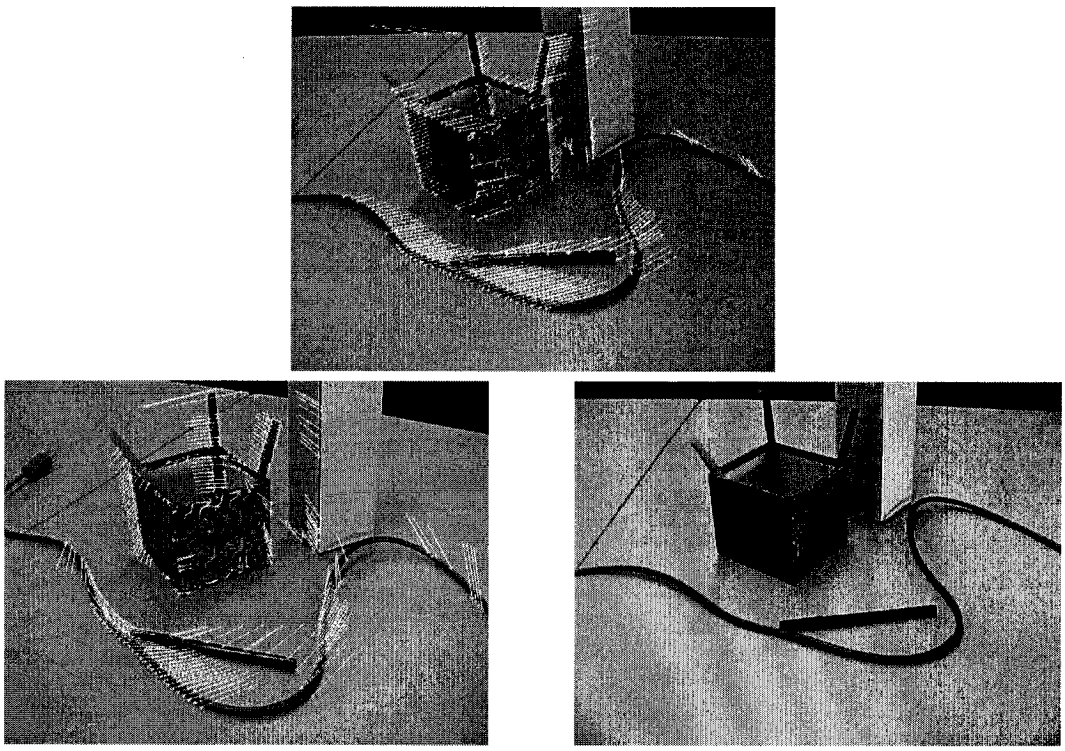


Figure 4.8: Matches found between epipolar gradient feature points. The displacement vectors for matches between images 1-2 are shown on image 1 (bottom-left), and the ones between images 2-3 are shown on image 2 (top).

among the different surfaces, and often lie on the borders between them.

Figures 4.8 and 4.9 also show matches found using the proposed approach and the Harris/correlation approach respectively on simple images of a few objects. Displacements between corresponding feature points are also shown. Here, 273 matches were found using the proposed method. With the same number of feature points, the Harris detector with correlation only found 64, and it was not possible to modify the thresholds to accept more matches without introducing a significant number of mismatches. The success of the proposed method, in contrast to the Harris/correlation approach can be attributed to the fact that the scene objects contain few clear corners, and little textural information, but still enough significant edges to permit their

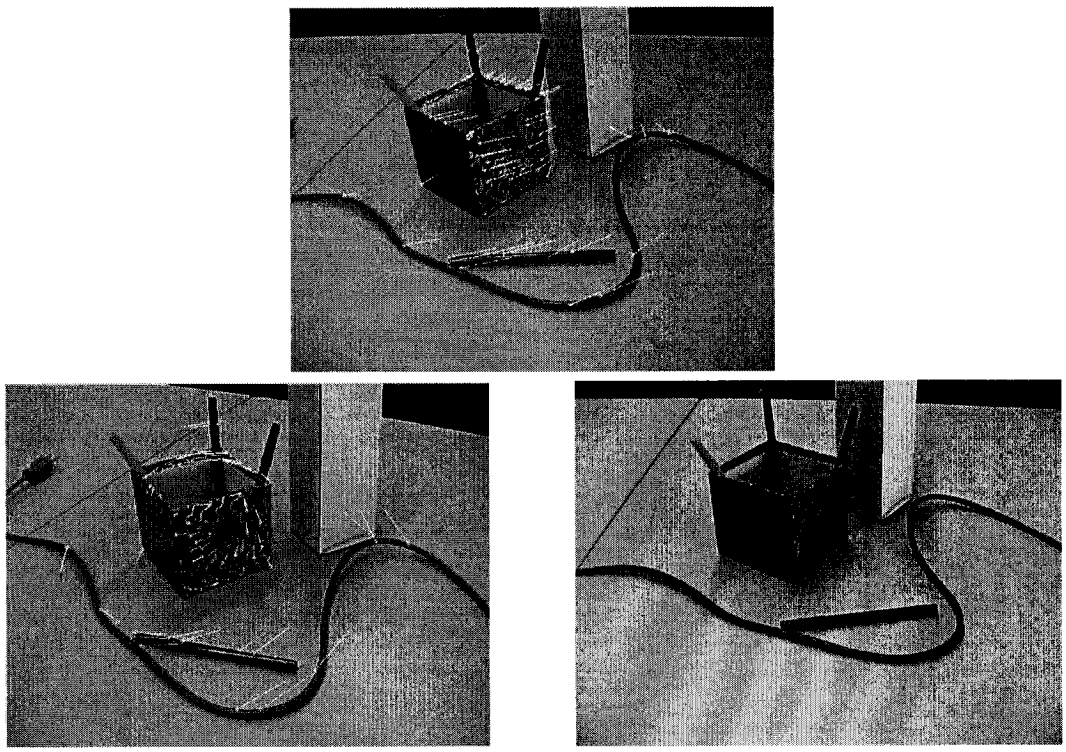


Figure 4.9: Matches found between Harris feature points. The displacement vectors for matches between images 1-2 are shown on image 1 (bottom-left), and the ones between images 2-3 are shown on image 2 (top).

detection as epipolar gradient features.

4.6 Reconstruction

To demonstrate the usefulness of the proposed matching approach to fast model building, the correspondences shown in Figures 4.6 and 4.7 were used, together with the known calibration parameters, to construct models of the scene. The position of scene points in space were computed as the intersection of backprojected rays from the image points. Then, points from common planar surfaces were used to estimate the equation of these planes in space, and the sections of these planes defined by the

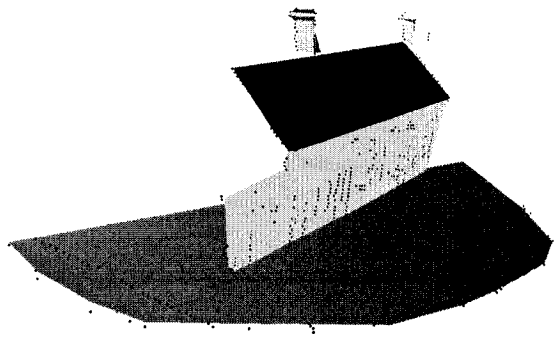


Figure 4.10: A model constructed from the matches shown in Figure 4.6 that were found using epipolar gradient feature points.

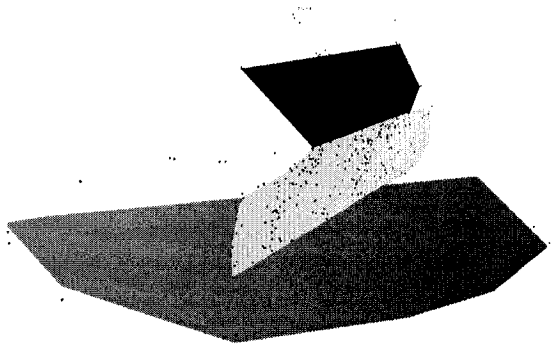


Figure 4.11: A model constructed from the matches shown in Figure 4.7 that were found using Harris feature points.

points were drawn in Figure 4.11 and 4.10. These figures also show, as black dots, the computed locations of matched points.

It can be seen that the model generated from the proposed matching method is far more representative of the scene, mainly as the points used to generate it contain more relevant information. These are distributed more evenly among scene objects, thus allowing for instance the drawing of some parts of the chimneys. They also define more precisely the borders of objects, as it can be seen, for example, that the areas of the front wall and roof do not cover the entire areas of the actual front wall and roof in the other model.

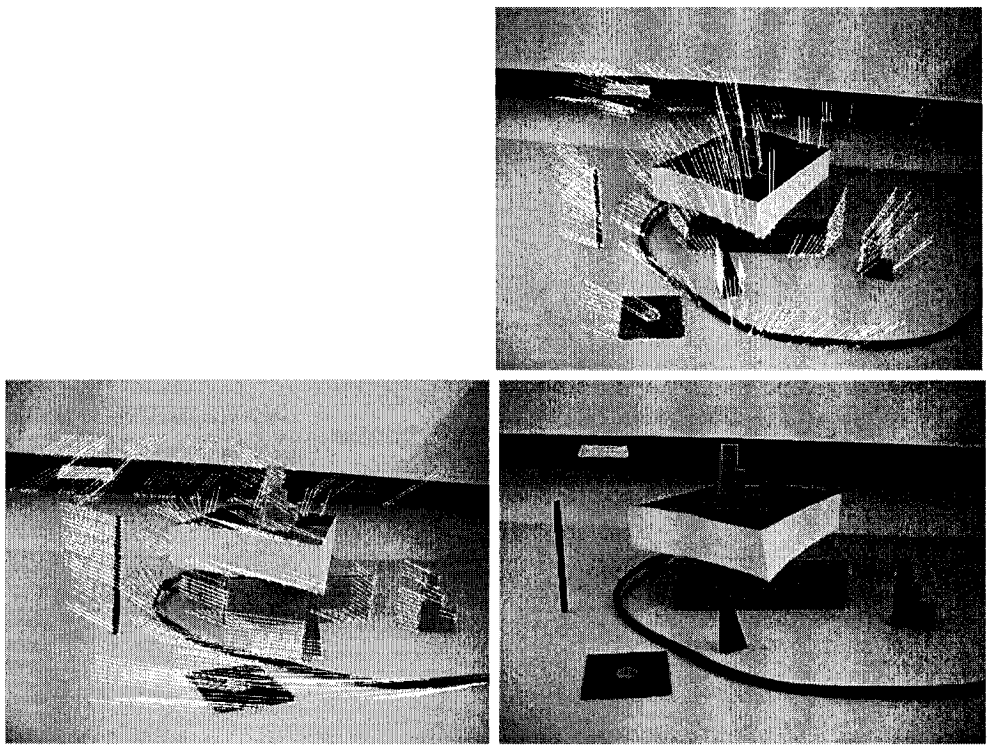


Figure 4.12: Matches between epipolar gradient feature points that were detected in two directions, with the cameras in a L-shape configuration. Displacement vectors between the bottom-left/bottom-right images are shown on the bottom-left image, and between the top/bottom-right images, on the top image.

4.7 Camera Configurations

So far, one apparent weakness of the matching scheme described in this chapter is that it only finds matches among points that are located on edges that are more or less perpendicular to epipolar lines. To obtain a more complete model from these matches, points on other edges would be desirable. These will be easily obtained, although at some computational cost, as long as the camera configuration is carefully selected.

Figure 4.12 shows pictures taken from viewpoints in an *L-shape* configuration. In the bottom-right view, the epipolar lines from the bottom-left view had a generally

horizontal orientation, while the epipolar lines from the top view had a generally vertical orientation. Thus, by running the proposed matching scheme twice, using epipolar geometries between the bottom-left/bottom-right, and then top/bottom-right views, matches were found on image edges orientated in two main directions.

4.8 Conclusion

In this chapter, a method for finding correspondences between calibrated narrowly separated views was described. The method was shown to be especially useful when the correspondences are needed for fast reconstruction. The main contributions of this work are:

- The introduction of epipolar gradient features, which significantly increase the number, and the quality, of the distribution of matched points.
- The introduction of the idea of edge transfer, as a similarity measure between feature points. This measure is invariant to image perspective deformation and fast to compute.
- The experimental demonstration that the matching scheme based on epipolar gradient features and edge transfer finds more and better distributed matches than a scheme based on the Harris detector and correlation.
- The experimental demonstration that this matching scheme is more appropriate for quick reconstruction from sparse matches.

Part III

Matching Uncalibrated Widely Separated Views

Chapter 5

A Survey of Widely Separated View Matching

5.1 Introduction

Widely separated views normally refers to views taken from viewpoints that are separated by a significant baseline, and between which there is a large rotation. It is still considered an open problem, to which no fully satisfactory solution exists. Nevertheless, several partial solutions can be found in the literature, and are briefly described in this chapter.

The most important characteristic of feature-based matching methods, in the context of widely separated views, is the manner in which they compare features to determine similarity. These methods must be robust to the large perspective distortion that may exist between features. Two main approaches exist. The first one consists in using descriptions of the features that are partially invariant to perspective distortion. The main weakness of this approach is that making descriptors increasingly invariant necessarily reduces their discriminating power. This introduces increased ambiguity in matching, and consequently, eventually increases the number of mismatches. The other approach consists in finding an approximation of the perspective distortion of

image regions around the features. Then these regions can be normalized, as to eliminate the distortion. However, such an approach is necessarily costly, as every pair of feature points must be normalized differently.

5.1.1 Summary

The extensiveness of the literature review presented here, and the fact that widely separated view matching is a very active field of research, warrants its inclusion as an independent chapter. Some new approaches to the widely separated views problem will then be presented in Chapters 6 and 7.

A version of the literature review presented in this chapter will be published in:

> Etienne Vincent, Robert Laganière and Gerhard Roth,
> **Widely Separated View Matching : A Literature Review,**
> to be published in *in Image and Vision Computing Journal.*

Section 5.2 considers the problem of feature detection in the widely separated view context. The following sections explore how robustness to different classes of neighborhood transformations can be achieved. Section 5.4 looks at changes in scale. Sections 5.5, 5.6 and 5.7, look at different refinements of perspective deformation: similarities, affinities, and homographies. The metrics used to compare invariant characterizations of feature points are then studied in Section 5.8. Section 5.9 looks at changes in illumination. Finally, the robust estimation of a camera system's geometry is examined in Section 5.10.

5.2 Feature Detection

A first consideration, in widely separated view feature matching, is the choice of a feature extractor. Most often, feature points are used. In Section 1.7, the important properties of feature point detectors were enumerated.

The most widely used feature point detector, when it comes to matching, is the Harris detector [39], which was described in Subsection 2.3.1. This section will describe other feature detectors that are partially invariant to perspective distortion.

5.2.1 Scale Invariant Corner Detection

Lowe [66] claims that the repeatability of Harris detectors is poor when large changes in scale occur. he suggests a feature point detector that selects locations at extrema of a difference of Gaussian function as the solution. To be invariant to changes of scale, the extrema are taken over scale space. This is done by building an image pyramid of difference of Gaussian images, where each level is computed from the resampled version of the Gaussian image used in the previous level. Extrema are obtained by comparing each pixel with its neighbors in the same level, and with the closest pixel location in the level above and below. The size of the feature point neighborhood used to characterize the feature, can then be chosen as proportional to the scale where the extremum was found. The feature points are then compared using the method described in Subsection 5.5.1.

Another option is to use a Harris detector at several scales. Dufournaud et al. [22] use the fact that in the event of a change in scale, image points are related by some linear transformation $\mathbf{x}' = \frac{1}{s}\mathbf{x} + \mathbf{t}$. Thus derivatives are related by $I'_x = sI_x$, and $I'_y = sI_y$. This implies that at a scale factor s, the Matrix (2.1) used to define the Harris detector in Subsection 2.3.1 can be replaced by:

$$C(x, y, s) = s^2 S * (\nabla I(x, y) \cdot \nabla I^\top(x, y)) \tag{5.1}$$

Here the standard deviation of the Gaussian kernels used for differentiation must also be multiplied by s. However, a feature point must now be compared with all the points detected in another image at several different scales, thus greatly increasing the complexity of the search for correspondence.

Mikolajczyk and Schmid [74] present a feature point detector that achieves invariance to affine transformations. An iterative scheme is used, which is initialized using

the method described by Dufournaud et al. described above. Feature points are successively recomputed along with an affinely invariant neighborhood around them. At each step, the feature point's neighborhood is normalized using the second moment matrix as done by Lindeberg and Gårding in [63] (see Subsection 5.6.2). The scale of the feature point is then determined using scale selection, as done by Lindeberg in [62] (see Subsection 5.4.2). Finally, the location of the point is refined as a local maximum of the Harris detector.

5.2.2 Other Features

Matas et al. introduced a type of feature which they call maximally stable external regions (MSERs). These are the connected components of images containing pixels which are all significantly darker (or brighter) than the pixels on the region's border. MSERs are good candidates for matching, as they are naturally invariant to perspective distortion, and monotonic image intensity changes.

Tuytelaars and Van Gool [111] noticed that since corners often lie close to the borders of scene objects, their neighborhoods are likely to undergo non-homographic transformations when there is a change in viewpoint. They therefore suggest using local image intensity extrema, as feature points, in addition to Harris features. These are detected through non-maximum suppression following image smoothing. These features are less likely to be localized accurately, but are also unlikely to lie on non-planar surfaces. These local intensity extrema are then matched using a method described in Subsection 5.6.2.

Alhichri and Kamel [2] use a different approach. Their features consist of empty spaces between edges. Specifically, they use virtual circles, i.e. the circles of locally maximal radii that do not contain edge points. These features are stable, only if the ones whose circumference touch the most edges are used. Drawbacks of these features are that they require stable edge detection, and that they can only be used to model similarity transformations between the images, as more general image transformations will not preserve circularity.

5.3 Types of Neighborhood Transformations

In Section 1.3, different classes of approximation for local perspective distortion were described: homographies, affinities, and similarities. Similarities can be further decomposed into a translation, change in scale, and rotation. The translation is determined by the relative position of corresponding feature points, so only the change in scale and rotation remain to be dealt with. Robustness to these are studied in Sections 5.4 and 5.5 respectively. Then, robustness to affinities is reviewed in Section 5.6, and to homographies in Section 5.7.

Of course, not every feature point neighborhood is planar. Thus not all perspective distortions can be approximated as a homography. A change in viewpoint might cause all or part of a feature point's image neighborhood to become occluded or severely deformed. Unfortunately, it would be nearly impossible to make a matching process tolerant to such effects. The number of Degrees of freedom (DOF) to which the process would need to be invariant would be so high as to make the resulting comparison meaningless.

Some attempts to deal with more complicated transformations of specific types can nevertheless be found in the tracking literature. For instance, a common case is the change of the background of an object due to a change in viewpoint, which is studied by Darrell in [18]. In this case, feature point neighborhoods must be segmented and correlation is applied to the foreground region only.

5.4 Robustness to Changes in Scale

In the context of applications such as 3D reconstruction, or robot navigation, change in scale is often insignificant, unlike in other applications such as image retrieval, or object recognition. Scale changes are commonly treated separately from other components of perspective distortion. There are two general approaches to matching over different scales. A feature can be compared to others at several scales, then, only the scale resulting in the best correspondence is kept. Alternatively, some heuristic

can be used to select a single scale at which to characterize each feature, to be used in the later comparison.

5.4.1 Scale Space

This first approach is implemented by computing the scale space of a given image. This is the image pyramid where the successive levels are blurred and subsampled versions of the preceding ones. Then, correspondence is established by comparing a region of an image, with regions taken at all levels of the scale space in the other. Schmid and Mohr [97] integrate this multiscale approach in their matching algorithm. They compute some differential invariants (see Subsection 5.5.2) that are used to characterize feature points at each level of the scale space. The scale giving the highest similarity score is the one at which the features correspond.

Hansen and Morse [36] prefer to use correlation between the scale traces of points. The scale trace is a vector of pixel intensities measured at different scales. The scale trace of corresponding points, in images that differ in scale only, should be shifted (or scaled) versions of each other. Thus, correlation between shifted versions of these scale traces will peak with the right shift. The scale trace can also be computed for measures other than pixel intensities, such as gradient intensities, or derivatives in the x and y directions.

5.4.2 Scale Selection

The other approach to obtaining robustness to changes in scale, referred to as *scale selection*, was first advocated by Lindeberg [62]. He suggests a heuristic principle to select the *natural scale* which characterizes a feature point neighborhood. Normally, the amplitude of derivatives decreases with the scale at which they are computed. However, when the scale is known, derivatives can be expressed in a coordinate systems normalized by it. Lindeberg proposes to evaluate a non-linear combination of normalized derivatives, at different scales, on the feature points. The natural scale used to characterize a feature point is then the one where an extremum is reached.

The idea is that the graph of the combination of derivatives evaluated at some feature point, over different scales, should be a scaled version of the graph generated for the same feature point, but viewed at a different scale in another image. Thus, the ratio of the scales where extrema are reached, in these graphs, should correspond to the ratio of scales between the views. Hence, the scales where extrema are reached naturally describe the points. The choice of combination of normalized derivatives used by Lindeberg depends on the application.

Baumberg [6], uses this approach to determine the scale of feature points using the determinant and trace of Matrix (2.1) as the combination of scale-normalized derivatives. Mikolajczyk and Schmid [73] use local extrema of a scale-normalized Laplacian instead:

$$Lap(\mathbf{x}, \sigma) = \sigma^2(I_{xx}(\mathbf{x}, \sigma) + I_{yy}(\mathbf{x}, \sigma)) \tag{5.2}$$

where the derivatives at scale σ are taken by convolution with derivatives of Gaussian filters of standard deviation σ. There, the scale-normalized Laplacian was compared to a few other functions, and found to be the best one for feature point scale selection. This normalized Laplacian is also used for scale selection by Hall et al. in [35]. The idea of scale selection is also exploited by Fdez-Valdivia et al. in [25], where Gabor filters are used to identify the natural scale.

5.5 Robustness to Local Similarities

Many authors use invariance to similarities as an approximation for invariance to changes in viewpoint. As seen previously, a similarity consists in a translation, rotation, and change of scale. Matching, itself, essentially consists in finding the translation of a feature point between two views, and in the previous section, changes in scale were investigated, so all that remains is making the matching process robust to image rotation. This section reviews ways of achieving this.

5.5.1 Alignment

The simplest way of obtaining invariance to local rotation is to normalize the local neighborhoods, by rotating them to align the intensity gradients. For instance, Mikolajczyk and Schmid [73], and Hall et al. [35] use Gaussian derivatives in their characterization of feature point neighborhoods. To achieve invariance to rotation, derivative filters are steered in the direction of the local intensity gradient, which is obtained as the peak of a histogram of local gradient orientations. Steerable filters are described in [27], where it is shown how to compute filters with arbitrary directions as a linear combination of a set of basis filters. i.e. A Gaussian derivative filter G_i^θ of order i, and steered in the direction θ can be expressed as:

$$G_i^\theta = \sum_{j=1}^{N} k_j(\theta)\, G_i^j \tag{5.3}$$

where N is usually a small number which depends on i (2 for $i = 1$, 3 for $i = 2,\ldots$), the G_i^j are N basis filters of order i, and the $k_j(\theta)$ are multipliers for basis vectors which depend on the order and direction.

Lowe [66], who is interested in object recognition, presents a similarity measure which is modeled on the response of complex cells in the primary visual cortex. Feature points are described by image gradients, and evaluated on their neighboring pixels. To achieve invariance to rotation, the gradients are computed in a coordinate frame determined by the gradient direction at the feature point. By looking only at gradients, he claims that, as in the primate vision system, the obtained characterization is partially invariant to affine image transformations. Robustness to changes in local geometry is achieved by using many orientation planes. However, this gives the feature point characterization a perhaps unnecessarily high dimension.

Tuytelaars and Van Gool [111] align a different measure instead. They first produce affinely invariant elliptical regions around feature points (as described in Subsection 5.6.2), and then normalize these to circles. But before comparing the circles, a rotation is found to align them. A photometrically invariant version of the major axis of inertia is used. It is defined as the line passing through the center of the circle

at an angle θ satisfying:

$$\tan^2(\theta) + \frac{M_{20} - M_{02}}{M_{11}} \tan(\theta) - 1 = 0 \tag{5.4}$$

where $M_{pq} = \int I(x,y) x^p y^q \, dx dy$ is the moment of degree n and order $(p+q)$ taken over the circular region.

5.5.2 Differential Invariants

The idea of matching differential invariants evaluated at feature points is based on the work of Hilbert [43], who shows that any differential invariant of finite order can be expressed as a polynomial on a set of irreducible invariants.

Montesinos et al. [77] claim that the following invariants work reasonably well for gray level images:

$$I, \quad I_{\eta\eta} + I_{\xi\xi}, \quad I_\eta, \quad \frac{I_{\xi\xi}}{I_\eta}, \quad \frac{I_{\xi\eta}}{I_\eta} \tag{5.5}$$

where the η, ξ coordinate frame is defined by the unit vector $\eta = \frac{\nabla I}{\|\nabla I\|}$ and such that $\xi \perp \eta$.

Schmid et al. [97] use differential invariants to match interest points between images for the purpose of image retrieval. Each feature point is described by a vector of 9 differential quantities that are invariant to image rotation, and given as follows (using Einstein summation convention):

$$\begin{gathered} I, \quad I_i I_i, \quad I_i I_{ij} I_j, \quad I_{ii}, \quad I_{ij} I_{ji}, \\ \epsilon_{ij}(I_{jkl} I_i L_k L_l - I_{jkk} I_i I_l I_l), \quad I_{iij} I_j I_k I_k - I_{ijk} I_i I_j I_k, \quad -\epsilon_{ij} I_{jkl} I_i I_k I_l, \quad I_{ijk} I_i I_j I_k \end{gathered} \tag{5.6}$$

where ϵ is the 2D antisymmetric epsilon tensor, and with $i, j, k, l \in \{x, y\}$. Gaussian kernels must be employed for the computation of the derivatives, otherwise they are very unstable. The main drawback is that these kernels are then very large for higher derivatives, and might cover too large areas of the image. The technique might work well over large planar objects, but not necessarily elsewhere.

Montesinos et al. [77] advocate the use of color information. The following invariants are used which only use first order derivatives:

$$R, \quad \|\nabla R\|^2, \quad G, \quad \|\nabla G\|^2, \quad B, \quad \|\nabla B\|^2, \quad \nabla R \cdot \nabla G, \quad \nabla R \cdot \nabla B \tag{5.7}$$

They claim that first order differential invariants are sufficient in this case because of the additional color information. The matching process is thus faster, and much more robust to noise. The authors also claim that their process is robust to changes in viewpoint, but only really demonstrate it on narrowly separated image pairs, and pairs that differ mostly by simple image rotation.

According to Baumberg [6], rotational differential invariants have not been demonstrated to work with wide changes in viewpoint. Another drawback of these characterizations, also valid for other approaches based on the comparison of invariant descriptors, is that reducing a feature point neighborhood's description to a small number of scalar values reduces greatly the discriminating power of the comparison.

5.5.3 Window Functions

Baumberg [6] uses a variant of the Fourier-Mellin transformation. He computes a set of complex-valued quantities $u_{n,m}$ over each feature point neighborhood, and in each color plane:

$$u_{n,m} = \int \frac{\mathrm{d}^n}{\mathrm{d}r^n} G_\sigma(r)\, e^{i\phi m}\, I(r,\phi)\, r\, \mathrm{d}r\mathrm{d}\phi \tag{5.8}$$

Where G_σ is a Gaussian window function, and (r,ϕ) are polar coordinates centered on the feature point. A rotation of angle θ results in a scaling by a factor $e^{i\theta m}$ of the coefficients. Thus, a rotation invariant characterization is obtained by normalizing the $u_{n,m}$ coefficients by $\frac{u_{0,m}}{|u_{0,m}|}$. Baumberg used thirteen such rotation invariants, resulting from different values for n and m (43 on color images).

Schmid [96] uses convolution by a set of isotropic operators similar to Gabor filters, which are spatial bandpass filters, combining frequency and scale as the product of Gaussian and cosine transform filters:

$$F(x,y,\tau,\sigma) = F_0(\tau,\sigma) + \cos\left(\frac{\pi\tau\sqrt{x^2+y^2}}{\sigma}\right) e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5.9}$$

Where σ represents the scale, and τ is the number of cycles of the harmonic function within the Gaussian envelope of the filter. Thirteen such filters are used, obtained

with different values for τ and σ. The result is a rotation invariant characterization. Schmid states that these invariants provide an improved performance over her differential rotation invariants described in Subsection 5.5.2.

5.5.4 Consistency of Neighboring Pairs

Jung and Lacroix [51], avoid the problem of finding an invariant characterization by considering only the intensity of feature points themselves, and not their neighborhood. However, they require that the displacement of adjacent corresponding points agree with some similarity transformation. Initially, only feature point cornerness, defined as the ratio of the eigenvalues of Matrix (2.1) are compared, as described in Subsection 2.5.1. Then, similarity transformations are computed for groups of adjacent matched feature points. Correspondences are considered valid when they are surrounded by many other pairs agreeing with the estimated transformation. A correlation score is also used, but only considering the pixel intensities of feature points in the groups.

5.6 Robustness to Local Affinities

Making a matching process invariant to local similarities might not be enough to achieve sufficient robustness in many applications. Local affinities are more general, and thus provide a better approximation of perspective deformation. For this reason, several methods which are robust to affine transformations have been proposed in the literature.

5.6.1 Invariant Descriptors

In [95], Schmid suggests the use of affine differential invariants given, using Einstein summation convention, as:

$$I,\quad I_iI_j(H^{-1})^{ij},\quad I_{ijl}I_lI_mL_n(I^{-1})^{il}(I^{-1})^{jm}(I^{-1})^{kn},\quad I_{ijk}I_l(I^{-1})^{ij}(I^{-1})^{kl},\\ I_{ijk}I_{lmn}(I^{-1})^{ij}(I^{-1})^{kl}(I^{-1})^{mn},\quad I_{ijk}I_{lmn}(I^{-1})^{il}(I^{-1})^{jm}(I^{-1})^{kn} \tag{5.10}$$

where H^{-1} is the inverse of the Hessian, and with $i,j,k,l,m,n \in \{x,y\}$. However, Schmid explains that these invariants are too difficult to incorporate into a multiscale framework, and thus discards them.

Mindru et al. [75] present affine moment invariants for image regions with the intended goal of identifying color labels and logos. A general color moment is defined as:

$$M_{pq}^{abc} = \iint x^p\, y^q\, R(x,y)^a\, G(x,y)^b\, B(x,y)^c\, \mathrm{d}x\mathrm{d}y \tag{5.11}$$

where the integration is over the considered image region, and, $R(x,y)$, $G(x,y)$, and $B(x,y)$ are the intensities of the red, green and blue channels respectively at coordinates (x,y). Only invariants made of first order and second degree moments were considered (i.e. $p+q \leq 1$ and $a+b+c \leq 2$) since moments of higher order or degree would be unstable. Twenty four rational functions of the general color moments are used to describe feature point neighborhoods in an affinely invariant way:

$$\begin{array}{c}
\frac{M_{00}^2 M_{00}^0}{(M_{00}^1)^2},\quad \frac{M_{00}^{11} M_{00}^{00}}{M_{00}^{10} M_{00}^{01}},\\
\frac{M_{10}^{10} M_{01}^{01} M_{00}^{00} + M_{10}^{01} M_{01}^{00} M_{00}^{10} + M_{10}^{00} M_{01}^{10} M_{00}^{01}}{M_{00}^{10} M_{00}^{01} M_{00}^{00}} \;_\; \frac{M_{10}^{10} M_{01}^{00} M_{00}^{01} + M_{10}^{01} M_{01}^{10} M_{00}^{00} + M_{10}^{00} M_{01}^{01} M_{00}^{10}}{M_{00}^{10} M_{00}^{01} M_{00}^{00}},\\
\frac{M_{10}^2 M_{01}^0 M_{00}^1 + M_{10}^1 M_{01}^2 M_{00}^0 + M_{10}^0 M_{01}^1 M_{00}^2}{M_{00}^2 M_{00}^1 M_{00}^0} \;_\; \frac{M_{10}^2 M_{01}^1 M_{00}^0 + M_{10}^1 M_{01}^0 M_{00}^2 + M_{10}^0 M_{01}^2 M_{00}^1}{M_{00}^2 M_{00}^1 M_{00}^0},\\
\frac{M_{10}^{11} M_{01}^{00} M_{00}^{10} + M_{10}^{10} M_{01}^{11} M_{00}^{00} + M_{10}^{00} M_{01}^{10} M_{00}^{11}}{M_{00}^{11} M_{00}^{10} M_{00}^{00}} \;_\; \frac{M_{10}^{11} M_{01}^{10} M_{00}^{00} + M_{10}^{10} M_{01}^{11} M_{00}^{00} + M_{10}^{00} M_{01}^{11} M_{00}^{10}}{M_{00}^{11} M_{00}^{10} M_{00}^{00}},\\
\frac{M_{10}^{11} M_{01}^{00} M_{00}^{01} + M_{10}^{01} M_{01}^{11} M_{00}^{00} + M_{10}^{00} M_{01}^{01} M_{00}^{11}}{M_{00}^{11} M_{00}^{01} M_{00}^{00}} \;_\; \frac{M_{10}^{11} M_{01}^{01} M_{00}^{00} + M_{10}^{01} M_{01}^{11} M_{00}^{00} + M_{10}^{00} M_{01}^{11} M_{00}^{01}}{M_{00}^{11} M_{00}^{01} M_{00}^{00}},\\
\frac{M_{10}^{02} M_{01}^{00} M_{00}^{10} + M_{10}^{10} M_{01}^{02} M_{00}^{00} + M_{10}^{00} M_{01}^{10} M_{00}^{02}}{M_{00}^{02} M_{00}^{10} M_{00}^{00}} \;_\; \frac{M_{10}^{02} M_{01}^{10} M_{00}^{00} + M_{10}^{10} M_{01}^{02} M_{00}^{00} + M_{10}^{00} M_{01}^{02} M_{00}^{10}}{M_{00}^{02} M_{00}^{10} M_{00}^{00}},\\
\frac{M_{10}^{20} M_{01}^{01} M_{00}^{00} + M_{10}^{01} M_{01}^{00} M_{00}^{20} + M_{10}^{00} M_{01}^{20} M_{00}^{01}}{M_{00}^{02} M_{00}^{10} M_{00}^{00}} \;_\; \frac{M_{10}^{20} M_{01}^{00} M_{00}^{01} + M_{10}^{01} M_{01}^{20} M_{00}^{00} + M_{10}^{00} M_{01}^{01} M_{00}^{20}}{M_{00}^{20} M_{00}^{01} M_{00}^{00}},
\end{array} \tag{5.12}$$

where M_{pq}^{i} stands for M_{pq}^{i00}, M_{pq}^{0i0}, or M_{pq}^{00i}; and M_{pq}^{ij} stands for M_{pq}^{ij0}, M_{pq}^{i0j}, or M_{pq}^{0ij}, depending on which color bands are used. All possible combinations of color bands

are used for the first six invariants, while the last one is only taken on RB, and the second last one only on RG and GB.

Tuytelaars and Van Gool [111] also use a descriptor based on generalized color moments. However, they work on image regions that were previously normalized (as explained in the next subsection). They can therefore use simpler descriptors as their main concern is now robustness to noise. They chose eighteen simple rational functions of the second order and first degree moments:

$$\begin{array}{cccccccccc} \frac{M_{00}^{110}}{M_{00}^{000}}, & \frac{M_{00}^{011}}{M_{00}^{000}}, & \frac{M_{00}^{101}}{M_{00}^{000}}, & \frac{M_{10}^{100}}{M_{00}^{100}}, & \frac{M_{10}^{010}}{M_{00}^{010}}, & \frac{M_{10}^{001}}{M_{00}^{001}}, & \frac{M_{01}^{100}}{M_{00}^{100}}, & \frac{M_{01}^{010}}{M_{00}^{010}}, & \frac{M_{01}^{001}}{M_{00}^{001}}, \\ \frac{M_{11}^{100}}{M_{00}^{100}}, & \frac{M_{11}^{010}}{M_{00}^{010}}, & \frac{M_{11}^{001}}{M_{00}^{001}}, & \frac{M_{20}^{100}}{M_{00}^{100}}, & \frac{M_{20}^{010}}{M_{00}^{010}}, & \frac{M_{20}^{001}}{M_{00}^{001}}, & \frac{M_{02}^{100}}{M_{00}^{100}}, & \frac{M_{02}^{010}}{M_{00}^{010}}, & \frac{M_{02}^{001}}{M_{00}^{001}} \end{array} \tag{5.13}$$

The above descriptors of Mindru et al., and Tuytelaars and Van Gool, are very sensitive to the deformation of image point neighborhoods between views. The moments on which they are based must be computed on precisely corresponding image regions. A way to obtain such regions is described in the next subsection.

5.6.2 Neighborhood Normalization

Tuytelaars et al. [112] use the two edges going through a corner point to select a region around it in an affinely invariant way. That is, they select the regions such that when two of them are on parts of two images related by an affine transformation, the corresponding image regions will be shaped as to be exact transformed versions of each other. This is done by selecting a point along each of the edges going through the corner, and then using the parallelogram defined by the two resulting line segments. The length of the segments are chosen as those where an affinely invariant function reaches an extremum, when computed over the parallelogram. Two examples of functions on which an extremum may be sought are:

$$\frac{\int I(x,y)\,\mathrm{d}x\mathrm{d}y}{\int \mathrm{d}x\mathrm{d}y} \quad \text{and} \quad \frac{\det(\mathbf{p}-\mathbf{q} \;\; \mathbf{p}-\mathbf{p}_g)}{\det(\mathbf{p}-\mathbf{p}_1 \;\; \mathbf{p}-\mathbf{p}_2)} \tag{5.14}$$

where $\mathbf{p}$ is the feature point, $\mathbf{p}_1$ and $\mathbf{p}_2$ are the points defining the line segments on the two edges, $\mathbf{q}$ is the other corner of the parallelogram defined by $\mathbf{p}$, $\mathbf{p}_1$ and $\mathbf{p}_2$, and, $\mathbf{p}_g$ is the center of gravity of the image intensities in the parallelogram. In [111], the following functions are also added:

$$\begin{matrix} \left|\frac{\det(\mathbf{p}_1-\mathbf{p}_g \quad \mathbf{p}_2-\mathbf{p}_g)}{\det(\mathbf{p}-\mathbf{p}_1 \quad \mathbf{p}-\mathbf{p}_2)}\right| \frac{M^1}{\sqrt{M^2M^0-M^1M^1}} \\ \left|\frac{\det(\mathbf{p}_1-\mathbf{p}_g \quad \mathbf{q}-\mathbf{p}_g)}{\det(\mathbf{p}-\mathbf{p}_1 \quad \mathbf{p}-\mathbf{p}_2)}\right| \frac{M^1}{\sqrt{M^2M^0-M^1M^1}} \end{matrix} \tag{5.15}$$

where $M^n = \int I^n \, \mathrm{d}x\mathrm{d}y$ are the moments of degree n and order 0, taken over the parallelogram. These two functions are minimized when p_g is closest to the diagonals of the parallelogram, a situation that is preserved under affine deformation.

Unfortunately, there can be a problem when the extrema of the functions are located on ridges of similar values (as opposed to well defined extrema). This is solved by taking the intersection of two such ridges for different functions. In the case where the edges meeting at the corner point are both curves (as opposed to lines), the process can be simplified. It can be reduced to the simultaneous search for the two points along the edges, from a two dimensional search along both edges. This is done using an affinely invariant way to relate points on both edges, that is by imposing:

$$\int |\det(\mathbf{p}_1'(s_1) \quad \mathbf{p}-\mathbf{p}_1(s_1))| \, \mathrm{d}s_1 = \int |\det(\mathbf{p}_2'(s_2) \quad \mathbf{p}-\mathbf{p}_2(s_2))| \, \mathrm{d}s_2 \tag{5.16}$$

where $\mathbf{p}_1(s_1)$ is a parameterization of the curved edge.

An important problem with the general approach of invariant regions defined by intersecting edges is its reliance on the edges. These can be unstable or absent in some feature point neighborhoods. For this reason, Tuytelaars and Van Gool suggest a complimentary method of selecting image neighborhoods in an affinely invariant manner, when the first one fails. It uses several rays emanating from the feature point, evaluating the following function along each:

$$f_I(t) = \frac{|I(t)-I(0)|}{\max\left(\frac{\int_0^t |I(t)-I(0)| \, \mathrm{d}t}{t}, d\right)} \tag{5.17}$$

where t is the distance from the feature point along the ray, and d is a small constant preventing division by zero. This function reaches a maximum when the intensity changes dramatically along the ray. These maxima are affine invariants. Thus after the maxima are found on all rays, they define an affinely invariant region which is then described by a fitted ellipse. Finally, the ellipse is enlarged to ensure enough contrast within the invariant region. These last regions are especially suitable when intensity extrema are used as the feature points, as described in Subsection 5.2.2.

As seen in the previous subsection, the affinely invariant regions described above can be matched using the moment descriptors of Equations (5.12) or (5.13), after normalization. Correspondences can then also be validated by applying correlation between the normalized regions. Hence, both invariant quantities and intensity correlation measures can be obtained from invariant regions.

Lindeberg and Gårding [63] have developed another method that iteratively finds affinely invariant blob-like feature point neighborhoods. These correspond to the non-isotropic Gaussian filters that make Matrix (2.1) weakly isotropic (proportional to the identity matrix), when they are used as the smoothing operator S for its computation. Non-isotropic Gaussian kernels are generated by:

$$G_\Sigma = \frac{e^{-\frac{1}{2}\mathbf{x}^\top \Sigma^{-1}\mathbf{x}}}{2\pi\sqrt{\det \Sigma}} \tag{5.18}$$

where Σ is a symmetric positive definite 2×2 matrix. When feature point neighborhoods are normalized to make these blob-like regions circular (to make Σ diagonal), the local texture becomes weakly isotropic, meaning that it has a weakly isotropic second moment matrix, and thus, covariance matrix. When the neighborhoods of corresponding feature points are made weakly isotropic in this way, they become similar, up to a rotation. An important drawback of using this technique is that the iterative scheme which it uses might not converge. Baumberg [6] proposes a simplified version of this scheme, but that requires that there be no change in scale between the images. Baumberg then eliminates the rotation which still exists between corresponding feature point neighborhoods using the Fourier-Mellin transformation, as was described in Subsection 5.5.3.

5.6.3 Matching Higher Level Structures

Tell and Carlsson [105] find potential matches using the intensity profiles between feature points. For all pairs of neighboring feature points in the images, the intensity profile, $p(i)$, along the line segment joining the points is extracted and characterized using the following scale invariant Fourier coefficients:

$$\frac{1}{N}\sum_{i=0}^{N-1} p(i)\cdot \sin\frac{2\pi mi}{N} \tag{5.19}$$

where N is the length of the profile, and $m \in \{1,2,3\}$. More coefficients are also obtained with a cosine.

These coefficients are invariant to affine transformations of the line segments, if these segments lie on planar surfaces. Matches can be obtained through a voting scheme using the number of pairs of agreeing intensity profiles. This method seems to work well, however, the authors state that it usually only finds a small number of matches between two images. The obvious limitation is that the method will only match points located on relatively large planar surfaces.

Another approach which uses higher level structures consists in first finding image regions of similar texture, and then matching feature points within these regions. Schaffalizky and Zisserman [92] have presented a method for texture matching which is invariant to affine transformations, and can be used towards feature point matching.

First images are segmented in regions having relatively uniform texture. For a region Ω, the texture is then described with the second moment matrix:

$$M_I = \int_\Omega \begin{pmatrix} I_xI_x & I_xI_y \\ I_xI_y & I_yI_y \end{pmatrix} \frac{\mathrm{d}x\mathrm{d}y}{|\Omega|} \tag{5.20}$$

where $|\Omega|$ denotes the area of Ω. The regions is then normalized, to make its second moment matrix the identity. The normalization A can be obtained from $A^\top A = M_I$. To match normalized regions with similar texture, a rotationally invariant bank of local filters is used, the region's pixels are labelled with filters having the highest response. This is done as did Malik et al. in [70], where directional Gaussian derivative

operators were used with different levels of smoothing. The resulting histograms of filter responses for each pixel are compared to determine similarity. Finally, when it comes to finding corresponding points between matched regions, the affine normalization associated with the regions can be used to normalize the regions before correlation of feature point neighborhoods. A clear drawback, is that the image must contain large regions of stationary texture. Also, segmentation of the image is required as an input, which is hard to obtain automatically.

Berg and Malik [8] chose to use very large windows for matching. They explain that when an image region is affinely distorted, the amount of geometric distortion will vary linearly with distance from the center points. Thus, it is suggested that a Gaussian blur operator should be applied to feature point neighborhoods, where the blur increases with distance from the feature point. Since this blurring technique works better with images containing sparse data, several channels corresponding to oriented edge responses are used instead of the actual images. Results are shown where the geometric blur method proves superior when applied to large (i.e. 81×81) neighborhoods. However, very few matching points are found and they are all localized on very distant objects, that appear only moderately distorted between the views.

5.7 Robustness to Local Homographies

It was seen in the previous two sections how similarities or affinities can be used to approximate the distortion of small planar regions due to changes in viewpoint. However, the true transformation should be a local homography. Nevertheless, as general homographies possess 8 DOF, they are difficult to approximate. For the same reason, it is difficult to make feature point characterizations invariant to them.

One attempt to estimate local homographies was made by Pritchett and Zisserman [85] who were interested in reconstruction. They use local planar homographies to warp image patches around feature points prior to applying correlation. This results

in a similarity measure that is completely invariant to changes in viewpoint. The difficulty, however, lies in estimating the local homographies.Pritchett and Zisserman rely on the computation of homographies induced by sets of four matched coplanar lines. This works well when significant planar objects containing line features are found in the scene. Other special configurations could define local homographies, but the main drawback of such an approach is that it relies on the presence of suitable structures in the scene.

Super and Klarquist [104] have noticed that when there is a large perspective deformation between views, image patches corresponding to a same scene surface are not sampled at the same rate. Hence, some high frequency information in one view might be absent in the other. They solve this by sampling the images in a consistent fashion, according to a hypothesis of the 3D scene structure. To obtain this scene structure hypothesis, they conduct a coarse-to-fine search over all possible planar surface shapes to find the best 3D structure that would generate the two views. They thus determine simultaneously correspondence and structure, for an image pair. This is done through an area based method that matches image patches. However, the extensive search they use to find the best homographies between image patches makes the approach highly impractical.

5.8 Comparing Invariant Vectors

Many techniques described in previous sections introduce invariant characterizations to describe feature points. Typically, these characterizations consist of a vector of invariant attributes. In order to determine correspondence between the features, these vectors must be compared. The choice of the metric used to compare characterization vectors is now addressed.

5.8.1 Mahalanobis Distance

Mahalanobis distances are used in several works from the previous sections, including [73, 97, 105, 111]. For two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, the distance is defined as:

$$d_M(\mathbf{v}_1, \mathbf{v}_2) = \sqrt{(\mathbf{v}_1 - \mathbf{v}_2)^\top \Lambda^{-1} (\mathbf{v}_1 - \mathbf{v}_2)} \tag{5.21}$$

where Λ is the covariance matrix of the vector components. This distance should provide a good measure of the similarity of vectors, as long as an appropriate estimate of Λ is available. Covariance may be empirically estimated by following many feature points over several views, and keeping track of its average resulting value in the characterization vector components. In [112], Tuytelaars et al. use a Mahalanobis distance together with a hashing technique to reduce the search complexity.

Baumberg [6] also uses the Mahalanobis distance, but defines the ambiguity of matches as the difference in this measure when using a point's strongest match, and when using its second strongest match. It is then only this ambiguity measure that is used to determine if two points correspond. Indeed, two points are more likely to represent the same scene feature if they are matched unambiguously. A similar idea was described by Biber and Straßer in [11].

5.8.2 Other Metrics

The estimation of the covariance matrix needed when using Mahalanobis distances requires large training sets. To avoid the need of gathering such data, Montesinos et al. [77] suggest a simpler scheme. They resize the components of feature vectors to a given interval, and use a simple Euclidean distance to measure similarity. They claim that although sub-optimal, this measure produces a high proportion of correct matches when used with their invariant characterization.

For Lowe [66], the characterizations being compared have a much higher dimension, and thus, finding the best match for each feature point becomes prohibitively complex. An approach based on the Hough transform is thus developed to cluster similar features at a larger scale, before finding the best match within the clusters.

5.9 Robustness to Illumination Changes

Objects generally reflect light differently in different directions. Also, if the images being compared are taken at different times, illumination may differ. For these reasons, in many applications, it is advantageous to make widely separated view matching techniques robust to changes in illumination.

5.9.1 Illumination Invariance

A first strategy to deal with illumination changes is to find ways to represent images in a manner that is invariant to illumination variations. This is the idea behind ordinal measures developed by Bhat and Nayar [10], where correspondence is established using intensity ranks, rather than the intensity values. Matching then consists in measuring the closeness between pairs of ranking matrices, which represent the relative ordering of the intensity values in image windows. The proposed distance metric relies on the estimation of the maximum number of misplaced preceding elements over each element of the ranking matrix. Scherer et al. [94] have however demonstrated that this measure has a low discriminatory power, that is, produces numerous ambiguous matches. To mitigate this problem, they add another term to the Bhat-Nayar ordinal measure.

Another way to obtain illumination invariance is to exclude low frequency information from the images. According to Stockham [103], low frequencies should correspond to the illumination part of the signal, while the high frequency components correspond to the reflectance. Alternatively, only high frequency components could be used, such as gradients. Lowe [66] uses gradients of neighboring points for matching, as described in Subsection 5.5.1. The process becomes robust to illumination change when the gradient magnitudes are thresholded at 10% of their maximum possible value, thus eliminating lower frequency variations. The use of contour maps by Govindu and Shekhar [33], rather than image intensities, is another way to obtain invariance to illumination changes by using only high frequencies in the images.

Nevertheless, any attempt to separate illumination changes from the reflectance information necessarily results in a loss of information, and should thus only be used in applications where it is warranted.

5.9.2 Intensity Change Modelling

The other approach is to fit an explicit illumination model to the image data, such as those described by Lai and Fang [58], Viola and Wells [123], or Wang et al. [124]. These sophisticated models can be used in conjunction with energy minimization schemes, to align images taken with very large changes in illumination. However, in the context of widely separated views, illumination variations are often less important. Since accurate models of illumination variation increase the complexity of matching procedures, simple heuristic are often sought instead.

For simplicity, changes in pixel intensity due to changes in illumination can be modelled by a linear transformation (a photometric offset and scaling):

$$I' = aI + b \tag{5.22}$$

Mindru et al. [75] only take account of the offset, and thus normalize the pixel's intensities by substracting the average intensity in their neighborhoods. Schmid [96] also removes the offset, but by convolving with a filter designed to always give a zero DC component (see Equation (5.9)).

Otherwise, in order to make matching invariant to changes in illumination, many authors first normalize the images by re-expressing the pixel intensities, through a linear transformation using the extrema of their neighboring grey values [6, 77, 105]. Pixel intensities should then be invariant to linear changes in illumination, as long as feature point neighborhoods with corresponding shape are compared, a situation that is hard to achieve in the widely separated view context.

5.10 Robust Estimators of the System Geometry

Once a set of candidate correspondences has been established using the tools presented in the previous sections, a robust estimation scheme should be used to extract the epipolar or trifocal geometry relating the views. This extracted geometry is needed to filter out the mismatches that are always present. A way of increasing the tolerance of a matching process to more difficult matching situations, including widely separated views, is to make the estimator more robust to outliers.

5.10.1 LMedS and RANSAC

The most commonly used estimators are the Least Median Squares estimator (LMedS) and the Random Sample Consensus estimator (RANSAC). LMedS, used in [77, 127] is less appropriate for widely separated view matching, as it requires that mismatches constitute no more than 50% of the set of candidate feature matches that it uses to estimate a camera system's geometry.

RANSAC, already described in Subsection 2.7.1, is an approach that can succeed with candidate match sets containing more outliers, and has been commonly used in the widely separated view matching literature, such as in [7, 72, 73, 85, 92, 105, 112]. Cham and Cipola [15], introduced a Bayesian version of RANSAC, which takes account of the correlation scores between features in the estimation process. This improves the estimate in more difficult cases, such as with widely separated views. They show examples where mosaics are produced despite significant changes of views, through homography estimation with this Bayesian RANSAC.

5.10.2 IMPSAC

Torr and Davidson [106], have developed an estimator, called Importance Sampling Consensus (IMPSAC), which they claim makes the matching process invariant to changes in viewpoint. It uses a coarse to fine approach were the estimated epipolar

geometry at the coarser levels are used in the estimates at finer ones. The information to be passed to finer levels is represented as a set of particles which are obtained through importance sampling, while RANSAC is used to generate this sampling function.

The coarse to fine process restricts greatly the number of possible matches for a feature point, as they will now be constrained to a small portion of the image. In addition to the epipolar geometry, a global homography is also transmitted to finer levels which attempts to approximate the global image transformation. This homography can be used to warp image neighborhoods in order to make correlation more effective. The algorithm appears to effectively find several matches on some test image pairs.

5.10.3 Estimation Without Previous Correspondence

Georgis et al. [30], propose a very different approach to matching that is based entirely on a robust estimator, and uses no similarity measure between feature points. However, the initial knowledge of 4 coplanar point correspondences is assumed. The idea is that an epipolar line in the right image, for any point in the left one, passes through the true correspondence of this left image point, as well as its *v*irtual correspondence (its image through the known homography defined by the 4 coplanar points). Thus, for correct matches, the lines defined by joining the true correspondences to the virtual ones will be the epipolar lines, and will intersect in a single point (the epipole).

A cost function is designed to evaluate a set of potential correspondences by measuring how close the resulting epipolar lines are to intersecting in a single point. Matching has then been reduced to an optimization problem solved by randomly selecting 3 points in each image until the resulting epipolar lines intersect in a single point.

Since the epipolar geometry can be determined by seven point correspondences, and here, four are assumed to be known. The method is essentially a modified

RANSAC for finding the remaining three. However, an important distinction is that the points pairs are randomly formed, rather then being from a candidate match set built using a similarity measure. This makes the scheme quite powerful, as the authors use it to match points in images taken by cameras separated by 90° of rotation. The main drawback is that 4 coplanar matches must be initially available, and presumably manually selected.

5.11 Conclusion

This chapter presented a survey of the different approaches that have been proposed to solve the correspondence problem between widely separated views. The different approaches proposed in the literature were grouped in terms of their robustness to the different classes of transformations: changes in scale, similarities, affinities, homographies, and changes in illumination. Such a comprehensive review is not, to our knowledge, available elsewhere, and thus, constitutes an important contribution to the field.

Chapter 6

Uncalibrated Widely Separated View Corner Matching

6.1 Introduction

Uncalibrated widely separated view matching is the most difficult matching situation. Firstly, calibration is unavailable to guide matching as was the case in Chapters 3 and 4, and point displacements between the views might vary wildly. Thus, the search for correspondences must cover the whole image, leaving ample opportunity for ambiguity. In fact, the increased likelihood of occlusions and the varying area covered by the images mean that many feature points do not have a correspondence in the other image at all. Secondly, image regions can be severely deformed by the changes in viewpoint. Thus, similarity measures based on direct correlation of image regions fail.

The regions in different views, around points located on the same planar scene areas, are related by homographies (assuming there are no occlusions). Two of the eight DOF of a homography are already known, as the translation parameters are defined by the coordinates of the points being compared. However, a six DOF family of possible transformations between the two image regions remains.

Although no completely satisfactory solution currently exists for the problem,

widely separated view matching could be useful to many applications. It could be because no other views are available. In other cases, it could be due to attempts in minimizing the number of images that cover a given scene. Elsewhere, more widely separated views might be sought because they bring an increased accuracy to reconstruction and depth estimation. Indexing in image databases is another application of widely separated view matching that is not an extension of the narrowly separated case. The idea is that images can be found to be similar when they contain similar feature points. Thus, an application can search for similar images in an image library by looking for images with several consistent matching feature points. In order to be useful, such an application must be able to deal with significantly different views of objects.

The main challenge in widely separated view matching is making the similarity measure used to compare image feature points tolerant to projective distortion. This can be done in two ways, either by making the measure invariant to some class of projective transformations, or by estimating and eliminating the distortion before comparing feature points. Making a similarity measure invariant to projective distortion is a very delicate task, as there is a clear tradeoff between the level of invariance achieved, and the discriminating power of the measure. Estimating the local projective deformation between possibly corresponding feature point neighborhoods is nonetheless practically impossible, as six DOF need to be accounted for. This is why an affine approximation is often sought instead, requiring only four additional DOF.

6.1.1 Summary

This chapter will introduce a method of comparing corner point neighborhoods by first estimating an affine transformation between them using the shape of the corners, then warping one of the corner neighborhoods by this transformation, and finally correlating the warped image regions. But first, a corner detector will be introduced which can yield the required corner shape information. This corner detector's use, however, is not necessarily limited to the widely separated view context.

The work described in this chapter has largely been published. Some of the results were published in:

Robert Laganière and Etienne Vincent,
Wedge-based Corner Model for Widely Separated Views Matching,
in Proc. 10^{th} *International Conference on Image Processing*, vol. 1, pp. 277-280, Barcelona, Spain, September 2003.

Then, a longer journal version was accepted to appear in:

Etienne Vincent and Robert Laganière,
Detecting and Matching Feature Points,
in Journal of Visual Communication and Image Representation, to be published.

The main contribution of the work described in this chapter is the introduction of a new feature point detector, and the demonstration of the possibility of using the shape of feature points in estimating the local perspective distortion between different views of feature points. Section 6.2 describes a corner point model, and a procedure to find image points whose neighborhood fit that model. Section 6.3 proposes experiments to evaluate the feature point detector. Then, in Section 6.4, it is seen how the corner information provided by this detector can be used towards uncalibrated widely separated view matching. Finally, Section 6.5 concludes with a more detailed list of the contributions of this chapter.

6.1.2 Previous Work

The previous Chapter was a detailed survey of the literature on widely separated view matching. The method used for estimating local affine transformations between image point neighborhoods that will be presented here is based on corner points and the lines which define them. Similarly, the shape of corner points was used in matching

by Darrell [18], but towards an application in tracking. There, the goal was not the estimation of a perspective deformation, but the extraction of the foreground object in order to consider it only when establishing correspondence.

The approach proposed here is somewhat more similar to the approach of Tuytelaars et al. [112], where affine-invariant regions are extracted starting from the intersection of image edges. Finding two corresponding regions extracted in an affine-invariant way amounts to discovering an affine approximation of the homography relating the regions. However, despite the fact that the two approaches make use of the lines which define junction points, they are very different in their use of these lines, and in that Tuytelaars et al. use moment invariants to compare the extracted image regions.

In [128], Zoghlami et al. also use the shape of corner points to recover homographies. However, they are interested in global homographies for building mosaics. They use the corner model introduced by Deriche and Blaszka in [20] to describe corner points. Then, they compute homographies from the four lines that constitute the edges of two such corner points.

A great part of this chapter is also devoted to feature point detection. The most commonly used feature point detectors rely on image intensity derivatives to identify high curvature points. These will be reliable, as long as they include some smoothing operation to reduce sensitivity to noise. The most popular detector in this category was proposed by Harris and Stephens [39], and is described in Subsection 2.3.1.

In other approaches, detection is based on the direct comparison of intensity values inside small predefined windows. The SUSAN operator [102], proposed by Smith and Brady, fits into this category, and was described in Subsection 2.3.2.

Finally, some detectors define feature point models, and search for image regions which match them. These include the detectors of Deriche and Blaszka [20], Guiducci, [34], and Rohr [88]. The selected parameterized models generally represent ideal corners, and the detection consists in determining the value of the model's parameters which best fit the underlying intensity pattern. Depending on the approach taken to

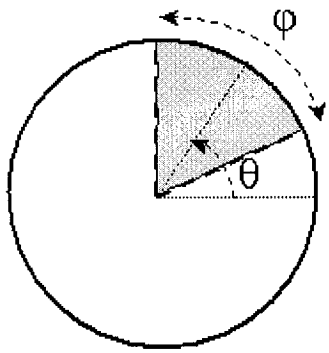


Figure 6.1: The corner model: a wedge W_θ^φ at angular position θ and having an angular width of φ.

optimize these parameters, the complexity of this kind of detector can be quite high.

6.2 A Wedge-Based Corner Detector

6.2.1 The Corner Model

The corner detector introduced in this chapter uses an intensity-based approach that makes use of a simple corner model to detect corner shapes from intensity patterns. The goal was to devise a stable and efficient corner detector that would later yield the corner shape information needed for the approach to matching that will be presented in Section 6.4. The corner model consists of a wedge (the corner), having its origin on the center of a circular neighborhood (the background). This idealized corner is described by two parameters: an angular position, θ, and an angular width, φ (see Figure 6.1). Furthermore, the wedge's width must be within a reasonable angular range, i.e. $\varphi_{min} < \varphi < \varphi_{max}$.

This model is similar to the one used by Rohr [88], who however also explicitly includes the strength of a Gaussian blurring on the corner, and the image intensity of the wedge and of the background, which are specifically required to be uniform. Rohr's

corner model was in turn inspired by Berzins [9], who worked on the problem that Laplacian edge detectors encounter near corners. Guiducci [34] uses a model which is similar to Rohr's, but only looks at the difference in intensity between the wedge and its background. The model is then fitted analytically from image derivatives rather than directly from the image intensities. Deriche and Blaszka [20] use the same model as Rohr, but consider a more efficient implementation to the fitting. On the other hand, Marr [71] proposes a much simpler model, only allowing for wedges with an angular width of 90°.

To identify corners in an image, each pixel location should be examined by comparing its surrounding circular neighborhood with the ideal corner model. To do so, the following process is proposed:

1. The circular window's intensity mean and variance are computed. The variance must be over a given threshold σ^2_{min}. This is to limit the sensitivity of the corner finder by discarding low contrast areas which do not correspond to corners but might otherwise generate false positives.

2. The circular area around a potential corner is segmented into a background and a foreground, following a simple classification scheme described in Subsection 6.2.2.

3. The corner model is fitted to the extracted foreground, by finding the values for φ and θ that best approximate the area, as explained in Subsection 6.2.3. The strength of the corner is then determined by comparing the segmented area to the parameterized corner model.

Applying this procedure to all pixels results in a corner map where each feature point is associated with a corner strength value, and the parameters of the fitted model at that location. The final set of corners is obtained by imposing a threshold on the corner strength values, which must be preceded by non-maxima suppression to eliminate clusters. The non-maxima suppression step simply involves discarding all

corner locations immediately adjacent to other locations with higher corner strength scores.

The efficiency of a detector is essentially determined by the ease with which corner information is extracted from raw image intensity values. This is accomplished here with a simple but effective foreground/background classification scheme. The result is a segmented area from which the ideal model can be incrementally reconstructed. This is explained in the next two subsections.

6.2.2 Background/Foreground Segmentation

To obtain a simple foreground/background segmentation of circular areas around potential corners, the surrounding pixels are classified as having an intensity above, or below the mean intensity in the area. The group covering the smallest area becomes the foreground, and the other group forms the background. This segmentation is similar to the truncated blocks method which was used in Subsections 2.5.2 and 2.5.3 in matching constraints which depended on the shape of feature point neighborhoods.

In practice, the smallest region of the segmented area is not necessarily its foreground (in the case of a concave foreground element). In fact, both areas might be part of the same scene element (in the case where the corner is part of a color pattern on a single object). The terms foreground/background are simply used because they represent the common situation of corners defined by a convex foreground object. Different types of corners might be more easily matched, but aside from that, the fact that the foreground and background areas of corners might be misclassified has no effect on the following matching step described in Section 6.4.

Instead of using a hard threshold, a sigmoidal function is used to determine the degree to which each pixel belongs to one of the two classes. The use of this sigmoidal function gives more refined segmentations. It has the form:

$$sig(I(x,y),\overline{I}) = \frac{1}{1+e^{s(I(x,y)-\overline{I})}} \tag{6.1}$$

where $I(x,y)$ is the pixel location being considered, $\overline{I}$ is the average intensity in the

considered circular region, and s is a constant whose sign is adjusted such that pixels belonging to the foreground obtain a segmented value greater than 0.5. According to our wedge model, if the center pixel does not belong to the foreground region, then there is no corner at this location, so the algorithm proceeds with the next pixel. Note that while the use of a sigmoid function greatly improves the stability of the operator, the specific value of s that is chosen does not bear much influence on the final result. In practice, values in the range $[0.1, 1.0]$ gave good results.

In a similar fashion, the SUSAN corner detector (see Subsection 2.3.2) relies on a simple segmentation to obtain univalue segments. Corners are then identified from the properties of the resulting nuclei, but this is done without explicitly comparing them to a preestablished corner model.

6.2.3 Corner Model Extraction

The model which best fits the segmented area is found from the following criterion:

$$c_s(x, y, W_\theta^\varphi) = \iint_{\mathcal{C}_{xy}} |W_\theta^\varphi(i, j) - sig(I(x+i, y+j), \overline{I})| \, \mathrm{d}i \, \mathrm{d}j \tag{6.2}$$

where $\mathcal{C}_{xy}$ is a small circular window centered on the point (x, y), W_θ^φ is a binary map of the wedge model being fitted as in Figure 6.1, $sig()$ is the sigmoid function defined in Equation (6.1), and $\overline{I}$ is the average pixel intensity in $\mathcal{C}_{xy}$. In the context of feature point detection, it would be too costly to find the optimal θ and φ for this pseudo Hamming distance using functional minimization. Instead, we propose to use the following strategy to approximate the optimal values:

1. The circular area around a potential corner is subdivided into small wedges, $W_{n\Delta\theta}^{\varphi_{min}}$ having a width of φ_{min} and rotated around the circular area by increments of $\Delta\theta$. Note that for a better accuracy in the delimitation of the wedge model, $\Delta\theta$ will generally be smaller that φ_{min}, implying that adjacent elementary wedges will overlap.

2. The foreground coverage of these elementary wedges is then computed:

$$c_{cover}(x, y, W_{n\Delta\theta}^{\varphi_{min}}) = \iint_{W_{n\Delta\theta}^{\varphi_{min}}} sig(I(x+i, y+j), \overline{I})\, \mathrm{d}i\, \mathrm{d}j \tag{6.3}$$

   An elementary wedge will be considered as a part of the potential corner's foreground if its corner coverage is greater than some predetermined threshold value, c_{min}. This use of wedges is similar to the rotated wedge filtering presented by Yu et al. in [125], where it is used to compute intensity mean values, from which one-dimensional angular derivatives are computed.

3. The wedge with the highest coverage is selected.

4. All adjacent foreground wedges are retained (those for which $c_{cover}(x, y, W_{n\Delta\theta}^{\varphi_{min}}) > c_{min}$). The corner model is then the one formed by the union of all the selected elementary wedges.

The union of all the wedges selected in the above procedure determines W_θ^φ, which is used in Equation (6.2) to compute the strength of the detected corner. Thus, φ is the angle spanned by the union of all selected $W_{n\Delta\theta}^{\varphi_{min}}$, and θ is their bisector. However, before going further, it is verified that φ is compatible with the corner model described in Subsection 6.2.3, i.e. $\varphi_{min} < \varphi < \varphi_{max}$.

6.2.4 Controllability of the Wedge-Based Detector

The wedge-based detector is controlled by several parameters, but each has a specific role. Usually, all these parameters can be fixed except c_{min}, which is used as the main control parameter to determine how many points are selected. The parameters are:

1. The minimum acceptable variance σ_{min}^2, which controls the sensitivity of the operator. This parameter is normally fixed, but could be modified to deal with images containing more noise or artifacts. A value of 150 was used to produce the presented results.

2. The step angle $\Delta\theta$, that discretizes the space of admissible angles for a wedge corner. The smaller this angle, the more precise the identification of the edges of the corner. A value of 10° was used to produce the presented results.

3. The minimal and maximal angle width of corners φ_{min} and φ_{max}, which define the range of acceptable wedge angular widths that can be attributed to a corner. These control parameters can be selected once and for all. The presented results used values of 30° and 120°.

4. The radius of the wedge model is the last control parameter. it should be relatively small to reduce the computation time. The presented results were obtained with a radius of 7.

5. The corner strength threshold c_{min}. This is the main control parameter, and the only one that needs to be tuned according to the problem at hand. Its value is chosen as a compromise between selecting more corners, and selecting better corners. Values between 0.75 and 0.9 are reasonable.

6.3 Experimental Comparisons

In order to test the proposed corner detector, experiments that compare the wedge-based corner detector to the SUSAN (see Subsection 2.3.2) and Harris (see Subsection 2.3.1) detectors were conducted. Results are presented in the following two subsections.

6.3.1 Corner Localization

The first set of experiments measures the robustness and the accuracy of each detector. To this end, a synthetic image was used, where the position of each corner (here the vertices of synthetic shapes) is known.

Figures 6.2 and 6.3 show the original test image, and the feature points found by the three different detectors. Note the variability in contrast conditions at different

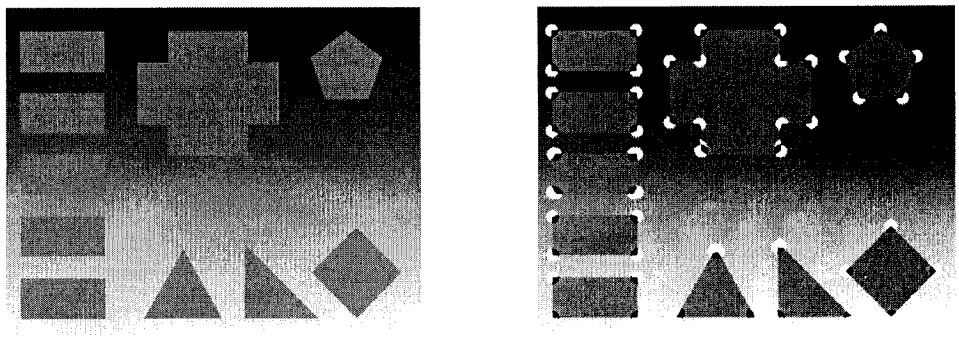
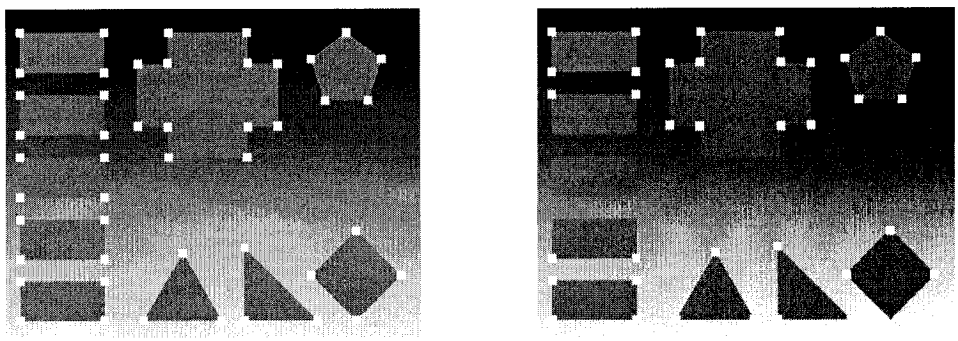


Figure 6.2: A synthetic image, and the wedge models of corners detected on it. The black foreground areas of the corners, as described in Subsection 6.2.2, do not necessarily correspond to the physical foregrounds.



(a) (b)

Figure 6.3: Corners detected on a synthetic image. (a): using SUSAN. (b) using Harris.

corner locations, caused by the smooth gray level transition of the background. This allows a comparison of the respective sensitivity of the detectors, each of them having been tuned to obtain the best possible results. In Figure 6.3 (b), note that it was impossible to find a threshold for the Harris detector that would allow the detection of the corners located in the low contrast area without significantly increasing the number of false positives. In this area, there is a difference of 11 between the intensity values of the background and the shapes.

The behavior of the same corner detectors in the presence of noise was also tested. Figures 6.4 and 6.5 show the feature points detected after a Gaussian noise of standard

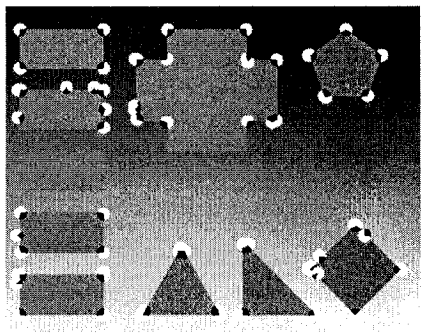
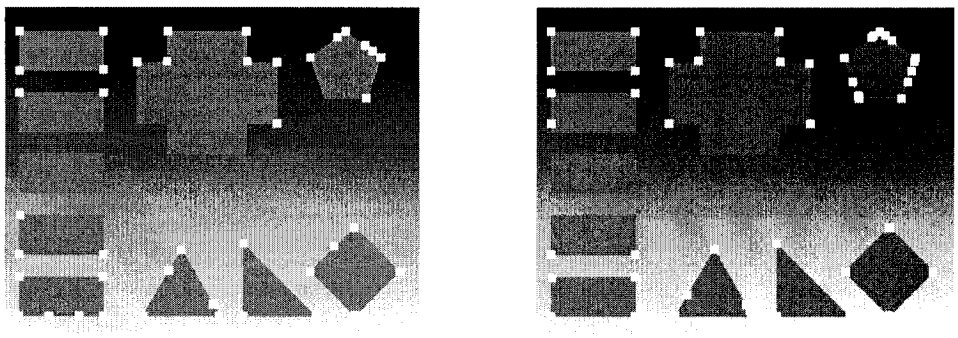


Figure 6.4: Wedge model of corners detected on a synthetic image with added Gaussian noise of standard deviation 50.



(a) (b)

Figure 6.5: Corners detected on a synthetic image with added Gaussian noise of standard deviation 50. (a): using SUSAN. (b) using Harris.

deviation 50 was added to the original synthetic image. The detection thresholds for these experiments were again chosen to visually give the best results. Then, for each vertex in the synthetic image of Figure 6.2, the closest detected corner was determined, and the distances between the vertices and their closest detected corner were measured. These measurements were taken for all three detectors, and for noisy versions of the test image. Results are presented in Figures 6.6 and 6.7, where the graphs show, for the different noise levels, the number of vertices having a detected corner within 1, 2, 3 and 4 pixels.

In light of the first experiments with the noiseless image, SUSAN appears to be a

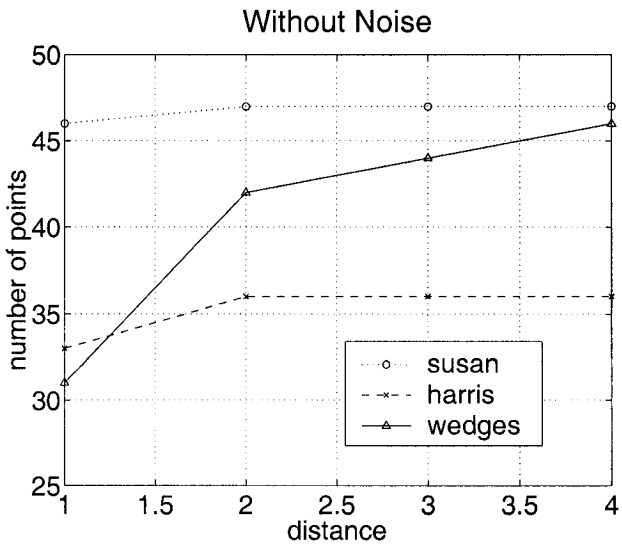


Figure 6.6: Evaluating the accuracy of different corner detectors by counting the number of vertices of the test image of Figure 6.2 with a detected corner at a distance smaller than 1, 2 , 3 and 4 pixels.

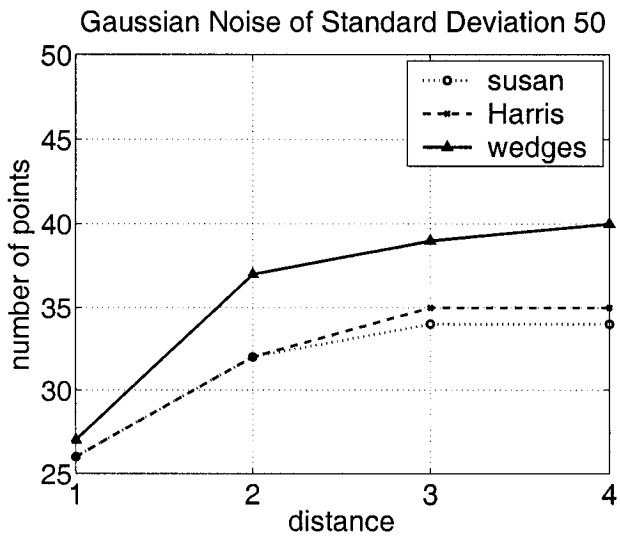


Figure 6.7: Result of the same experiment as in Figure 6.6, but done on an image with added Gaussian noise of standard deviation 50.

more accurate detector. However, as the level of noise is increased, the wedge-based detector leads to superior results. The poor performance of the Harris detector, in the presence of noise, can be attributed to its reliance on image intensity derivatives which are unstable with respect to noise. As for the segmentation used by SUSAN, it selects, as foreground, intensities in a small range around the central pixel's intensity. This represents a narrower range of possible intensities than the segmentation described in Subsection 6.2.2, making the segmentation into USANs more sensitive to noise. The fact that the segmentation described in Subsection 6.2.2 is not binary, but rather based on the sigmoid function of Equation (6.1) also reduces the effect of misclassifications within the segmentation. Finally, USANs are less stable because they depend on the intensity of the model's central pixel which is unstable in noisy circumstances.

However, it must be noted that with the noisy image, the proposed detector produced more false positives. For example, the wedge-based detector found points within 3 pixels, for 39 out of the 47 true corners in the image, as opposed to 35 and 34 for Harris and SUSAN respectively. But also found 22 false positives, compared to 13 and 8 for Harris and SUSAN respectively. It should nevertheless be noted that each detector produced about the same number of real false positives, while most of the wedge-based detector's and some of Harris' false positives were found in clusters around the true corners. These clusters survive the step of non-maxima suppression since they are formed of points which are not immediately adjacent, but separated by one or two pixels with lower corner strength. Fortunately, when the goal is matching, the number of false positives is far less important than the robustness of a detector.

6.3.2 Repeatability of the Corner Detection

In this subsection, a second set of experiments concerning the stability of the different detectors is presented. Repeatability was previously introduced in Subsection 2.3.3, and consists in the proportion of features that are consistently detected in different views of a scene.

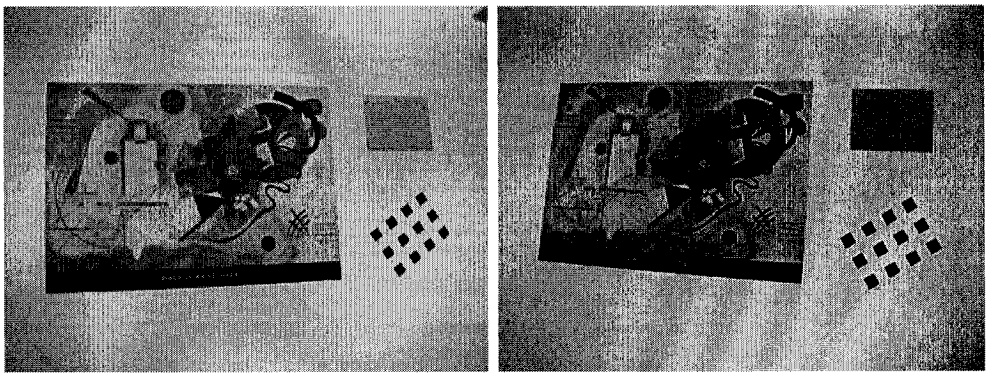


Figure 6.8: Images related by a homography: Two views of a plane.

Figure 6.9: Images related by a homography: Two views separated by pure rotation.

To evaluate repeatability, the location of a corresponding point in a second view must be known for every feature point in a first view of the same scene. It is easy to determine this correspondence when the views are related by a homographic transformation, using Equation (1.6). It was seen in Section 2.8 that a homography relates two views of a planar surface. This relation also holds between images where the difference in viewpoint corresponds to a pure rotation around the camera's focal point. Two image pairs, corresponding to these two situations where the views are related by homographies, are shown in Figures 6.8 and 6.9. Having estimated the values of $\mathbf{H}$ for these image pairs, it is easy to determine the expected position of a corresponding point in the other image using Equation (1.6), and to see if a feature point was also

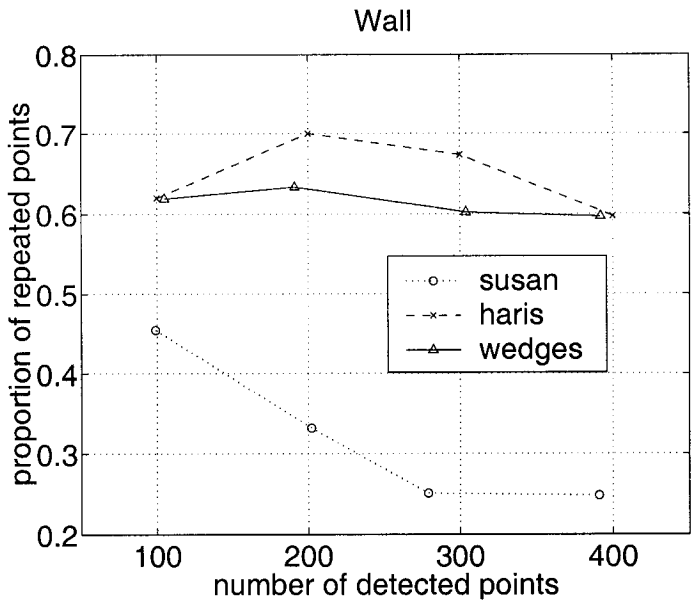


Figure 6.10: Evaluating the stability of corner detectors using their repeatability, on the images of Figure 6.8.

detected at that location.

The computed repeatabilities for each considered corner detector are given in Figures 6.10 and 6.11. These show the proportion of detected feature points in the left views for which the corresponding points in the right image have also been detected. These repeatability rates were computed for different corner acceptance thresholds. In this case, SUSAN clearly demonstrates the worst performance, while the other two detectors present a similar behavior.

The novel feature point detector introduced in this thesis proceeds in two steps: It first performs a simple segmentation based on the intensity values found in the vicinity of each considered point. Then, it tries to fit a simple wedge corner model to the resulting segmented area. The combination of these two approaches contribute to the effectiveness of the detector. Indeed, intensity-based corner finders, like the SUSAN operator, tend to show good accuracy, a property also inherited by the operator proposed here. On the other hand, the stability of such approaches is usually poor

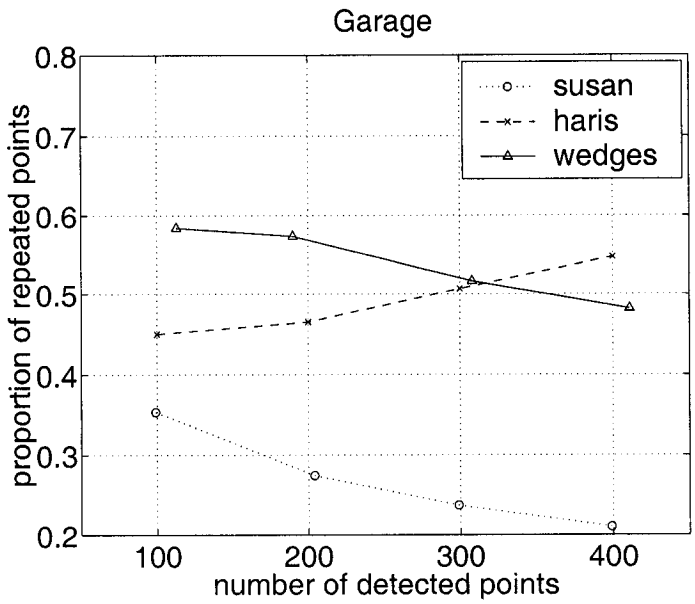


Figure 6.11: Evaluating the stability of corner detectors using their repeatability, on the images of Figure 6.9.

because intensity patterns can undergo important modifications when geometrical transformations are applied. This is not the case of model-based approaches such as the one presented here, or derivative-based approaches such as the Harris detector, which normally exhibit good stability. However, derivative-based approaches are usually more sensitive to noise than model-based approaches. These observations were confirmed by the experiments of this section.

6.4 Matching Feature Points

Here, the goal of feature point detection is in selecting points towards sparse matching. When the change in viewpoint between images is small, matching can be accomplished effectively using VNC and some simple constraints, as shown in Chapter 2. However, in the case of widely separated views, the situation is more difficult, as there might be significant deformation between the images to be matched. Then, normalized

correlation between windows around feature points may not provide a meaningful measure of similarity. It will now be shown how the corner shape information conveyed by the wedge-based corner detector could contribute to improve matching results. A corner detector that provides rich descriptive information about its detected corners can be advantageous.

6.4.1 Local Affine Transformations

The local 2D projective transformation between two image patches around corners will now be recovered. This transformation will then be used to warp the image patches prior to correlation. The local neighborhoods of points, as in most widely separated view matching works, are assumed to be locally planar, so that perspective distortion can be approximated as a homography.

To further simplify the problem, it is often assumed that shear effects between views can be neglected, and that simple similarity transformations (rotation+scale) can be used, as in the methods surveyed in Section 5.5. However, a better approximation of the homography due to perspective effects consists of an affine transformation, used in works reviewed in Section 5.6, and described by Equation (1.8). These transformations approximate viewpoint distortion with a translation, a rotation, and a non-isotropic scaling (skew).

When matching is done between feature points, the translation between them is determined by their relative position in the respective images. As explained previously, 4 DOF remain, in order to completely describe an affinity. Some extra information can be provided by the wedge corner model. Assuming that the difference in scale between the two images may be neglected, the affinity can be estimated using the two following simplifications:

1. The rotation angle is given by the difference in angular position between the two detected wedges, $\theta - \theta'$.

2. The non-isotropic component of the affinity is applied in the bisector direction

of the rotated wedge, with a ratio given by the angular widths of the two wedges, $\varphi : \varphi'$.

Under these assumptions, given a point (x, y), with a corner model represented by θ and φ, which corresponds to a point $(x+t_1, y+t_2)$, with a corner model represented by θ' and φ', then $(x+\cos(\theta+\varphi/2), y+\sin(\theta+\varphi/2))$ should correspond to $(x+t_1+\cos(\theta'+\varphi'/2), y+t_2+\sin(\theta'+\varphi'/2))$, and $(x+\cos(\theta-\varphi/2), y+\sin(\theta-\varphi/2))$ should correspond to $(x+t_1+\cos(\theta'-\varphi'/2), y+t_2+\sin(\theta'-\varphi'/2))$. By solving the resulting linear system of equations, the coefficients from Equation (1.8) are found to be:

$$\begin{aligned}
a_{11} &= \frac{\cos(\theta'+\varphi'/2)\sin(\theta-\varphi/2) - \cos(\theta'-\varphi'/2)\sin(\theta+\varphi/2)}{\cos(\theta+\varphi/2)\sin(\theta-\varphi/2) - \cos(\theta-\varphi/2)\sin(\theta+\varphi/2)} \\
a_{12} &= \frac{\cos(\theta+\varphi/2)\cos(\theta'-\varphi'/2) - \cos(\theta-\varphi/2)\cos(\theta'+\varphi'/2)}{\cos(\theta+\varphi/2)\sin(\theta-\varphi/2) - \cos(\theta-\varphi/2)\sin(\theta+\varphi/2)} \\
a_{21} &= \frac{\sin(\theta'+\varphi'/2)\sin(\theta-\varphi/2) - \sin(\theta'-\varphi'/2)\sin(\theta+\varphi/2)}{\cos(\theta+\varphi/2)\sin(\theta-\varphi/2) - \cos(\theta-\varphi/2)\sin(\theta+\varphi/2)} \\
a_{22} &= \frac{\cos(\theta+\varphi/2)\sin(\theta'-\varphi'/2) - \cos(\theta-\varphi/2)\sin(\theta'+\varphi'/2)}{\cos(\theta+\varphi/2)\sin(\theta-\varphi/2) - \cos(\theta-\varphi/2)\sin(\theta+\varphi/2)}
\end{aligned} \tag{6.4}$$

6.4.2 Matching Warped Neighborhoods

The proposed matching scheme therefore proceeds as follows:

1. Wedge corners are detected in each view.
2. For all possible pairs of wedge corners, the affine transformation between them is estimated using Equation (6.4).
3. VNC is applied to the intensity values found in the vicinity of the corner in the first view and the corresponding intensity values, in the second warped image.
4. All matches with a correlation score higher than a given threshold are retained.

This process is illustrated in Figure 6.12 where two views of a feature point are shown. In this work, VNC from Equation (2.4) was used. It is seen in Figure 6.12 that the use of the estimated affine transformation improves greatly this correlation score.

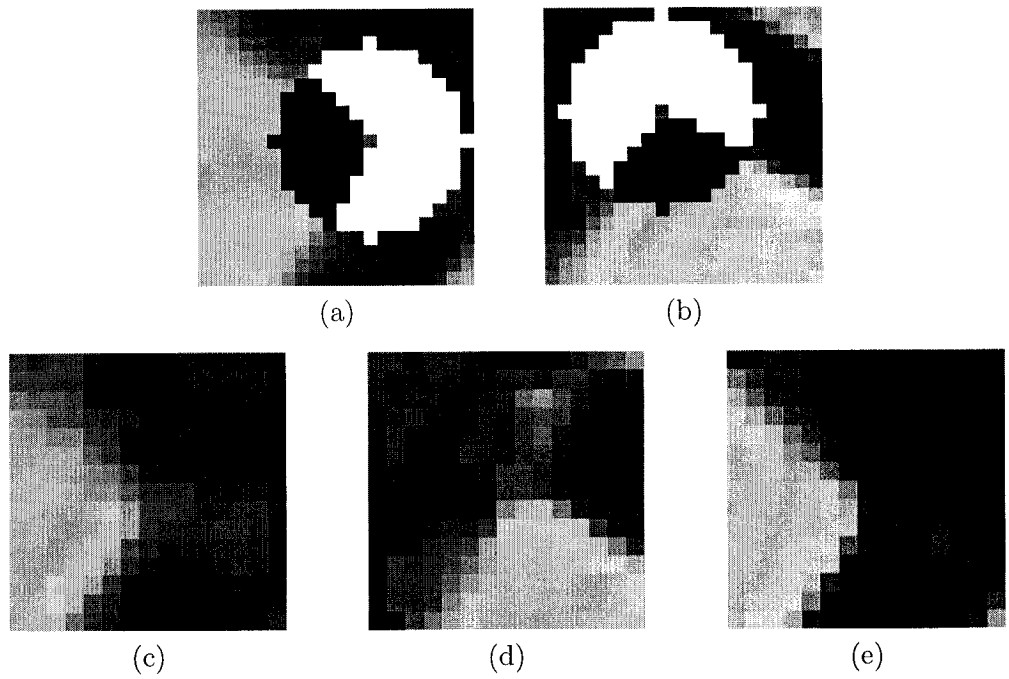
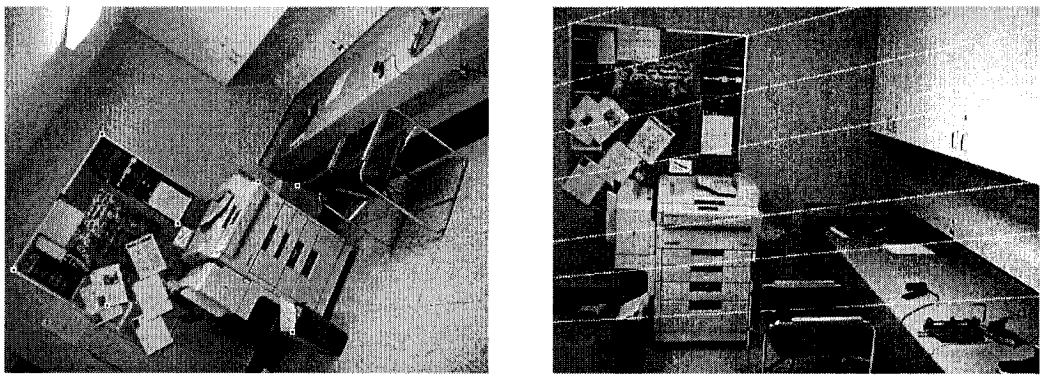


Figure 6.12: Matching of two corners with an affine warp. (a) and (b): The detected wedge models on the two corners. (c): The correlation window around corner (a). (d): The corresponding window around (b) before an affine transformation is applied (the correlation score is -0.31). (e): The window around (b) after applying the affine transformation defined by Equation (6.4) (correlation score is now 0.96).

Obviously, this simple approach will still produce a large number of false matches. However, in the case of widely separated views, it should produce many more good matches than would be obtained through direct correlation of the image patches. Mismatches can still be filtered out using some robust estimator of the camera system's geometry. In the following examples, a random sampling consensus (RANSAC) scheme was used to estimate a fundamental matrix or homography, towards filtering the match sets (see Subsections 2.7.1 and 2.8.2). Again, the software of Roth and Whitehead [90] was used to perform the RANSAC estimations.

Figure 6.13 shows a pair of images on which the matching procedure was applied. The resulting candidate match set contained 300 point pairs, and was used to estimate



(a) (b)

Figure 6.13: The Epipolar Geometry estimated from point correspondences obtained using the wedge-based corner model to warp corner neighborhoods. (a) One view with some selected points. (b) The corresponding epipolar lines on the second view.

the epipolar geometry. A fundamental matrix was found which agreed with 35 of the candidate matches. However, when 300 different pairs were used, that were found through direct correlation without prior warping, too few were found to be correct matches. No accurate fundamental matrix could be obtained, no matter how many RANSAC iterations were attempted.

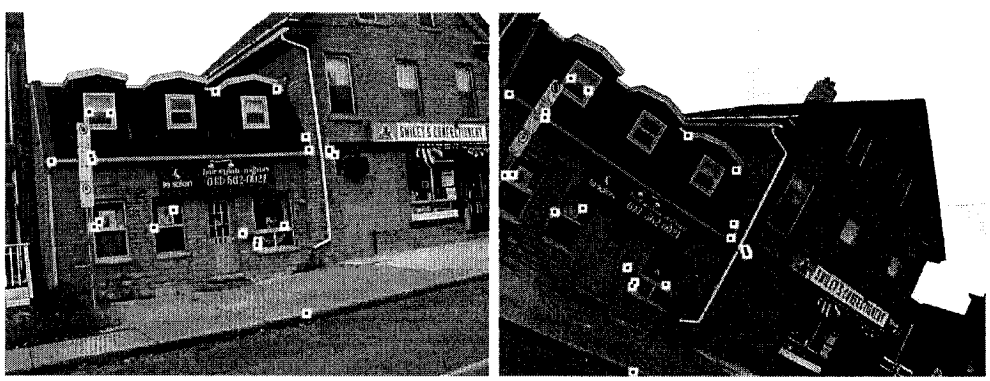


Figure 6.14: Two views obtained through pure rotation of a camera.

The same approach was also applied to the image pair shown in Figure 6.14. In this case, the two images are related by a pure rotation. This time, after applying the

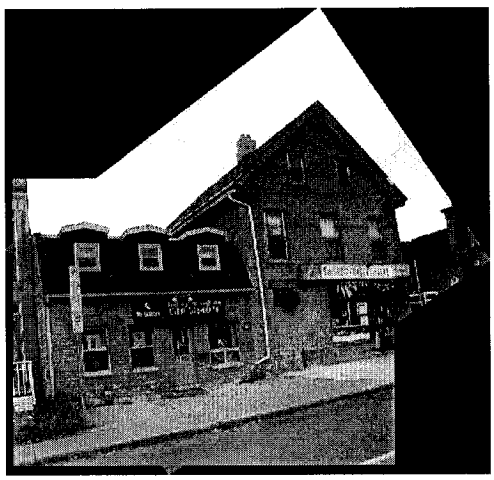


Figure 6.15: Homography estimation: A mosaic constructed using the homography detected between the views in Figure 6.14, using the wedge-based corner model to warp corner neighborhoods.

proposed matching scheme, a homographic transformation was estimated from 300 point pairs. This homography agreed with 21 candidate matches that are drawn in Figure 6.14. A mosaic is shown in Figure 6.15 which was constructed from the computed homography. Again, using the best 300 matches obtained by direct correlation did not produce valid results.

6.5 Conclusion

In this chapter, a new corner detector was described. Then, its use in recovering an approximation of local homographies in image pairs was introduced. The main contributions of this work are:

- The development of a new wedge-based corner detector. This includes the description of a corner model, and a method to fit this model.
- The introduction of a segmentation method for small image regions towards fitting the corner model.

- The experimental demonstration that the wedge-based detector is accurate in the presence of noise.
- The experimental demonstration of the high repeatability of the wedge-based detector.
- The introduction of the idea of using the shape of feature points towards the estimation of local perspective distortion between corresponding feature point neighborhoods.
- The description of a method for estimating the local affine transformation between point neighborhoods.
- The demonstration of the usefulness of the wedge-based feature point detector, together with local affine transformation estimation method in generating mosaics, and fundamental matrices estimates.

Part IV

Matching Calibrated Widely Separated Views

Chapter 7

Calibrated Widely Separated View Junction Point Matching

7.1 Introduction

Matching feature points between calibrated images can be seen as an extension of the problem addressed in Part II, but where the process must be made robust to greater image deformation. In Chapter 6, a solution to the matching problem was suggested that estimated local perspective distortion from the shape of corner points. Here, it will be seen how this process can be improved with the knowledge of the epipolar geometry.

If the scene neighborhood of a feature point is planar, local deformation between the corresponding image regions will be a homography. This homography must agree with the epipolar geometry via Equation (2.15). Therefore, if an attempt is made to estimate the local perspective distortion, as was done in Section 6.4, the knowledge of the image pair's epipolar geometry will yield further constraints on the local homography.

7.1.1 Summary

This chapter proposes a method for establishing point correspondences between widely separated views when the underlying epipolar geometry is known. This might be needed to interpret a dynamic scene from images taken by fixed calibrated cameras, or towards epipolar geometry recovery from approximate camera positions and orientations, as will be proposed.

Essentially, the correspondence problem will be solved by estimating the local perspective distortion between junction point neighborhoods using the shape of these junctions, and the camera system's epipolar geometry. Unlike most proposed solutions to making matching robust to perspective deformation, the proposed formulation is fully invariant to projective deformation, rather than to an affine approximation of this deformation.

Most of the work presented in this chapter was submitted for publication in:

> Etienne Vincent and Robert Laganière,
> **Junction Matching and Fundamental Matrix Recovery in Widely Separated Views**,
> *submitted to British Machine Vision Conference*, 2004.

A longer journal version is also planned.

The main contribution of the work described in this chapter is the development of a method for estimating local homographies that approximate perspective distortion, and its application to calibrated widely separated view matching. Section 7.2 presents the constraints that can be used in estimating local homographies. Section 7.3 describes the proposed approach to computing them between junction points. Section 7.4 describes the proposed matching process. Section 7.5 introduces an application to fundamental matrix recovery, in the case where a crude estimate of camera positions and orientations is available. Then, Section 7.6 examines the extent to which the estimates would be sensitive to errors in the position and orientation of the cameras. Finally, Section 7.7 concludes with a more detailed list of the contributions of this

chapter.

7.1.2 Previous Work

It can be noticed, in the literature survey presented in Chapter 5, that all the discussed matching schemes were set in the uncalibrated context. Indeed, little has been proposed for the calibrated case. This is in part why it will be attempted to justify working on this problem by introducing an application to fundamental matrix recovery in Section 7.5.

Some reference should nevertheless be made to the literature on rectification towards matching, which is the process most similar to the method introduced here. Rectification consists in finding some *global* transformation that can partially eliminate the distortion between two images. Such a transformation would be computed from calibration and pose information. This is unlike the *local* homographies that are sought in this chapter.

The idea of rectification is to find a transformation which makes corresponding epipolar lines collinear and horizontal. This has the advantage of making the subsequent search for correspondence easier, as corresponding feature points in a rectified image pair will have the same vertical coordinate. However, their image neighborhoods will not necessarily be similar. The methods that have been proposed are based on different criteria, in attempts to minimize the distortion between rectified images. These include the works of Gluckman and Nayar [31], Hartley [41], Loop and Zhang [64], Pollefeys et al. [84], Robert et al. [87], and Roy et al. [91] which use different criteria in attempts to minimize the distortion in rectified images. However, it is always possible that some significant distortion remains, as image pairs can not generally be related by a single global homography.

The work of Shen and Palmer [101] does propose a way of matching junction points in a calibrated setting. They use a more complex model of junctions, which includes the endpoints of the line segments meeting at the junctions. Two such line segments in different views are compatible if their endpoints agree with the epipolar geometry.

Then, junction points are compatible if they agree with the epipolar geometry, and the line segments which define them agree. Finally, the topology of junction point groupings is used to determine matches. That is, groups of junction points related by sharing common line segments, must define similar graphs in the two views. Some examples are shown, where the method finds a limited number of matching junction points, on carefully chosen image pairs.

7.2 Homography Estimation

As in Subsection 6.4.1, it will be assumed that the neighborhoods of most image points are approximately planar. Thus, two views of these neighborhoods would be related by a homography. It would then be possible, if such homographies were recovered, to warp neighborhoods of feature points before comparing them, and thus solve the widely separated viewpoints problem. In Subsection 6.4.1, only an affine approximations of the local homographies were found, but here, having access to the image pair's epipolar geometry, the full homography will be recovered.

Estimating a local homography is not an easy task, as it involves solving for eight DOF (see Section 1.3). However, if the epipolar geometry of an image pair is known, much can be said about the possible planar homographies that are compatible with it. If a homography $\mathbf{H}$ is compatible with a fundamental matrix $\mathbf{F}$, the matrix $\mathbf{H}^\top\mathbf{F}$ will be skew-symmetric, that is, $\mathbf{H}$ and $\mathbf{F}$ are related by Equation (2.15). As explained by Luong and Viéville [69], this results in six homogeneous linear constraints on $\mathbf{H}$, five of which are linearly independent, thus leaving only three more DOF of $\mathbf{H}$ to be determined.

A homography describes the relationship between views of corresponding image points lying on a planar surface, but it also describes the relationship between corresponding lines. Thus, if $\mathbf{k}$ and $\mathbf{k}'$ are projections of a same line in space, lying on a plane whose views are related by a homography $\mathbf{H}$, they will be related by

$$\mathbf{k} = \mathbf{H}^\top\mathbf{k}' \tag{7.1}$$

where lines, as usual, are described by the vector whose dot product with the points lying on them is null.

If two image lines $\mathbf{k}$ and $\mathbf{k}'$, are known to correspond, Equation (7.1) adds three constraints (two of which are independent) on $\mathbf{H}$. Thus, the fundamental matrix, and two line correspondences lying on a common planar area are slightly more than enough to determine the homography between the views of that planar area, as they provide $5 + 2 + 2$ constraints on the 8 DOF of the homography.

7.3 Using Junction Points

Junction points can be used as the feature points to be matched between two views. Here, a junction point is defined as the intersection of two non-collinear image line segments. Once they are known, these lines can also be used in the matching process. The idea is that when two corresponding junction points are compared, the lines defining the junctions should also correspond. These line correspondences can then be used to provide constraints on the homography relating the junction point neighborhoods.

In Subsection 6.4.1, a simple way to exploit these constraints was presented, for the case where no other information is available. Homographies were approximated as affine transformations which align the lines defining junctions, while assuming that lengths are preserved along these lines, thus yielding Equation (6.4).

However, when the epipolar geometry of the image pair is available, Equation (2.15) can also be used to constrain the homography. In this case, the system is over-constrained, and thus a minimization is required. Good results were obtained when it was assumed that the correspondence between the lines defining junctions were exact. Therefore, the homography was approximated as the one agreeing with the line correspondences, but minimizing the constraints defined by the epipolar geometry. That is, the homography was defined as the $\mathbf{H}$ which minimizes:

$$\mathbf{H}^\top\mathbf{F} + \mathbf{F}^\top\mathbf{H} \text{ subject to } \mathbf{k}_i = \mathbf{H}^\top\mathbf{k}'_i \tag{7.2}$$

where $i \in \{1, 2\}$, $\mathbf{k}_i$ and $\mathbf{k}'_i$ being the lines defining the junctions. A least-squares solution to this constrained system of linear equations can be found directly using singular value decomposition.

The constraints $\mathbf{k}_i = \mathbf{H}^\top \mathbf{k}'_i$ result in a system of linear equations on the entries of $\mathbf{H}$, of the form $\mathbf{Ch} = \mathbf{0}$. Let $\mathbf{C} = \mathbf{UDV}^\top$ be the singular value decomposition (SVD) of $\mathbf{C}$, and $\mathbf{C}^\perp$ be the result of deleting all columns of $\mathbf{V}$ corresponding to non-zero values in the diagonal matrix $\mathbf{D}$. There should be r such entries, where r is the rank of $\mathbf{C}$ ($r = 5$ in this case). Then, any vector $\mathbf{h}$ satisfying $\mathbf{Ch} = \mathbf{0}$ can be written as $\mathbf{h} = \mathbf{C}^\perp \mathbf{h}'$ for some $\mathbf{h}'$.

The problem then consists in solving with respect to the constraints coming from Equation (2.15), but within the orthogonal complement of $\mathbf{C}$, the set of all $\mathbf{h}$, such that $\mathbf{Ch} = \mathbf{0}$. This is done with a system of constraints of the form $\mathbf{AC}^\perp \mathbf{h}' = \mathbf{0}$, where $\mathbf{A}$ represents the system of linear constraints defined by Equation (2.15). This simple overconstrained system of homogeneous linear equations can be solved as the last column of $\mathbf{V}'$, in the singular value decomposition $\mathbf{AC}^\perp = \mathbf{U}'\mathbf{D}'\mathbf{V}'^\top$. Once $\mathbf{h}'$ is found in this way, the entries of $\mathbf{H}$ are simply determined as $\mathbf{h} = \mathbf{C}^\perp \mathbf{h}'$.

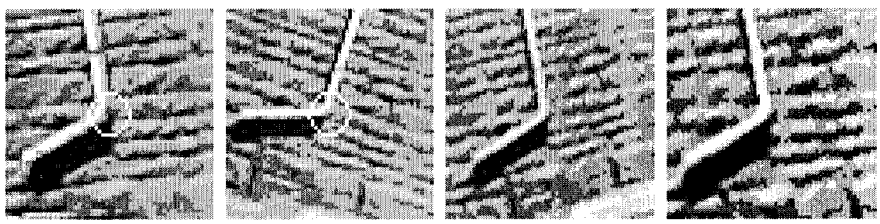
Thus, a good estimate of $\mathbf{H}$ is easily found using only the singular value decomposition and basic matrix operations. All that is required is the SVD of the 9×9 zero-filled matrix $\mathbf{C}$, the product of the 6×9 matrix $\mathbf{A}$ with the 9×5 matrix $\mathbf{C}^\perp$, the SVD of the resulting 6×5 matrix, and finally, the product of $\mathbf{C}^\perp$ with the 5-vector $\mathbf{h}'$. The solution is thus direct and can be efficiently implemented.

	unwarped	**affine warp**	**homog. warp**
Figure 7.2	0.034	0.331	0.675
Figure 7.3	0.180	0.546	0.784

Table 7.1: Correlation between junction neighborhoods shown in Figures 7.2 and 7.3

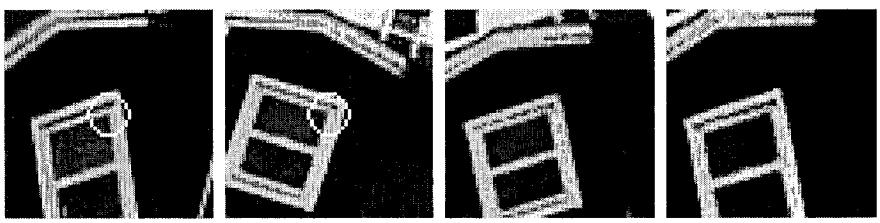
To illustrate the effectiveness of the proposed warping scheme, Figure 7.1 shows two views of some buildings, along with epipolar lines. Figures 7.2 and 7.3 show closeups of some neighborhoods of image junction points that are to be matched. In these figures, images (a) and (b) show the original closeups. It is seen that they are

Figure 7.1: A pair of images with some corresponding epipolar lines. The squares show the regions used in Figures 7.2 and 7.3.



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

Figure 7.2: (a) and (b): closeups of image regions centered on a junction point from each image in Figure 7.1. (c): image (b) warped by a transformation computed using only the junction lines with Equation (6.4). (d): image (b) warped by a homography computed using Equation (7.2).



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

Figure 7.3: (a) and (b): closeups of image regions centered on a junction point from each image in Figure 7.1. (c): image (b) warped by a transformation computed using only the junction lines with Equation (6.4). (d): image (b) warped by a homography computed using Equation (7.2).

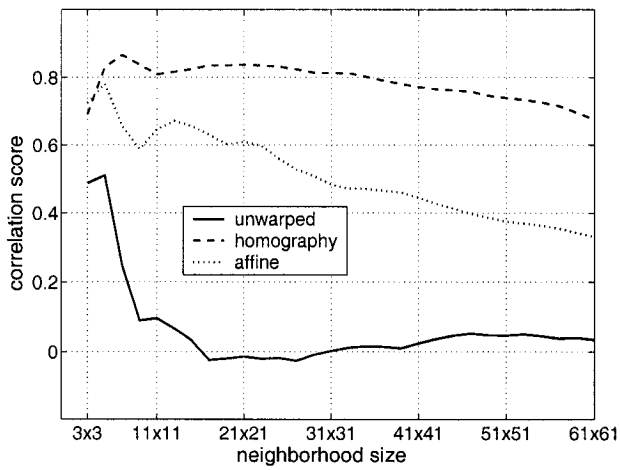


Figure 7.4: The correlation scores between the regions shown in Figure 7.2 taken over neighborhoods of different size around the junction points.

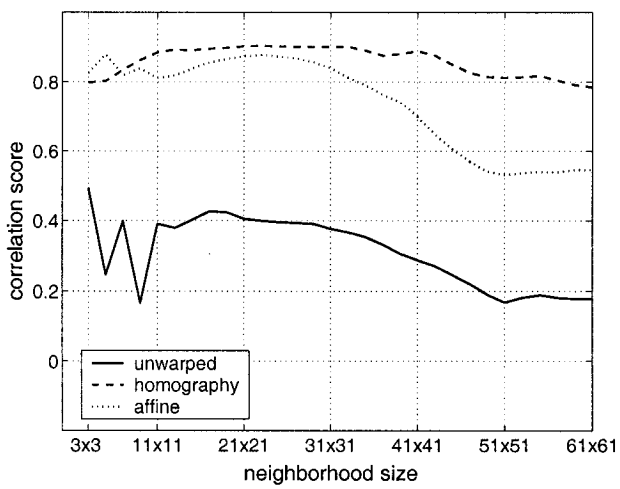


Figure 7.5: The correlation scores between the regions shown in Figure 7.3 taken over neighborhoods of different size around the junction points.

significantly different. In fact, they exhibit VNC scores of 0.03 and 0.18 respectively (see Table 7.1). If the images (b) are warped, prior to correlation, the correlation scores can be greatly improved. The images (d) show the regions from the images (b) warped by homographies obtained from Equation (7.2), and are obviously very similar to the corresponding images (a).

For comparison, the images (c) show the result of warping the images (b) by the affine transformation computed only from the junction lines using Equation (6.4), and not the epipolar geometry, as was done in Subsection 6.4.1. In Figure 7.3, it can be seen how this transformation incorrectly preserves the length of the line segments defining the junctions, while they are stretched appropriately in the images (d). The correlation scores between warped right image regions and the corresponding left image regions are also shown in Table 7.1. The method proposed in this chapter is a clear improvement.

Figures 7.4 and 7.5 show correlation scores between images (a) of the junction neighborhoods shown in Figures 7.2 and 7.3, and the warped regions around feature points shown in images (b), but taken over neighborhoods of different size. It can be seen that the matching scores are improved when using the proposed method defined by Equation (7.2), over the simpler method defined by Equation (6.4), or over using no warp at all.

7.4 Finding Correspondences

The problem of finding corresponding junction points in pairs of images can now be addressed. Firstly, junction points must be detected in both images. Many methods for identifying and characterizing junction points have been proposed, notably [28, 48, 89, 125]. These may be based on differential operators, junction model fitting, or on the use of edge maps. The method used here was developed by Laganière and Elias, and is described in [56]. Of course, the corner detector described in Chapter 6 could also have been used. The detector from [56] was selected here, as an implementation

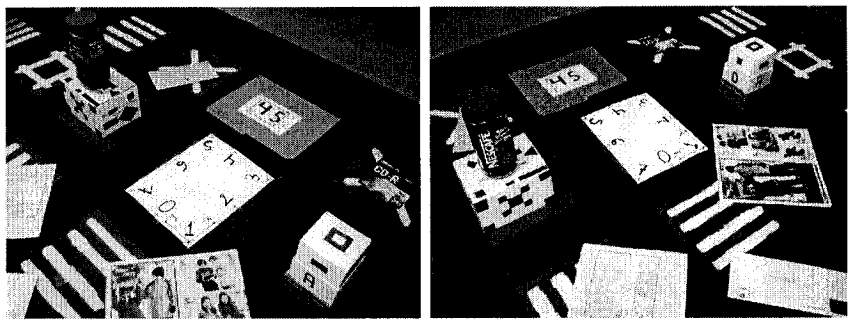


Figure 7.6: An image pair used to test the proposed matching scheme (kindly provided by Hassan Hajjdiab).

was readily available, and because it returns the accurate junction line orientation information that is needed. However, it is expected that any junction point detector would produce acceptable results. The main goal of this section is to show how the shape of the junctions can be used in matching, no matter the junction detector used.

The epipolar geometry of the pairs to be matched is now assumed to be available, so as in Chapter 3, it can be used to guide the search for matches. Thus, only junction points in the second image which lie close to the epipolar line of a junction point in the first image are considered as possible matches. For all candidate junction point matches, a homography is computed according to Equation (7.2). Then, as with the affine transformations described in Subsection 6.4.1, the homography is used to warp the second image, and the region surrounding the junction in this warped image is compared to the region surrounding the junction in the first image. VNC is used to this end, as described in Subsection 2.4.2, and only candidate matches exhibiting a correlation score above some specified threshold are kept. In addition, the simple and advantageous constraints of uniqueness and symmetry are imposed, as described in Subsections 2.4.3 and 2.4.4.

The image pair shown in Figure 7.6 was used as an example. Figure 7.7 shows the results of applying the described matching scheme between some junction points extracted on the images, with 19×19 correlation windows, and different correlation

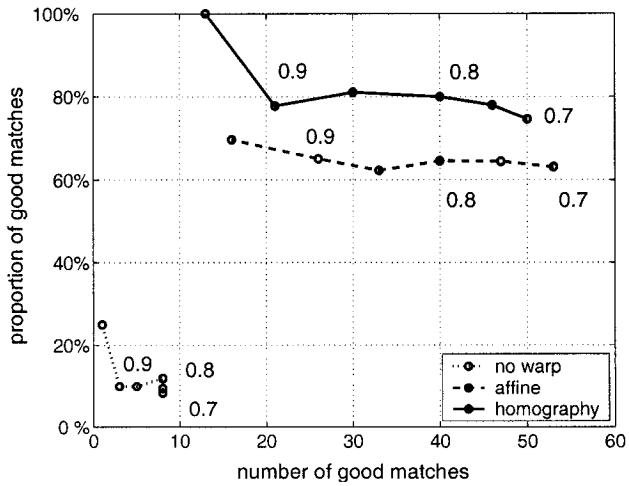


Figure 7.7: The result of matching between the images of Figure 7.6, with different correlation thresholds, on 19 × 19 windows. The number and percentage of good matches obtained are shown.

score thresholds. The same scheme was also applied but using the affine transformation defined by Equation (6.4), instead of the proposed homography. The resulting match sets were of different size and contained varying proportions of mismatches. As in the graphs of Chapter 2, each point on the graph represents the result of an experiment with a different VNC threshold. The vertical axis records the proportion of accurate matches in the result, and the horizontal axis records their number. Matching is successful if it results in a high number of junction pairs, among which few are false matches. It is seen that the homography estimation method described in this chapter compares favorably with the simpler approximations of perspective distortion, which did not require any calibration information.

For comparison, similar experiments were conducted where correlation was applied without any warping. Then, for example, with a correlation score threshold of 0.7, 16 pairs were obtained, 15 of which were mismatches. However, correlation without warping will work better on smaller image patches. Therefore, to obtain a fairer comparison, the values in Figure 7.7, where no warping was used, were obtained with smaller 3 × 3 neighborhoods. These results are nevertheless far inferior to the use of

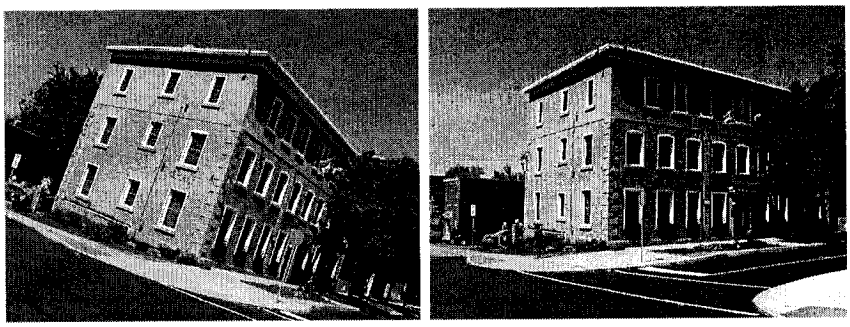


Figure 7.8: An image pair used to test the proposed matching scheme.

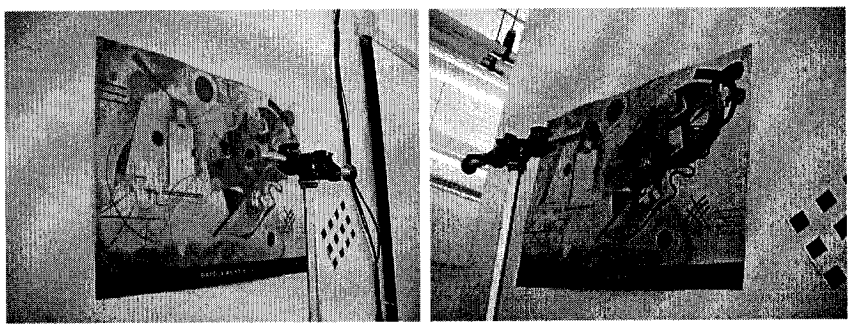


Figure 7.9: An image pair used to test the proposed matching scheme.

the homography or its affine approximation.

Table 7.2 shows the result of applying the matching scheme presented in this section to the four widely separated image pairs shown in Figures 7.1, 7.6, 7.8 and 7.9, with a correlation threshold of 0.7 and using 19×19 neighborhoods. The results of the same scheme, but with the homography estimates replaced by the affine transformations described by Equation (6.4) are also recorded for comparison. There is again a significant improvement, when using the calibrated matching method. The results of the same experiments, but where correlation was applied without prior warping are also shown in Table 7.2. For these experiments, the correlation window size was again reduced to 3×3, but still produced far inferior results.

	No Warp	Using Equation (6.4) (affine warp)	Using Equation (7.2) (homography warp)
Figure 7.1	16.0%	65.5%	71.4%
Figure 7.6	8.2%	63.1%	74.6%
Figure 7.8	33.8%	79.4%	96.4%
Figure 7.9	42.9%	41.9%	57.8%

Table 7.2: Proportion of good matches found between widely separated views, using the no warp, and the warps defined by Equations (6.4) and (7.2) .

7.5 Epipolar Geometry Estimation

Fundamental matrix estimation is a key step in many applications, such as calibration, and 3D reconstruction. In Section 3.2, this was done using a calibration pattern. In Section 2.7, it was shown how RANSAC estimation schemes can be used to estimate the fundamental matrix from point correspondences. Unfortunately, with very widely separated views, it is very difficult to estimate a fundamental matrix because it is difficult to obtain an accurate set of point correspondences. A solution was proposed in Subsection 6.4.2, but a different approach will be proposed here, which will work in the case of very widely separated views, but when a rough approximation of the epipolar geometry is available.

Fundamental matrices can be computed from the camera positions, orientations, and internal parameters. However, this type of procedure is quite sensitive to errors in the parameters, making these direct approaches impractical. The solution proposed here is to first estimate the various extrinsic parameters with some positional sensor, and use them to find a crude estimate of the fundamental matrix. Then, this estimate can be used, with the proposed matching scheme, to find candidate matches between the views. These candidate matches can in turn be used to produce a refined estimate of the sought fundamental matrix.

Figure 7.10 shows an initial estimate of an image pair's epipolar geometry which was obtained from approximate manual measurements of the camera positions and orientations. The proposed matching method was used to improve this fundamental

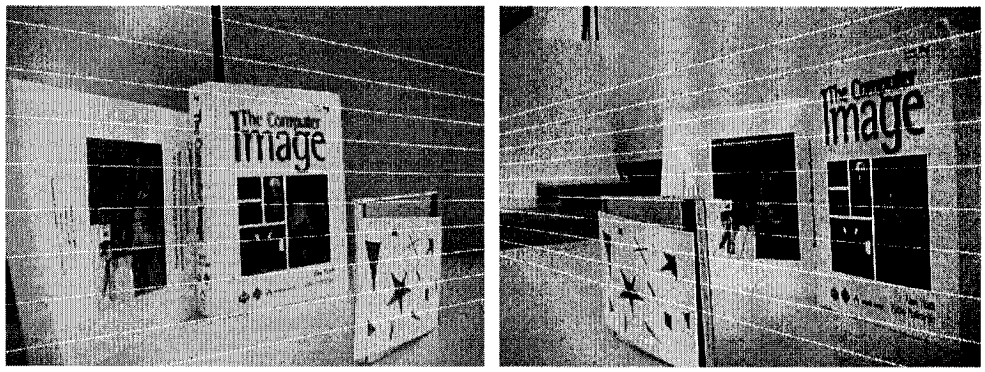


Figure 7.10: An approximation of the epipolar geometry obtained from rough estimates of the cameras relative positions and orientations.

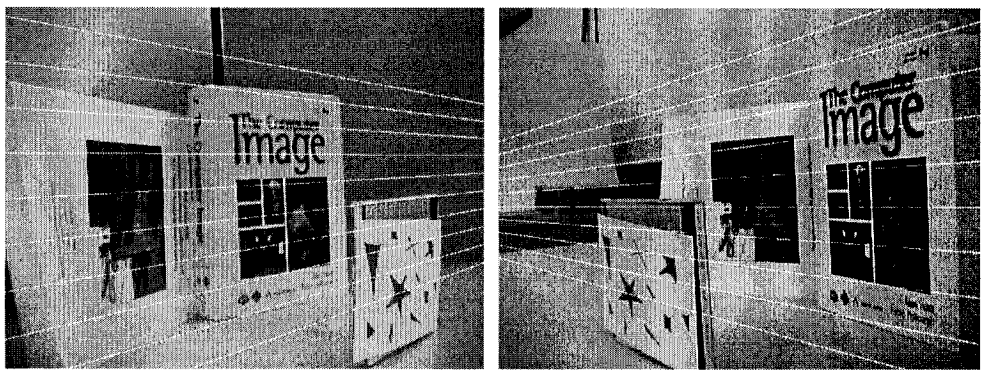


Figure 7.11: Epipolar geometry recovered using matches found through homography warps of junction neighborhoods from the estimate shown in Figure 7.10.

matrix estimate. Four iterations were actually required, where junction points were matched and a fundamental matrix estimated from the result. This was done with the help of the RANSAC-based software of Roth and Whitehead [90], which was described in Subsection 2.2.1. The degree to which corresponding points must agree with the putative epipolar geometries was gradually increased. In the end, a set of 35 matches containing 10 mismatches was obtained and used to compute the final epipolar geometry shown in Figure 7.11. It can be seen that the refined estimate is much better than the original, as the epipolar lines now go through the same points in both views.

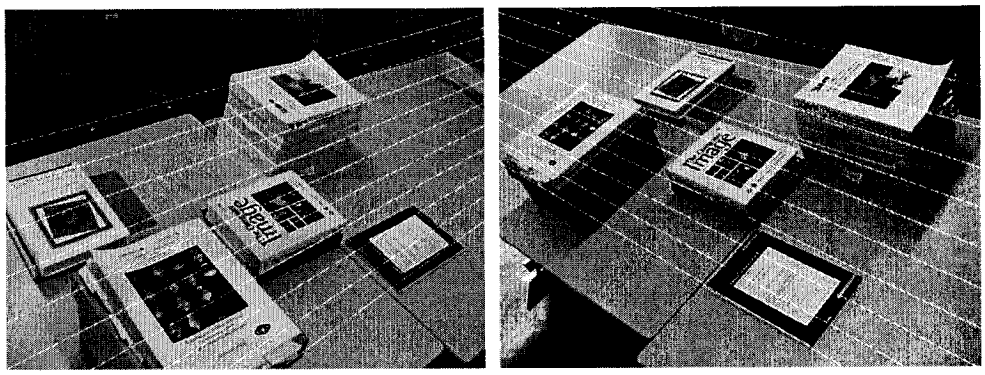


Figure 7.12: An approximation of the epipolar geometry as the epipolar geometry of images taken before a small displacement of one camera.

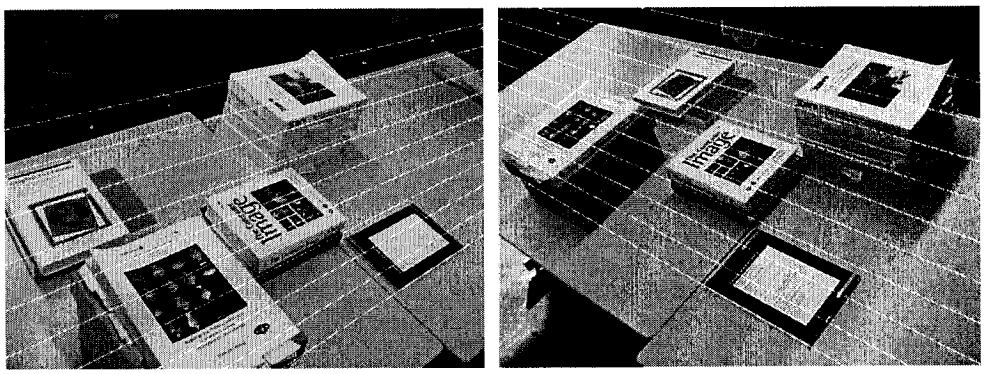


Figure 7.13: Epipolar geometry recovered using matches found through homography warps of junction neighborhoods from the estimate shown in Figure 7.12.

Figures 7.12 and 7.13 show the result of a similar experiment. This time, however, the image pair is the result of a small displacement of one of the cameras in a previously calibrated image pair. Thus, the previous pair's epipolar geometry can serve as a starting point in estimating the new epipolar geometry. Note that it would be impossible to extract enough good matches automatically without using some warping of the junction point neighborhoods, as the views are very widely separated. From the 47 junction point pairs obtained, 16 agreed with the refined epipolar geometry shown in Figure 7.13.

7.6 Tolerance Analysis

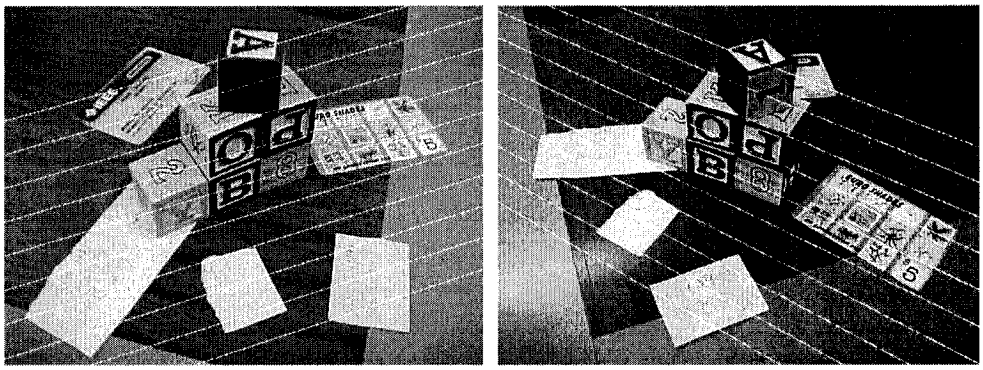


Figure 7.14: An image pair, and its epipolar geometry, which was calibrated and used to test the robustness of the proposed fundamental matrix recovery scheme. This image and its calibration parameters were kindly provided by Rimon Elias.

If the fundamental matrix estimation scheme presented in the previous section is to be used in practice, the matching scheme it relies on must be shown to be tolerant to errors in the original estimate of the fundamental matrix. To demonstrate such tolerance, the image pair shown in Figure 7.14 was used. The correct epipolar geometry is also shown in Figure 7.14. A precise estimate of this image pair's calibration parameters (camera position, orientation and internal parameters) was used to precisely estimate the epipolar geometry relating these views. Then, experiments were conducted where these parameters were perturbed, and the effect on the number of valid matches found by the proposed matching scheme was recorded.

Some results of these trials are recorded in Table 7.3, as well as Figures 7.16 and 7.17. Each point in the graphs of these figures represents an experiment with a particular set of perturbed parameters. Each experiment was conducted with calibration parameters perturbed such that the camera positions are randomly distributed in a cube of 30cm diameter centered on the real position, and such that the camera orientations have their three parameters within 3° of their accurately estimated values.

difference in				F	number	prop.
pos.	tilt	pan	swing	deform.	correct	correct
10.3cm	+1.1°	+1.3°	-1.0°	29.8	27	35.1%
12.7cm	-2.1°	+1.5°	+1.3°	57.5	18	24.7%
5.3cm	-2.3°	-1.1°	-1.4°	45.3	25	34.7%
9.8cm	-1.3°	+0.3°	-0.8°	20.2	24	28.9%
9.1cm	+3.0°	-2.2°	+0.4°	51.9	18	24.3%
7.9cm	+1.2°	-1.5°	-1.8°	51.6	25	30.5%
14.7cm	-1.0°	-0.5°	-1.2°	15.6	28	34.6%
9.8cm	+1.1°	-1.0°	+0.3°	46.8	23	28.4%
14.4cm	-2.7°	-2.9°	+1.5°	28.8	25	32.9%
3.7cm	+0.6°	+1.6°	-1.7°	22.1	25	30.1%
16.9cm	-1.4°	+2.0°	0.5°	14.7	31	35.6%
4.8cm	+0.5°	-1.9°	+0.4°	30.3	25	28.4%
15.4cm	+2.9°	+3.0°	-0.4°	58.3	14	20.6%
9.3cm	+2.0°	+0.5°	-1.0°	59.1	20	25.3%
6.0cm	-1.1°	-1.3°	-1.2°	22.6	24	29.6%
5.3cm	-0.2°	+1.8°	-1.9°	45.5	22	29.3%
18.9cm	-0.4°	+2.5°	-2.9°	28.9	25	34.2%
4.1cm	+1.6°	-1.4°	+1.2°	42.3	21	29.2%
10.2cm	-1.4°	+0.4°	+2.5°	24.6	25	30.9%
8.6cm	-1.6°	-1.1°	+1.3°	22.6	26	29.3%

Table 7.3: Experiments with perturbed parameters on the images of Figure 7.14. The columns represent the camera displacement, change in tilt, pan and swing angles, the resulting **F** deformation described in the text, and the number and proportion of correct matches found after matching junction points between the two images using the perturbed fundamental matrix, while warping with homographies.

For each of these experiments, a measure of the deformation of the resulting fundamental matrix was recorded as the maximal distance between a left-image point's epipolar line, and its corresponding point in the right image. Figure 7.15 shows an example of a perturbed epipolar geometry which corresponds to the test described in the first row of Table 7.3. On this figure an example of the computation of the deformation measure is also shown. It is seen that the feature point which is shown on the left image, has an epipolar line in the right image which runs 29.8 pixels from its expected position on the corresponding point. Since this is the epipolar line in that

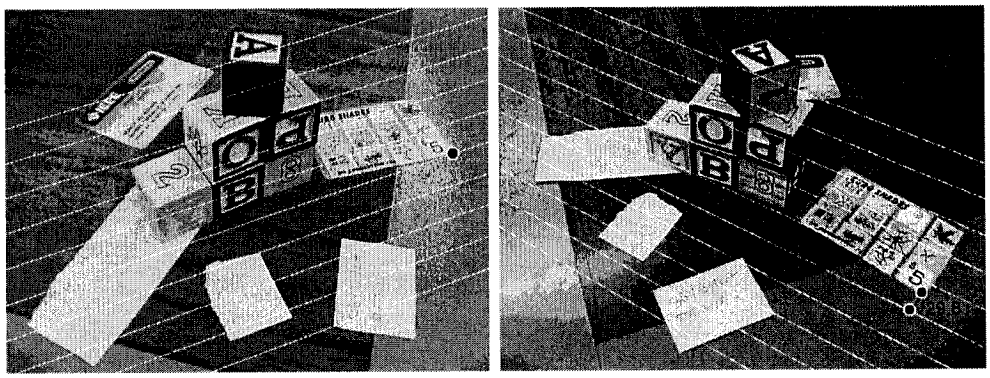


Figure 7.15: An example of a fundamental matrix obtained with perturbed parameters. The points used to evaluate the fundamental matrix deformation are also shown.

pair of images which runs the furthest from one of the feature points it should run through, the distance is taken as the measure of the fundamental matrix deformation.

In the experiments, the accurate matches found by the proposed matching scheme were also counted through visual inspection of the resulting match set. The points in Figure 7.16, have their x-coordinate as the given measure of deformation, and their y-coordinate as the number of accurate matches that were found. In Figure 7.17, the y-coordinates represent the proportion of accurate matches in the resulting match set.

It is seen that for all experiments, an adequate number of matches were found (between 14 and 31), enough to allow for a proper fundamental matrix estimation using a RANSAC-based scheme. In each of these experiments, the number of mismatches also remained reasonable, as it varied between 47 and 63, resulting in a proportion of good matches that was between 20.6% and 35.6%. It is also seen, as expected, that the number of good matches that were found diminishes with the fundamental matrix's deformation, but not too drastically. That is to say, an exact fundamental matrix would be easily recovered with random sampling, given the obtained proportion of correct matches.

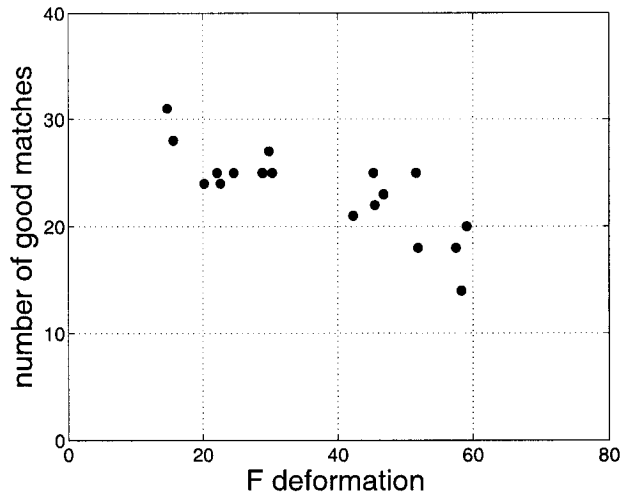


Figure 7.16: Tolerance of the fundamental matrix recovery scheme. Each point corresponds to an experiment with a given set of perturbed parameters. The deformation is the maximal distance between a point in the image pair and the epipolar line on which it should lie. The number of correct matches found is shown as a function of the fundamental matrix deformation.

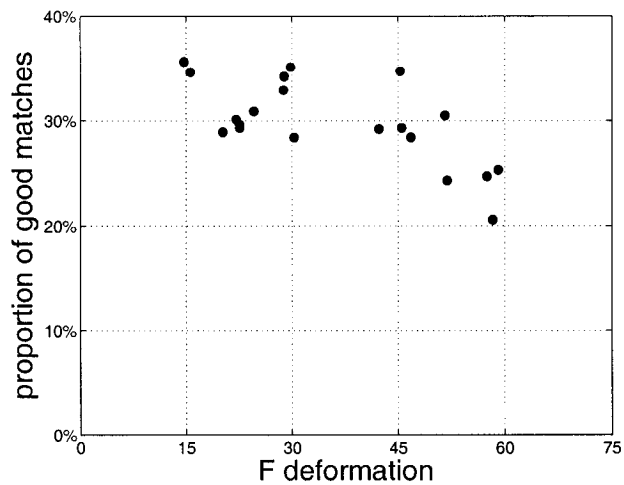


Figure 7.17: Tolerance of the fundamental matrix recovery scheme. The proportion of correct matches found is shown as a function of the fundamental matrix deformation.

7.7 Conclusion

In this chapter, a method for finding correspondences between calibrated widely separated views was described. An application to fundamental matrix recovery was also introduced. The main contributions of this work are:

- A solution to the matching problem in the calibrated widely separated view situation.
- The introduction of a new way of estimating local homographies from the epipolar geometry and the shape of image junctions. Unlike most proposed approaches, this method gives a full homography, not an affine approximation. It is also straightforward and non-iterative.
- The demonstration that the use of the proposed local homography estimation technique produces better results than direct correlation of image neighborhoods, or than the affine transformations described for the uncalibrated case in Section 6.4.1.
- The experimental demonstration that the method is tolerant to inexact fundamental matrices, generated from perturbed camera parameters.
- The introduction of a measure of fundamental matrix deformation, as the maximal distance from an image point to the epipolar line computed using its corresponding point.
- The introduction of a method for fundamental matrix estimation, given an initial approximation.
- The demonstration that the proposed fundamental matrix estimation scheme could be used in practice to calibrate difficult image pairs, when an approximation of the camera positions and orientations can be obtained from positional sensors.

Chapter 8

Conclusion

Feature-based Matching is certainly among the most important and fundamental problems in computer vision. This work targeted the problem under different circumstances: in a calibrated or uncalibrated setup, and between widely or narrowly separated views.

This conclusion chapter will offer a summary of the problems that were addressed in this work, and their relevance for computer vision applications. It will also review the solutions that were proposed, with their advantages and disadvantages, and finally suggest possible directions for future work.

8.1 Justification

Part I In the uncalibrated context, feature-based matching is definitely useful towards calibration. Calibration is usually obtained from point correspondences, and when these correspondences are found automatically from feature points on ordinary scene objects, we speak of autocalibration. Calibration, itself, is needed towards depth estimation, reconstruction, taking measurements in images, virtual viewpoint generation, or the insertion of virtual objects.

The alternatives to autocalibration can be impractical, in many application contexts. Using fixed cameras whose positions and orientation were precisely measured

is not always possible. The use of a calibration pattern is often impractical. Finally, manual extraction of point correspondences is obviously an unpleasant task.

Part II Sparse feature matching is less common in the calibrated case, where dense matching of every possible image point is more common. It can nevertheless be significantly faster to match only key image feature points. If these key points contain most of the structure information, interesting results could be produced while avoiding the overwhelming complexity of treating a dense disparity map. Possible applications for this include fast reconstruction of the basic scene structure, obstacle detection in navigation applications, or measurement from images, towards the insertion of virtual objects in dynamic scenes. The sparse correspondences obtained through calibrated feature matching could also constitute a starting point towards dense matching, when a dense disparity map is unavoidable. For example, in [60], Lhuillier and Quan obtain dense matches from sparse correspondences.

Part III Feature matches between widely separated views are needed mostly towards calibration, as was the case for uncalibrated narrowly separated views. In this context, autocalibration offers the same advantages.

Approaches developed for uncalibrated narrowly separated view matching will not work in the widely separated view context. This is because of the large perspective distortion that now exists between the images. However, it is often necessary to work with widely separated views. This could be because no other views are available (such as in image indexing applications), because of an attempt to minimize the number of images covering a scene, or to obtain more accurate reconstruction results, by separating the images to increase the angles between back-projected rays, and thus improve the localization of their intersection.

Part IV Solutions to matching in the calibrated widely separated view context would allow the extension of the applications suggested in Part II, to more difficult cases. However, the solution that was proposed in this work, despite successfully

identifying several correct matches, did not, in general, identify enough matches for applications related to reconstruction. A new application to fundamental matrix recovery was thus suggested, where the proposed technique was very adequate. The main alternatives to calibrated widely separated views matching towards fundamental matrix recovery would be the fastidious manual identification of correspondence, or the use of uncalibrated techniques, such as those mentioned in Part III. In this latter case, however, some potentially valuable information is obviously being ignored, when some camera position information is available.

8.2 Contributions

The conclusions of the previous chapters included detailed lists of important contributions of this work. A simple summary of these contributions is now included.

Part I Chapter 2 presented an empirical framework for evaluating matching constraints. It then used it to compare several strategies. It also introduced the background elimination constraints, and a non-iterative way of using constraints based on relative displacements. Finally, it briefly examined the advantages of using planar homography detection prior to epipolar geometry estimation.

Part II Chapter 3 presented a basic scheme for calibrated sparse matching. The use of a third image was introduced, as a way to confirm the validity of matches between the first two. Chapter 4 took the framework defined in Chapter 3, and proposed changes to improve its results, in view of applications in quick reconstruction. It introduced epipolar gradient features, and matching based on edge transfer.

Part III Chapter 5 consisted in a comprehensive survey of the widely separated view matching literature. Then, Chapter 6 introduced the wedge-based corner detector. It also presented the idea of using the lines that define a corner to constrain an

affine transformation which approximates perspective distortion between corresponding corner neighborhoods. It then showed how these local affinities can be used in matching widely separated views.

Part IV Chapter 7 expanded the idea of the previous chapter in the calibrated context. It derived more accurate approximations of local perspective distortion around junction points as homographies constrained by the epipolar geometry. It then justified the goal of widely separated view matching in the calibrated context with an application to fundamental matrix recovery.

8.3 Evaluation

Part I This work on uncalibrated narrowly separated view matching provided a very accurate evaluation of matching constraints, thanks to its use of ground truth sets. Such an approach is far superior to the use of homographies, or the epipolar geometry, in verifying matches. However, the disadvantage was the expense of producing such accurate sets.

Part II The solution proposed for the calibrated matching problem using epipolar gradient features and edge transfer proved very effective. It gives superior results to the use of more traditional tools in this context, when it comes the the number and distribution of feature point matches. The method can be applied in most calibrated settings, although it still requires the presence of lines in the scene.

However, it required the use of Gaussian filters, in estimating image derivatives accurately, and was thus slightly more expensive than the use of Harris detectors. Another small disadvantage is that it requires the epipolar geometry to be precisely known. If it was inaccurate, points further from computed epipolar lines would have to be considered, and these may be harder to distinguish from each other.

Theoretically, epipolar gradient features are not fully preserved by larger changes in viewpoint, and therefore would not be completely appropriate in that context,

although in practice, they are not much worse than Harris features in this respect.

Part III The proposed method, for matching points among uncalibrated widely separated views, allowed a solution to the problem in many difficult matching situations. Furthermore, the wedge-based detector provided a tool to acquire the required corner shape information.

Disadvantages of this approach are that it will only work on images containing clear corners (a fault of most feature-based matching schemes), and that it is not fully invariant to homographies of feature point neighborhoods, but rather to 4 DOF approximations. This can be seen in Figures 7.2 and 7.3 where the warped images are compared to those produced by full homographies.

Another problem that should be noticed, and that exists with all methods based on similarity measures that are made robust to wide changes in viewpoints, is a loss in discriminating power. One common example of this is found when two views of a square object are compared. Then, the corners of this square will often be wrongly paired with all of their four counterparts. This problem is however much worse when invariant characterizations are used, as opposed to the proposed warping scheme.

Part IV In the calibrated widely separated context, the proposed solution could relatively easily find matches where it would otherwise be impossible. It also allowed the recovery of the epipolar geometry in very difficult situations.

One of its disadvantages would be its need for the presence of easily detected junction points in the images. Mostly for this reason, it usually did not find large numbers of matches between considered image pairs.

Finally, it should be reminded that all the methods studied in this work eventually fail in difficult situations. They could encounter problems with non rigid movement, occluded objects, large changes in illumination, or repeated structures in the scene.

8.4 Future Work

Part I Although some satisfying solutions exist in the uncalibrated narrowly separated context, there is much room for improvement. One aspect which could be improved is in the formulation of optimization criteria to estimate calibration information from feature matches, after they are obtained. Such estimates are often highly unstable, despite their sometimes great complexity.

Another area of interest would also be in the manner of combining several constraints into a similarity criteria, Mahalanobis distances are the usual way of doing this. They account for the correlation between different criteria; but not for the relative uniqueness of characteristics in sets of potential matches, or the precision of measurements.

Part II Now that an efficient and accurate sparse feature matching scheme was proposed in the calibrated narrowly separated context, what is mostly needed are tools that will benefit from the information. For example, ways of producing visually pleasing models from sparse data would be needed.

Improvements could also be conceived in the way searches are conducted along epipolar lines. Feature points could be sorted, and placed in a tree structure, according to the epipolar geometry. Currently, when points are sought along an epipolar line, all feature points are tested, but such a structure would allow the reduction of the complexity of the search to logarithmic order.

Finally, the work presented in Chapter 4 has yet to be implemented in continuous mode. What is mostly needed to achieve this is the implementation of a procedure to obtain an accurate estimate of the trifocal tensor, good enough to allow edge transfer.

Part III This is the most active research area studied in this work. Many applications assume calibration, where no viable means of obtaining it exist. Widely separated view matching could allow auto-calibration in these situations.

The wedge-based corner detector could be improved with a more complex fitting

scheme. However, because of efficiency issues, this is not necessarily desirable. More precise subpixel localization of the feature points might also be achieved. When it comes to matching the detected corners, the proposed approach should eventually incorporate more constraints, such as unicity, symmetry, and some measure of consistency among neighboring pairs, perhaps related to the disparity gradient constraint.

We believe that solutions to the widely separated view problem based on warping and correlation have more potential than the more popular use of invariant characterization, because they allow more complete preservation of the information found in feature point neighborhoods. Using junction shapes, as was proposed, is promising, but could be further improved from the recovery of a full affine transformation, rather than the 4 DOF approximation found here. Other solutions that claim to achieve this are too complex and computationally expensive, and not general enough. What is needed is a way to take a measurement in an affine-invariant way, independently along each side of corresponding corners, to solve the 6 DOF of an affine transformation.

Part IV Calibrated widely separated view matching is a context that was introduced in this work, and appears very promising, as seen in the given examples. The challenge will be to find ways of obtaining more matches, probably through a more stable estimation of the local homographies. Better ways of using the constraints in the optimization scheme for the local transformations probably exist. These could be based on a minimization of reprojection errors, rather than the algebraic constraints that were used.

Bibliography

[1] M. Agrawal, L. Davis, Trinocular stereo using shortest paths and the ordering constraint *Proc. Stereo and Multi-Baseline Vision*, pp. 3-9, 2001.

[2] H. Alhichri, M. Kamel, Image Registration using the Hausdorff Fraction and Virtual Circles, *Proc. Int. Conf. on Image Processing*, vol. 2, pp. 367-370, 2001.

[3] P. Anandan, Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion, *Proc. DARPA Image Understanding Workshop* pp. 236-246, 1984.

[4] S. Arbouche, Feature Point Correspondences: A Matching Constraints Survey, M.C.S. Thesis, *University of Ottawa*, 1999.

[5] S. Barnard, W. Thompson, Disparity Analysis of Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, pp. 333-340, 1980.

[6] A. Baumberg, Reliable Feature Matching Across Widely Seperated Views, *Proc. Computer Vision and Pattern Recognition*, vol. 1, pp. 774-781, 2000.

[7] P. Beardsley, P. Torr, A. Zisserman, 3D Model Acquisition from Extended Image Sequences, *Proc. European Conf. on Computer Vision*, LNCS 1065, pp. 683-695, 1996.

[8] A. Berg, J. Malik, Geometric Blur for Template Matching, *Proc. Computer Vision and Pattern Recognition*, vol. 1, pp. 607-614, 2001.

[9] V. Berzins, Accuracy of Laplacian Edge Detectors, *Computer Vision, Graphics & Image Processing*, vol. 27, no. 2, pp. 195-210, 1984.

[10] D. Bhat, S. Nayar, Ordinal Measures for Image Correspondence, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 4, pp. 415-423, 1998.

[11] P. Biber, W. Straßer Solving the Correspondence Problem by Finding Unique Features, *Proc. Vision Interface*, 2003.

[12] T. Binford, T Levitt, Quasi-Invariants: Theory and Exploitation, *Proc. DARPA Image Understanding Workshop*, pp. 819-829, 1993.

[13] R. Bolles, R. Cain, Recognizing and Locating Partially Visible Objects, the Local-Feature-Focus Method, *Int. Jour. of Robotics Research*, vol. 5, no. 3, pp. 3-26, 1986.

[14] L. Brown, A Survey of Image Registration Techniques, *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.

[15] T.-J. Cham, R. Cipola, A Statistical Framework for Long-Range Feature Matching in Uncalibrated Image Mosaicing, *Proc. Computer Vision and Pattern Recognition*, pp. 442-447, 1998.

[16] C. Chou, Y. Chen, Moment-Preserving Pattern Matching, *Pattern Recognition*, vol. 23, no. 5, pp. 461-474, 1990.

[17] A. Criminisi, I. Reid, A. Zisserman, A Plane Measuring Device, *Image and Vision Computing*, vol. 17, no. 8, pp. 625-634, 1999.

[18] T. Darrell, A Radial Cumulative Similarity Transform for Robust Image Correspondence, *Proc. Computer Vision and Pattern Recogintion*, pp. 656-662, 1998.

[19] E. Delp, O. Mitchell, Image Compression Using Block Truncation Coding, *IEEE Trans. on Communication* vol. 27, no. 9, pp. 1335-1342, 1979.

[20] R. Deriche, T. Blaszka, Recovering and Characterizing Image Features Using an Efficient Model Based Approach, *Proc. Computer Vision and Pattern Recognition*, pp. 530-535, 1993.

[21] R. Deriche, Z. Zhang, Q.-T. Luong, O. Faugeras, Robust Recovery of the Epipolar Gepmetry for an Uncalibrated Stereo Rig, *Proc. European Conf. on Computer Vision*, LNCS 800, pp. 567-576, 1994.

[22] Y. Dufournaud, C. Schmid, R. Horaud, Matching Images with Different Resolutions, *Proc. Computer Vision and Pattern Recogintion*, vol. 1, pp. 612-618, 2000.

[23] O. Faugeras, M. Berthod, Improving Consistency and reducing Ambiguity in Stochastic Labeling: An Optimization Approach, em IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 3, pp. 412-423, 1981.

[24] O. Faugeras, Q. Luong, S. Maybank, Camera Self-calibration: Theory and Experiments, *Proc. European Conf. on Computer Vision*, LNCS 588, pp. 321-334, 1992.

[25] J. Fdez-Valdivia, J. Garcia, J. Martinez-Baena, X. Fdez-Vidal, The Selection of Natural Scales in 2D Images Using Adaptive Gabor Filtering, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 458-469, 1998.

[26] M. Fischler, R. Bolles, Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography, *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, 1981.

[27] W. Freeman, E. Adelson, The Design and Use of Steerable Filters, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, 1991.

[28] W. Föerstner, A Framework for Low Level Feature Extraction, *Proc. European Conf. on Computer Vision*, LNCS 801, pp. 383-394, 1994.

[29] P. Fua, Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities. *Proc. Int. Joint Conf. on Artifical Intelligence*, pp. 1292-1298, 1991.

[30] N. Georgis, M. Petrou, J. Kittler, On the Correspondence Problem for Wide Angular Separation of Non-coplanar Points, *Image and Vision Computing*, vol. 16, no. 1, pp. 35-41, 1998.

[31] J. Gluckman, S. Nayar, Rectifying transformations that minimize resampling effects *Proc. Computer Vision and Pattern Recognition*, vol. 1, pp. 111-117, 2001.

[32] A. Goshtasby, S. Gage, J. Bartholic, A Two-Stage Cross Correlation Approach to Template Matching, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 3, pp. 374-378, 1984.

[33] V. Govindu, C. Shekhar, Alignment using Distributions of Local Geometric Properties, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1031-1043.

[34] A. Guiducci, Corner Characterization by Differential Geometry Techniques, *Pattern Recognition Letters*, vol. 8, pp. 311-318, 1988.

[35] D. Hall, V. de Verdière, J. Crowley, Object Recognition Using Coloured Receptive Fields, *Proc. European Conf. on Computer Vision*, LNCS 1842, pp. 164-177, 2000.

[36] B. Hansen, B. Morse, Multiscale Image Registration Using Scale Trace Correlation, *Proc. Computer Vision and Pattern Recogintion*, vol. 2, pp. 202-208, 1999.

[37] R. Haralick, L. Shapiro, The Consistency Labeling Problem: Part II, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 2, no. 1, pp. 193-203, 1980.

[38] C. Harris, Determination of Ego-motion from Matched Points, *Proc. Alvey Vision Conf.*, pp. 189-192, 1987.

[39] C. Harris, M. Stephens, A Combined Corner and Edge Detector, *Proc. Alvey Vision Conf.*, pp. 147-151, 1988.

[40] R. Hartley. In Defense of the Eight-Point Algorithm, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 133-135, 1997.

[41] R. Hartley, Theory and Practice of Projective Rectification, *International Journal of Computer Vision*, vol. 35, no. 2, pp. 1-16, 1999.

[42] R. Hartley, A. Zisserman, Multiple View Geometry, first ed., Cambridge University Press, Cambridge, 2000.

[43] D. Hilbert, Theory of Algebraic Invariants, Cambridge Mathematica Library, Cambridge University Press, 1890.

[44] R. Horaud, O. Monga, Vision par ordinateur - Outils fondamentaux, second ed., Hermes, Paris, 1995.

[45] B. Horn, B. Schunk, Determining Optical Flow, *Artificial Intelligence*, vol. 20, pp. 199-228, 1981.

[46] Y. Hsieh, D. McKeown, F. Perlant, Performance Evaluation of Scene Registration and Stereo Matching for Cartographic Feature Extraction, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 214-237, 1992.

[47] X. Hu, N. Ahuja, Matching Point Features with Ordered Geometric, Rigidity, and Disparity Constraints. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 1041-1049, 1994.

[48] R. Hummel, L. Parida, D. Geiger, Junctions: Detection, Classification, and Reconstruction, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 687-698, 1998.

[49] R. Hummel, S. Zucker, On the Foundation of Relaxation Labeling Process, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 5, no. 3, pp. 267-286, 1983.

[50] G. Jones, Constraint, Optimization, and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features, *Computer Vision and Image Understanding*, vol. 65, no. 1, pp. 57-58, 1997.

[51] I.-K. Jung, S. Lacroix, A Robust Interest Points Matching Algorithm, *Proc. Int. Conf. on Computer Vision*, vol. 2, pp. 538-543, 2001.

[52] L. Kitchen, A. Rosenfeld, Discrete Relaxation for Matching Relational Structures, *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, no. 12, pp. 869-874, 1979.

[53] R. Klette, K.Schluns, A. Koschan, Computer Vision: Three-Dimensional Data from Images, Sprigner, 1996.

[54] A. Koschan, A Framework for Area Based and Feature Based Stereo Vision, *Machine Graphics & Vision*, vol. 2, no. 4, 1993.

[55] R. Laganière, A Morphological Operator for Corner Detection, *Pattern Recognition*, vol. 31, no. 11, pp. 1643-1652, 1998.

[56] R. Laganière, R. Elias, JUDOCA: A Fast Junction Detection Operator, *Proc. International Conference on Acoustics, Speech and Signal Processing* 2004.

[57] R. Laganière, E. Vincent, Wedge-based Corner Model for Widely Separated Views Matching, *Proc. Int. Conf. on Pattern Recognition*, vol. 3, pp. 672-675, 2002.

[58] S. Lai, M. Fang, Robust and Efficient Image Alignment with Spatially Varying Illumination Models, *Proc. Computer Vision and Pattern Recogintion*, vol. 2, pp. 167-172 1999.

[59] M. Lhuillier, L. Quan, Robust dense matching using local and global geometric constraints, Lhuillier, M.; Long Quan, *Proc. Int. Conf. on Pattern Recognition*, vol. 1, pp. 968-972, 2000.

[60] M. Lhuillier, L. Quan, Image-Based Rendering by Joint View Triangulation, Lhuillier, M.; Long Quan, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1051-1063, 2003.

[61] D. Liebowitz, A. Criminisi, A. Zisserman, Creating Architectural Models from Images, *Proc. Eurographics, Computer Graphics Forum*, vol 18, no. 3, pp. 39-50, 1999.

[62] T. Lindeberg, On Scale Selection for Differential Operators, *Proc. Scandinavian Conf. on Image Analysis*, pp. 857-866, 1993.

[63] T. Lindeberg, J. Gårding, Shape-Adapted Smoothing in Estimation of 3-D Depth Cues from Affine Distortions of Local 2-D Brightness Structure *Image and Vision Computing*, vol. 15, no. 6, pp. 415-434, 1997.

[64] C. Loop, Z. Zhang, Computing rectifying homographies for stereo vision, *Proc. Computer Vision and Pattern Recognition*, vol. 1, pp. 125-131, 1999.

[65] M. Lourakis, S. Orphanoudakis, Visual Detection of Obstacles Assuming a Locally Planar Ground, *Proc. Asian Conf. on Computer Vision*, LNCS 1352, pp. 527-534, 1998.

[66] D. Lowe, Object Recognition from Local Scale-Invariant Features, *Proc. Int. Conf. on Computer Vision*, vol. 2, pp. 1150-1157, 1999.

[67] Q.-T. Luong, R. Deriche, O. Faugeras, T. Papadopoulo, On determining the Fundamental matrix: analysis of different methods and experimental results, *INRIA Technical Report*, RR-1894, 1993.

[68] Q.-T. Luong, O. Faugeras, Determining the Fundamental Matrix with Planes: Unstability and New Algorithms, *Proc. Computer Vision and Pattern Recognition*, pp. 489-494, 1993.

[69] Q.-T. Luong, T. Viéville, Canonical Representations for the Geometries of Multiple Projective Views, *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 193-229, 1996.

[70] J. Malik, S. Belongie, J. Shi, T. Leung, Textons, Contours and Regions: Cue Integration in Image Segmentation, *Proc. Int. Conf. on Computer Vision*, vol. 2, pp. 918-925, 1999.

[71] D. Marr, Early Processing of Visual Information, *Phil. Trans. of the Royal Society of London*, vol. B-275, pp. 483-519, 1976.

[72] J. Matas, Š. Obdržálek, O. Chum, Local Affine Frames for Wide-Baseline Stereo, *Proc. Int. Conf. on Pattern Recognition*, vol. 4, pp. 363-366, 2002.

[73] K. Mikolajczyk, C. Schmid, Indexing based on Scale Invariant Interest Points, *Proc. Int. Conf. on Computer Vision*, vol. 1, pp. 525-531, 2001.

[74] K. Mikolajczyk, C. Schmid, An Affine Invariant Interest Point Detector, *Proc. European Conf. on Computer Vision*, LNCS 2350, pp. 128-142, 2002.

[75] F. Mindru, T. Moons, L. Van Gool, Recognizing Color Patterns Irrespective of Viewpoint and Illumination, *Proc. Computer Vision and Pattern Recogintion*, vol. 1, pp. 368-372, 1999.

[76] Model House Image Sequence, www.robots.ox.ac.uk/~vgg/data/

[77] P. Montesinos, V. Gouet, R. Deriche, D. Pelé, Matching Color Uncalibrated Images Using Differential Invariants, *Image and Vision Computing*, vol. 18, no. 9, pp. 659-671, 2000.

[78] H. Nishihara, T. Poggio, Stereo Vision for Robotics, *Proc. Int. Symp. on Robotics Research*, pp. 489-505, 1984.

[79] J. Noble, Finding Corners, *Image and Vision Computing*, vol. 6, no. 2, pp. 121-128, 1988.

[80] OpenCV Library, freely available at developer.intel.com

[81] Y. Ohta, T. Kanade, Stereo by Intra- and Inter-Scanline Search, em IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 7, no. 2, pp. 139-154, 1985.

[82] T. Papadopoulo, O. Faugeras, A New Characterization of the Trifocal Tensor, *Proc. European Conf. on Computer Vision*, LNCS 1406, pp. 109-123, 1998.

[83] L. Parida, D. Geiger, R. Hummel, Junctions: Detection, Classification, and Reconstruction, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 687-698, 1998.

[84] M. Pollefeys, R. Koch, L. Van Gool, A Simple and Efficient Rectification Method for General Motion, *Proc. Int. Conf. on Computer Vision*, pp. 496-501, 1999.

[85] P. Pritchett, A. Zisserman, Wide Baseline Stereo Matching, *Proc. Int. Conf. on Computer Vision*, pp. 754-760, 1998.

[86] L. Robert, M. Buffa, M. Hebert, Weakly-calibrated stereo perception for rover navigation Robert, *Proc. Int. Conf. on Computer Vision*, pp. 46-51, 1995

[87] L. Robert, C. Zeller, O. Faugeras, M. Hebert, Applications of Non-metric Vision to some Visually-guided Robotics Tasks, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, pp. 89-134, 1997.

[88] K. Rohr, Modelling and Identification of Characteristic Intensity Variations, *Image Vision Computing*, vol. 10, pp. 66-76, 1992.

[89] K. Rohr, Recognizing Corners by Fitting Parametric Models, *Int. Jour. of Computer Vision* , vol. 9, no. 3, pp. 213-230, 1992.

[90] G. Roth, A. Whitehead, Using Projective Vision to Find Camera Positions in an Image Sequence, *Proc. Vision Interface*, pp. 225-232, 2000.

[91] S. Roy, J. Meunier, I. Cox, Cylindrical Rectification to Minimize Epipolar Distortion, *Proc. Computer Vision and Pattern Recognition*, pp. 393-399, 1997.

[92] F. Schaffalitzky, A. Zisserman, Viewpoint Invariant Texture Matching and Wide Baseline Stereo, *Proc. Int. Conf. on Computer Vision*, vol. 2, pp. 636-643, 2001.

[93] D. Scharstein, R. Szeliski, R. Zabih, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *Proc. Stereo and Multi-Baseline Vision*, pp. 131-140, 2001.

[94] S. Scherer, P. Werth, A. Pinz, The Discriminatory Power of Ordinal Measures - Towards a New Coefficient, *Proc. Computer Vision and Pattern Recogintion*, vol. 1, pp. 76-81, 1999.

[95] C. Schmid, Appariement d'images par invarients locaux de niveaux de gris, Ph.D. Thesis, *Institut National Polytechnique de Grenoble*, 1996.

[96] C. Schmid, Constructing Models for Content-Based Image Retrieval, *Proc. Computer Vision and Pattern Recogintion*, vol. 2, pp. 39-45, 2001.

[97] C. Schmid, R. Mohr, Local Grayvalue Invariants for Image Retrieval, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530-535, 1997.

[98] C. Schmid, R. Mohr, C. Bauckhage, Comparing and Evaluating Interest Points, *Proc. Int. Conf. on Computer Vision*, pp. 230-235, 1998.

[99] C. Schmid, A. Zisserman, Automatic Line Matching across Views, *Proc. Computer Vision and Pattern Recogintion*, pp. 666-671, 1997.

[100] A. Shashua, Trilinearity in Visual Recognition by Alignment, *Proc. European Conf. on Computer Vision*, LNCS 800, pp. 479-484, 1994.

[101] X. Shen, P. Palmer, Uncertainty Propagation and the Matching of Junctions as Feature Groupings, em IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1381-1395, 2000.

[102] S. Smith, J. Brady, SUSAN - A New Approach to Low Level Image Processing, *Int. Journal of Computer Vision*, vol. 23, no. 1, pp. 45-78, 1995.

[103] T. Stockham, Image Processing in the Context of a Visual Model, *Proc. of IEEE*, vol. 60, no. 7, pp. 828-842, 1972.

[104] B. Super, W. Klarquist, Patch-Based Stereo in a General Binocular Viewing Geometry, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 247-252, 1997.

[105] D. Tell, S. Carlsson, Wide Baseline Point Matching Using Affine Invariants Computed from Intensity Profiles, *Proc. European Conf. on Computer Vision*, LNCS 1842, pp. 814-828, 2000.

[106] P. Torr, C. Davidson, IMPSAC: Synthesis of Importance Sampling and Random Sample Consensus, *Proc. European Conf. on Computer Vision*, LNCS 1843, pp. 819-833, 2000.

[107] P. Torr, A. Zisserman, Robust Computation and Parameterization of Multiple View Relations, *Proc. Int. Conf. on Computer Vision*, pp. 727-732, 1998.

[108] P. Torr, A. Zisserman, S. Maybank, Robust Detection of Degenerate Configurations for the Fundamental Matrix, *Computer Vision and Image Understanding*, vol. 71, no. 3, pp. 312-333, 1998.

[109] B. Triggs, Autocalibration from Planar Scenes, *Proc. European Conf. on Computer Vision*, LNCS 1406, pp. 89-105, 1998.

[110] E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision, *Prendice Hall*, 1998.

[111] T. Tuytelaars, L. Van Gool, Matching Widely Separated Views based on Affinely Invariant Neighborhoods, to appear in *Computer Vision and Image Understanding*, 2004.

[112] T. Tuytelaars, L. Van Gool, L. D'haene, R. Koch, Matching of Affinely Invariant Regions for Visual Servoing, *Proc. Int. Conf. on Robotics and Automation*, vol. 2, pp. 1601-1606, 1999.

[113] E. Vincent, R. Laganière, Junction Matching anf F Matrix Recovery in Widely Separated Views, submited to *Proc. Computer Vision and Pattern Recognition*, 2004.

[114] E. Vincent, R. Laganière, Models From Image Triplets using Epipolar Gradient Features, selected for a special issue of *Image and Vision Computing.*

[115] R. Laganière, E. Vincent, Detecting and Matching Feature Points, to be published in *Jour. of Visual Communication and Image Representation.*

[116] E. Vincent, R. Laganière, G. Roth, Widely Separated View Matching : A Litterature Review, submited to *Image and Vision Computing.*

[117] E. Vincent, R. Laganière, Matching with Epipolar Gradient Features and Edge Transfer, *Proc. Int. Conf. on Image Processing*, vol. 1, pp. 277-281, 2003

[118] E. Vincent, R. Laganière, Models From Image Triplets using Epipolar Gradient Features, *Proc. Vision, Video and Graphics*, pp. 143-150, 2003

[119] E. Vincent, R. Laganière, Matching Feature Points for Telerobotics, *Proc. Haptic Audio Video Environments*, pp. 13-18, 2002.

[120] E. Vincent, R. Laganière, An Empirical Study of Some Feature Matching Strategies, *Proc. Vision Interface*, pp. 139-145, 2002.

[121] E. Vincent, R. Laganière, Matching Feature Points in Stereo Pairs: A Comparative Study of Some Matching Strategies, *Machine Graphics & Vision*, vol. 10, no. 3, pp. 237-259, 2001.

[122] E. Vincent, R. Laganière, Detecting Planar Homographies in an Image Pair, *Proc. Image and Signal Processing and Analysis*, pp. 182-187, 2001.

[123] P. Viola, W. Wells, Alignment by Maximization of Mutual Information, *Proc. Int. Conf. on Computer Vision*, pp. 16-23, 1995.

[124] B. Wang, K. Sung, T. Ng The Localized Consistency Principle for Image Matching under Non-Uniform Illumination Variation and Affine Distortion, *Proc. European Conf. on Computer Vision*, LNCS 2350, pp. 205-219, 2002.

[125] W. Yu, K. Daniilidis, G. Sommer, Rotated Wedge Averaging Method for Junction Characterization, *Proc. Computer Vision and Pattern Recognition*, pp. 390-395, 1998.

[126] Z. Zhang, Le problème de la mise en correspondence: L'etat de l'art, *INRIA Technical Report*, RR-2146, 1993.

[127] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry, *Artificial Intelligence*, vol. 78, no. 1-2, pp. 87-119, 1995.

[128] I. Zoghlami, O. Faugeras, R. Deriche, Using Geometric Corners to Build a 2D Mosaic from a set of Images, *Proc. Computer Vision and Pattern Recognition*, pp. 420-425, 1997.