Correspondence-free Stereo Vision

by

YUAN, Ding

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

in

Automation and Computer-Aided Engineering

December 2003

ABSTRACT

Abstract of thesis entitled:

Correspondence-free Stereo Vision

Submitted by YUAN Ding

For the degree of Master of Philosophy

at The Chinese University of Hong Kong in December, 2003

Stereo vision is an important cue that allows 3D information of a scene to be reconstructed from two (or more) images visually. The established approach of stereo vision requires matching feature points across the input images such that each of the matched pair are projected by the same 3D entity in space. Such a correspondence problem is a classical problem of computer vision and has been studied for decades. Yet, extensive research shows a perfect solution to the correspondence problem is often difficult due to possible corruption of the image data by noise, and occurrence of complications like occlusions and repetitive patterns in the image data.

In an elegant work, Aloimonos and Herve (1990) presented an algorithm that could recover the 3D structure of a single planar surface without establishing point-to-point correspondences. The method was the only algorithm that could bypass the difficult stereo correspondence problem. Yet, their algorithm is restricted to images taken under the parallel-axis stereo geometry.

In this thesis we first generalize the parallel-axis stereo geometry into general stereo geometry, and propose a new algorithm for the recovery of a single planar surface. The

algorithm requires no point-to-point correspondence but only group-to-group ones. The algorithm puts no specific restriction on the stereo configuration of the cameras.

To deal with mistakes in feature group matching and so on, we also introduce the use of singular value decomposition and robust estimation techniques to enhance the performance of the algorithm.

The above algorithm is however applicable only to distant or planar scene. We also extend the algorithm to make it capable of reconstructing even multi-surface scene. As a whole, we have come up with a system that could reconstruct the 3D structure of polyhedral scene from images captured by arbitrarily positioned cameras, which is without the need of establishing point-to-point correspondences.

Our correspondence-free stereo vision method can be used in combination with the traditional stereo vision methods. It is particularly useful when one-to-one correspondences are difficult to obtain for reasons like the existence of repetitive patterns in the scene.

The performance of the system is illustrated with experimental results on synthetic and real scenes.

摘　要

立體視覺是從兩張（多張）圖像中使得景物三維信息得以重構的重要提示。已有的立體事覺需要在輸入的圖像中匹配特徵點，從而使得每一對匹配的特徵點都是由三維空間中同一物點投射而來的。在計算機視覺著一領域，匹配是一個已經被研究了幾十年的經典問題。然而，大量的研究顯示，由於噪聲引起的圖像數據破壞、或者在圖像中出現的複雜情況（例如：遮斷、重復的圖案）等原因，通常一組非常完美的匹配結果是很難獲得的。

Aloimonos and Herve (1990) 提出了一種不需要建立一一對應的特徵點，也可以恢復單一平面的三維的算法。到目前爲止，這種方法是唯一能夠繞開困難的匹配對應點問題的算法。可是，他們的算法把輸入的圖像局限在平行立體視中。

在本篇論文中 我們首先把平行立體視推廣到一般情況，提出新的恢復空間單一平面的不需要建立對應匹配點的算法。在這種算法中，不需要對相機對兒的相對位置做特殊的限制。

爲了對付特徵點群匹配的錯誤等影響因素，我們引入了奇異值分解和 robust estimation 的方法來提高算法的性能。

然而上述算法僅僅可以應用到遠處景物或者單一平面場景中。我們把這種算法進一步擴展，使它甚至可以應用到多個平面的場景恢復中。總體上，我們的系統可以在不需要建立一對應的特徵點的情況下，對於任意放置的的相機，從圖像中對多面體的場景進行三維結構恢復。

我們的不需要建立對應點的方法可以和傳統的立體視覺的方法聯合使用。當一特徵點——對應的關係非常難以確定的時候，例如，輸入的圖像是關於重復的紋理或單調的圖案的情況下，我們的算法就變得尤其有用了。

一些基於人造和真實的場景的實驗將說明我們的系統的性能。

ACKNOWLEDGEMENTS

I would like to sincerely thank my research supervisor, Professor CHUNG, Chi-kit Ronald, for his guidance, support, help and encouragement over the past two years.

I would also like to thank Professor LIU, Yun-hui, Professor WANG, Michael Yu and Professor QUAN, Long who are the thesis committee members, as well as other faculty members for their patient instruction throughout my academic study in the Chinese University of Hong Kong.

In addition, I would like express to my deep thanks to my colleague Mr. HE, Yong who gave me kindly help and good suggestions in my experiments. Moreover, I would like to thank Mr. WANG, Wei who lent a hand in 3D texture mapping even in his tough days.

Thanks to those colleagues and friends, CHUN Chun-nam, DAI Ruo-li and ZHAO, Yang, who share these happy days with me.

Finally I would like to thank my parents, my brother as well as my boyfriend for their love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

We live in a 3D world. The human vision system, in which two eyes located side-by-side in the front of their heads, is able to deal with 3D information. We use the term *stereo* in computer vision, as in human vision, to refer to the capability in inferring three-dimensional information of the structure and distance of a scene from two or more images taken from different viewpoints. This thesis aims at investigating how to achieve the three-dimensional information of the object from the two-dimensional image pairs, if without the process of correspondence establishment

In classical optics and computer graphics, the basic problem is to acquire the images of the three-dimensional objects. Thus, stereo vision in confronted with in the inverse problem of recovering the objects from images. During the projection from three-dimensional object to two-dimensional images, the depth information is lost, which causes the difficulties to the three-dimensional reconstruction of objects.

The established approach of stereo vision is first to establish correspondences across the input images, i.e., to estimate which feature points on the two images are projected by the same 3D point in space, then to determine 3D as the intersection of the corresponding lines of sight (through the corresponding image positions in the two images). The correspondence problem, which is a classical problem of computer vision, has been studied for decades. Now it is still an active research area, due to its practical difficulty.

So vision problem must often relay on some natural constraints. For instance, a number of constraints like the epipolar geometry and correspondence uniqueness, and heuristics like the disparity continuity and restricted disparity range, have been proposed to solve the correspondence problem. Unambiguous output could then be derived using these constraints that are some assumptions about the physical world. Yet, extensive research have shown that a perfect solution of the correspondence problem is generally difficult due to possible corruption of the image data by noise and occurrence of complications like occlusions and repetitive patterns in the image data.

In [1], Aloimonos and Herve presented a method of estimating 3D structure of a planar surface from stereo images without establishing point-to-point correspondences. The method was the only algorithm that could bypass the difficult stereo correspondence problem. However, the method is restricted to images taken under the parallel-axis stereo geometry in which the cameras have their optical axes parallel. Although some properties are of help to simplify mathematical models, parallel-axis stereo geometry has some limitations. For example, there are less overlaps in the two retinas under parallel-axis stereo system compared with general stereo system. We can also enlarge the overlap by decreasing the baseline, however, at the cost of less accuracy in depth computation. In the general stereo configuration, cameras do not necessarily have the image planes coplanar, nor do they have the optical axes parallel. In this thesis, we present an algorithm that is capable of handling image data taken from arbitrarily positioned cameras.

On the beginning, based on homography, we propose our new algorithm and generalized parallel-axis stereo geometry into general stereo geometry. Our method is

applicable to single plane surface recovery without one-to-one correspondence, however, for arbitrarily-positioned cameras. Our original algorithm can also be used under parallel-axis stereo geometry and has similar form as Aloimonos and Herve's. Not only does the algorithm, in a conceptual way, lump the two processes: (1) image rectification of the image data (to a form under the parallel-axis stereo geometry), and (2) application of the Aloimonos-Herve's algorithm (to the rectified data), into a single process, it offers a closed-form, one-shot solution to the stereo vision problem. As it is, the algorithm is applicable to both planar scenes and scenes that are pictured from a far distance (far compared with their depth range from the cameras).

The original algorithm mentioned above is capable of handling image data taken from arbitrarily positioned cameras, but it can only deal with a pure single planar surface or pure scene at infinity.

The real world is far more complicated than these cases, and there are usually some obstacles in the scene, besides the planar surface to be recovered, which is the more general case we hope to handle. Moreover, the feature points extracted in the two images are partly not corresponding due to the different contrast of the images and different viewpoints of the cameras. That is, we extract feature points in one image, but the correspondences of some of them in another image cannot be detected because of the changing lighting conditions and viewpoints during the image acquisition, and the deficiency of the feature detection algorithm. We indeed avoid the one-to-one correspondence establishment; nevertheless the imperfect features result in larger error.

For the sake of eliminating the effect caused by faulty features, we apply singular value decomposition (SVD) and robust estimation to get rid of wrong items form our

correspondence-free system. The improve algorithm is more robust and able to deal with the planar surface or the distant scene disturbed by some small obstacles.

Based on the improved algorithm, we go one step further and propose a novel stereo vision algorithm for recovering even scene consisting of multiple planar surfaces. Experimental results show that the algorithm is effective in recovering the 3D structure of polyhedral scene with neither the need of establishing point-to-point correspondence nor specific restriction on the stereo imaging geometry.

Our correspondence-free algorithm does not contradict the traditional correspondence methods. On the contrary, we hope that they can cooperate to create more efficient and reliable system to do 3D recovery in stereo vision.

The thesis is structured as follows. Chapter 1 (which is the current chapter) gives the motivation of our research work. In Chapter 2, some background related to stereo vision and traditional algorithms are briefly introduced in the first subsection, and then Aloimonos and Herve's correspondenceless algorithm for single planar surface recovery under parallel stereo geometry is reviewed. Chapter 3 is the core of the thesis, which gives our new correspondence-free algorithm handling the single planar surface for arbitrarily-positioned cameras. Then we introduce the use of singular value decomposition and robust estimation techniques to enhance the performance of our algorithm. The following in chapter 3 is about multiple planar surfaces reconstruction. Experimental results on testing our method are also shown. In last subsection of this chapter, comparison between correspondence and correspondence-free based algorithms is made in some experimental results. Finally chapter 4, the last part of the thesis, gives an overall conclusion of our research and recommends possible future work.

CHAPTER TWO

PREVIOUS WORK

The word "stereo" comes from the Greek word "stereos" which means firm or solid. With stereo vision we see an object as solid in three spatial dimensions--width, height and depth--or x, y and z. It is the added perception of the depth dimension that makes stereo vision so rich and special.

Stereo vision is an important technique, by means of which three-dimensional scene can be recovered from multiple images of the scene. In this thesis, my focus will be on the recovery of the scene from a pair of images, each image acquired from a different viewpoint.

For the stereo analysis, it is assumed that no object movements occur during the time interval of image acquisition of the two stereo images. Stereo analysis techniques aim to determine depth information or shape recovery from input image pairs.

2.1 Traditional Stereo Vision

In general, there are more or less the following steps covered by stereo vision [14]:

(1) Image acquisition: input image pairs acquired from different view points.

(2) Camera modeling: the determination of the intrinsic and extrinsic parameters of the cameras.

(3) Feature extraction: the calculation of significant image features like edges.

(4) Correspondence analysis: Determination of which feature point on the left image corresponds to which feature points on the right image.

(5) Triangulation: Starting with two corresponding feature points in the left and the right image, the corresponding depth can be calculated characterizing a certain point in the 3D scene space.

(6) Interpolation: The calculated points in 3D scene space have to be transformed into a certain representation of an object surface in 3D space.

Under the assumption of perspective projection, each image point is projected from the scene along the ray passing through the image point and its optical center. In fact, any point along this ray could probably be the scene point projected. However, if the corresponding image pints across the images are known, then the scene point must be at the intersection of the two rays. Therefore, once correspondences are determined, the three-dimensional positions of those feature points are easily computed from the pairs of feature points in both images, as illustrated in Figure 1. O and O' represent optical centers of the two cameras. The camera coordinates shown in the figure are (XYZ), ($X'Y'Z'$) respectively. Feature points p' and p on the left and right image are corresponding, that is, they are the projection of the same 3D scene point on the image pair. Once the correspondence (p and p') is determined, the 3D scene point P can be calculated via triangulation.

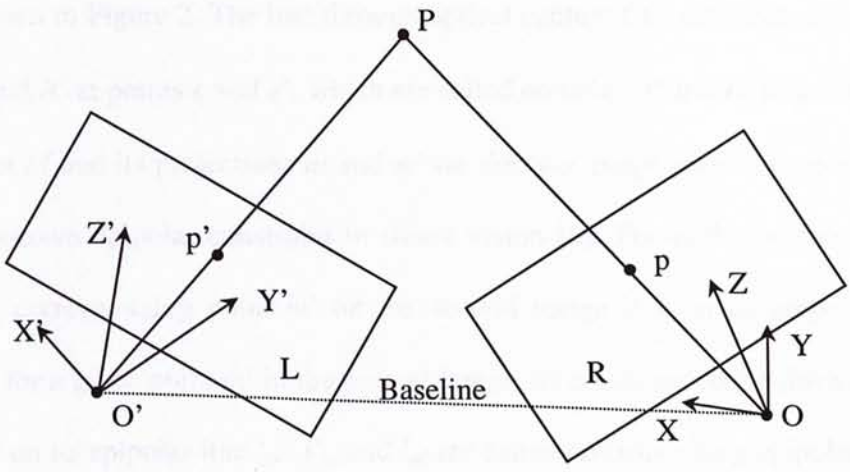

Figure 1: Stereo imaging system.

Among all the steps, correspondence establishment, determining which feature points on the input images are projected by the same 3D points in space, is a central and difficult problem of computer vision and has been studied for decades. A number of constraints like the epipolar geometry and correspondence uniqueness, and heuristics like the disparity continuity and restricted disparity range, have been proposed to solve the problem. Below, some classical constraints or assumptions and methods for correspondences establishment will be quoted.

2.1.1 Epipolar Constraint

The epipolar geometry is the intrinsic projective geometry between two views. It is independent of scene structure, and only depends on the cameras' internal parameters and relative pose [9].

Epipolar constraint was first described explicitly by Keating, Wolf, and Scarpace in 1975 [13]. Consider two images taken by two pinhole cameras with optical centers *C*

and C' shown in Figure 2. The line through optical centers CC' intersects the two image planes R and R' at points e and e', which are called epipoles. Plane Π, determined by the scene point M and its projections m and m' on the two image planes, is epipolar plane. The well-known epipolar constraint in stereo vision [6]: For each point m in the first image, its corresponding point m' on the second image lies on its epipolar line l'_m. Similarly, for a given point m' in the second image, its corresponding point m in the first image lies on its epipolar line $l_{m'}$. l'_m and $l_{m'}$ are called corresponding epipolar lines. The epipolar constraint can be expressed by a 3×3 matrix **F**, called fundamental matrix:

$$\mathbf{m'}^{\mathrm{T}}\mathbf{Fm} = 0 \text{ or } \mathbf{m}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{m'} = 0 \tag{2.1}$$

In another word, for a given point of projective coordinates **m** (or **m'**) in the first (or the second) image, its epipolar line $\mathbf{l'}_m$ (or $\mathbf{l}_{m'}$) in the second (or the first) image is given by:

$$\mathbf{l'}_m \cong \mathbf{F}\cdot\mathbf{m} \text{ (or } \mathbf{l}_{m'} \cong \mathbf{F}^{\mathrm{T}}\cdot\mathbf{m'}\text{)} \tag{2.2}$$

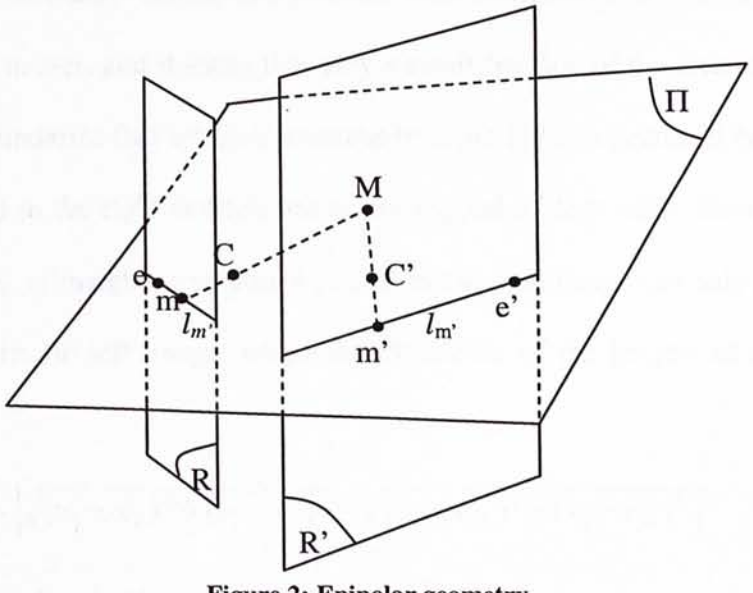

Figure 2: Epipolar geometry

Epipolar constraint reduces the search for a corresponding image point from a two-dimensional problem to a one-dimensional problem.

2.1.2 Some Constraints Based on Properties of Scene Objects

In order to reduce the ambiguities for the correspondence analysis processes, lots of constraints were proposed. Some of the constraints are so called geometric constraints since they are based on projective geometry, such as epipolar constraint, uniqueness constraint; while some other constraints are based on properties of scene objects. In the following subparagraph some typical ones of these constraints will be introduced briefly.

2.1.2.1 Continuity Constraint

The rule that continuity constraint cooperated with uniqueness constraint was first clearly proposed by Marr and Poggio in 1976. Continuity of disparities means that disparity varies smoothly almost everywhere. This condition is a consequence of the cohesiveness of matter, and it states that only a small fraction of the area of an image is composed of boundaries that are discontinuous in depth [17]. In details, if two points (u_1, v_1) and (u'_1, v'_1) in the right and left image correspond to each other, then we know an another point (u_2, v_2) neighboring points (u_1, v_1) in the right image can only match such a point (u'_2, v'_2) in the left image, where the difference of the lengths of two disparity vectors:

$$d_{diff} = \left| \sqrt{(u_1 - u'_1)^2 + (v_1 - v'_1)^2} - \sqrt{(u_2 - u'_2)^2 + (v_2 - v'_2)^2} \right| \tag{2.3}$$

is below a certain threshold.

2.1.2.2 Ordering of Projected Points in the Image Plane

Ordering constraint in another method to make the correspondence problem more unambiguous. It could be expressed as: image points that lie on one epipolar line in one stereo image are projected onto the corresponding epipolar line of the other image in exactly the same order [14].

Having been studied for many years, lots of research work is focused in this field. Certainly there are also some other constraints in solving the matching problem, such as compatibility of features, but these methods will not be listed in the thesis because of the limitation of the article. More details could be referred to some good publications [14], [23], [4], [20] etc.

2.1.3 Two Classes of Algorithms for Correspondence Establishment

Based on the epipolar constraint, searching for correspondences is reduced from a two-dimensional area to a one-dimensional line. The issues are how to determine and value the correspondences, which has been studied for decades, and still a hotspot in vision. Although many algorithms have been proposed in this field till now, however, basically there are two classes of algorithms for correspondences establishment: one is area-based method, the other one is feature-based method.

2.1.3.1 Area-Based Method

Aera-based method, such as [18] [16] [12], also called intensity-based method, is one of the oldest and still widely used method. This technique finds corresponding feature points on the basis of the similarity of the areas in the image pairs. In general, the

process consists of extracting feature in one image, and then searching its corresponding feature point on the conjugate epipolar line in another image. The feature extraction process is not necessary, and sometimes it can also be skipped. Aera-based method produces a dense set of correspondences.

For a given feature point in one image, its corresponding point in another image is sought on the basis of the similarity of the neighboring regions. So windows centered by these feature points should be selected at first. Assume that feature point (u_1, v_1) and (u'_1, v'_1) in the right and left images are a pair of corresponding candidate. Centered by the two feature points, we choose $m \times n$ windows named I_r (u_1, v_1) and I_l (u'_1, v'_1) for the points in their neighborhoods, respectively. Then whether the pseudo correspondence really exists relay on how similar the two chosen templates are. Many similarity functions have been proposed for this application. The most popular and basic two are the sum of squared differences (SSD) and the cross-correlation (CC). Suppose there are two templates $I_l(m,n)$ and $I_r(m,n)$ to be measured, then the similarity function could be the sum of squared differences (SSD) [19]:

$$SSD(\Delta m, \Delta n) = \sum_{i,j \in R} \sum [I_l(i, j) - I_r(i - \Delta m, j - \Delta n)]^2 \tag{2.4}$$

where $(\Delta m, \Delta n)$ is the disparity between the image locations of the two matched regions. The cross-correlation function can be defined as:

$$CC(\Delta m, \Delta n) = \sum_{i,j \in R} \sum I_l(i, j) I_r(i - \Delta m, j - \Delta n) \tag{2.5}$$

We notice the value of CC is dependent on the windows' size. Furthermore, (2.4) and (2.5) cannot overcome variant changes to image intensity such as those caused by changing lighting conditions across the image pairs. Hence, by normalizing the image

and feature vectors to unit length, the normalized cross-correlation (NCC) can conquer the difficulties:

$$\gamma(\Delta m, \Delta n) = \frac{\sum_{i,j}\left[I_l(i,j) - \bar{I}_l\right]\left[I_r(i-\Delta m, j-\Delta n) - \bar{I}_r\right]}{\left\{\sum_{i,j}\left[I_l(i,j) - \bar{I}_l\right]^2 \sum_{i,j}\left[I_r(i-\Delta m, j-\Delta n) - \bar{I}_r\right]^2\right\}^{0.5}} \tag{2.6}$$

where $\gamma(\Delta m, \Delta n)$ is named correlation coefficient, and $\bar{I}_l$, $\bar{I}_r$ are the mean of the templates. Area-based methods can handle dense maps, and are good for surface reconstruction. But these methods are generally sensitive to illumination variations and noises, and they are basically inadequate for the image pairs that are pictured at very different viewpoints.

Lots of research works aimed to improve the performance of the similarity measure. For example, Kanade and Okutomi proposed that the size of the matching window could be chosen adaptively on the basis of a local evaluation of the variation in both the intensity and the previously estimated disparity [12], since the size of matching window will greatly effect on the similarity measure. Veksler proposed an algorithm to choose a window size and shape by optimizing over a large class of "compact" windows. In his algorithm, the non-rectangular windows could be established efficiently [25].

2.1.3.2 Feature-Based Method

If a correspondence analysis technique for stereo images is not based on the comparison of the intensity values, but compares the image features' properties, then the technique is the feature-based method [7]. For instance, if the features were corners, the similarity would measure in terms of the corners' locations, and surrounding intensities used SSD or cross-correlation. If the features were some edge elements, lines or

contours, then the similarity measures would be the contrast, orientation, length, and the midpoints of the edges or lines.

Compared with the area-based method, feature-based methods make use of the sparse set of features selected from the dense image. Thus, feature-based methods are often preferred for its reduction of computing time. Moreover, feature-based methods do not entirely depend on image intensities, which is more robust to handle the photometric variations and noises during image acquisition. And accuracy of depth computation may be better than area-based method because the feature locations can be computed with subpixel accuracy.

Although studied for decades of years, correspondence establishment is still an active research area. Lots research work aims to make the matching problem more efficient and accurate and lots of publications have been published in this field. Besides the usual area-based or feature-based method, many algorithms bloom, for example, some of them turned to the third camera [26], and some algorithms combined area-based method with feature-based method to be more reliable than area-based method alone. Evaluations on the some of the present algorithms could be obtained in some papers about the comparison on the present matching algorithms, such as [15][22]. There are certainly some good matching software published, such as software *Image-matching* by Z. Y. Zhang [27].

2.2 Correspondenceless Stereo Vision Algorithm for Single Planar Surface Recovery under Parallel-axis Stereo Geometry

Aloimonos and Herve (1990) presented an algorithm [1] to recover planar surface under parallel-axis stereo geometry, as shown in Figure 3, without using any point-to point correspondence between the left and right images.

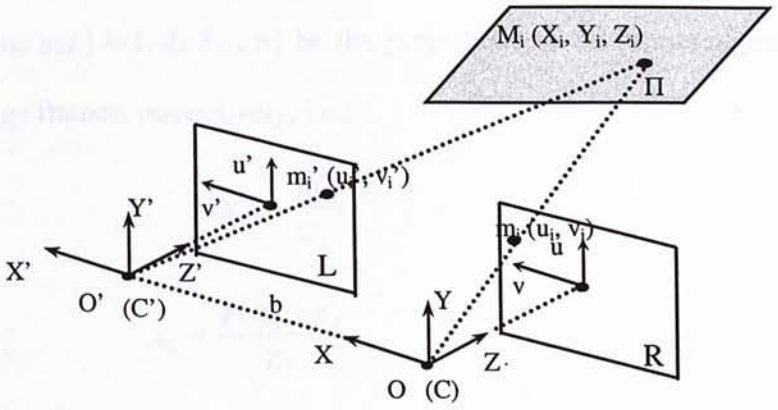

Figure 3: Parallel-axis stereo geometry

According to Aloimonos-Herve's work, the stereo system is set up as the following: the orthogonal world coordinate is consistent with the right camera coordinate (OXYZ); the optical centers of the two cameras, which have the same focal length f, are displaced by b (baseline) in the direction of X-axis; the left camera coordinate (O'X'Y'Z') is parallel to the right one with X-axis and X'-axis collinear; and each image plane is fixed at its focus perpendicular to the optical axis, as shown in Figure 3.

Consider a set of 3-D points $A=\{(X_i, Y_i, Z_i)| \ i=1, 2, 3...n\}$ lying on the same plane Π (Figure 3):

$$Z = pX+qY+c \tag{2.7}$$

Equality (2.7) can be rewritten as the following by dividing both sides of the equality by the term Zc:

$$\frac{1}{Z} = \frac{1}{c} - \frac{p}{c}\frac{X}{Z} - \frac{q}{c}\frac{Y}{Z} \tag{2.8}$$

Let I_r and I_l be the right and left image frames respectively. Finally let $\{(x_{ri}, y_{ri}) \mid i=1, 2, 3\ldots n\}$ and $\{(x_{li}, y_{li}) \mid i=1, 2, 3\ldots n\}$ be the projections of the points of set A on the right and left image frames, respectively, i.e.:

$$x_{ri} = \frac{fX_i}{Z_i} \tag{2.9}$$

$$x_{li} = \frac{f(X_i - b)}{Z_i} \tag{2.10}$$

$$\text{Then we have: } x_{ri} - x_{li} = \frac{fb}{Z_i} \tag{2.11}$$

$$y_{li} = y_{ri} = \frac{fY_i}{Z_i} \tag{2.12}$$

$i=1, 2, 3, \ldots, n$

where Z_i is the depth of the 3-D point having those projections.

$$\frac{1}{Z_i} = (f - px_{ri} - qy_{ri})\frac{1}{cf} \tag{2.13}$$

And now equality (2.11) and (2.13) can be jointed together by eliminating the term $1/Z_i$:

$$\frac{x_{ri} - x_{li}}{bf} = (f - px_{ri} - qy_{ri})\frac{1}{cf} \tag{2.14}$$

For any point in plane Π, equality (2.14) holds. Now there are n pairs of corresponding feature points across the two image frames, then certainly n equalities according to (2.14) are got. Apply Σ to get the summation of all the n equalities:

$$\sum_{i=1}^{n}\frac{x_{ri}}{bf}-\sum_{i=1}^{n}\frac{x_{li}}{bf}=\frac{1}{cf}\sum_{i=1}^{n}(f-px_{ri}-qy_{ri}) \tag{2.15}$$

The application of Σ can avoid point-to-point correspondence across images successfully. But three unknown parameters cannot be obtained uniquely from the only one linear equation (2.15). Then Aloimonos and Herve made full use of another important property of parallel-axis stereo geometry: $y_{ri} = y_{li}$, and this smart trick makes the problem of avoiding point-to-point correspondence feasible. They firstly multiplied both sides of equality (2.15) by the term of y_{ri}^{k}, where k is power:

$$\sum_{i=1}^{n}\frac{x_{ri}}{bf}y_{ri}^{k}-\sum_{i=1}^{n}\frac{x_{li}}{bf}y_{ri}^{k}=\frac{1}{cf}\sum_{i=1}^{n}(fy_{ri}^{k}-px_{ri}y_{ri}^{k}-qy_{ri}y_{ri}^{k}) \tag{2.16}$$

Then change term $\sum_{i=1}^{n}\frac{x_{li}}{bf}y_{ri}^{k}$ in (2.16) into $\sum_{i=1}^{n}\frac{x_{li}}{bf}y_{li}^{k}$, since $y_{ri}^{k} = y_{li}^{k}$ also holds, which avoids any point-to-point correspondence across the two images. Finally arrange equation (2.16) into a linear equation with respect to the three unknown parameters p, q, and c. If let power $k=k_1, k_2, k_3\,(k_1 \neq k_2 \neq k_3)$ respectively, we have three linear equations for three unknown parameters, and a linear system of equations in p, q, and c is obtained:

$$\begin{aligned}
&p\frac{1}{f}\sum_{i=1}^{n}x_{ri}y_{ri}^{k_1}+q\frac{1}{f}\sum_{i=1}^{n}y_{ri}^{k_1+1}+c\frac{1}{fb}\sum_{i=1}^{n}(x_{ri}y_{ri}^{k_1}-x_{li}y_{li}^{k_1})=\sum_{i=1}^{n}y_{ri}^{k_1}\\
&p\frac{1}{f}\sum_{i=1}^{n}x_{ri}y_{ri}^{k_2}+q\frac{1}{f}\sum_{i=1}^{n}y_{ri}^{k_2+1}+c\frac{1}{fb}\sum_{i=1}^{n}(x_{ri}y_{ri}^{k_2}-x_{li}y_{li}^{k_2})=\sum_{i=1}^{n}y_{ri}^{k_2}
\end{aligned} \tag{2.17}$$

$$p\frac{1}{f}\sum_{i=1}^{n} x_{ri} y_{ri}^{k_3} + q\frac{1}{f}\sum_{i=1}^{n} y_{ri}^{k_3+1} + c\frac{1}{fb}\sum_{i=1}^{n}(x_{ri} y_{ri}^{k_2} - x_{li} y_{li}^{k_3}) = \sum_{i=1}^{n} y_{ri}^{k_3}$$

The above method is feasible, yet Aloimonos and Herve proposed another method, which is combined with epipolar geometry. Assume we have two images taken by parallel-axis placed cameras, then the epipolar lines across the two images will be a group of parallel lines (Figure. 4).

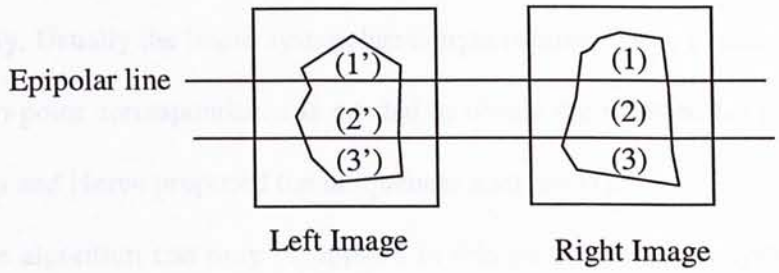

Figure 4: Epipolar lines under parallel-axis stereo geometry

These epipolar lines divide the images into different zones. For example, two epipolar lines will divide the images into three different zones: (1, 1'), (2, 2'), and (3, 3'). (Signs with prime and signs without prime are used to discriminate the zones in different images.) Under the previous condition, the group of n_1 feature points in (1) is corresponding to the group of n_1 feature points in (1'). Then apply equation (2.15) to the feature points in zone (1, 1'):

$$\sum_{i=1}^{n_2} x_{ri} y_{ri} + q\frac{1}{f}\sum_{i=1}^{n_2} y_{ri} + c\frac{1}{fb}\sum_{i=1}^{n_2}(x_{ri} y_{ri} - x_{li} y_{li}) = \sum_{i=1}^{n_2} y_{ri} \qquad (2.18)$$

Similarly, we obtain the other two equations with respect to zone (2, 2') and (3, 3'). Finally the just determined linear system of equations is constructed:

$$\sum_{i=1}^{n_2} x_{ri} y_{ri} + q\frac{1}{f}\sum_{i=1}^{n_2} y_{ri} + c\frac{1}{fb}\sum_{i=1}^{n_2}(x_{ri} y_{ri} - x_{li} y_{li}) = \sum_{i=1}^{n_2} y_{ri}$$

$$\sum_{i=1}^{n_2} x_{ri} y_{ri} + q\frac{1}{f}\sum_{i=1}^{n_2} y_{ri} + c\frac{1}{fb}\sum_{i=1}^{n_2}(x_{ri} y_{ri} - x_{li} y_{li}) = \sum_{i=1}^{n_2} y_{ri} \qquad (2.19)$$

$$\sum_{i=1}^{n_3} x_{ri} y_{ri} + q\frac{1}{f}\sum_{i=1}^{n_3} y_{ri} + c\frac{1}{fb}\sum_{i=1}^{n_3}(x_{ri} y_{ri} - x_{li} y_{li}) = \sum_{i=1}^{n_3} y_{ri}$$

where n_1, n_2, and n_3 represent the number of feature points in three different zones, respectively. Usually the linear system has unique solution for p, q, and c. We notice that no point-to-point correspondence is needed to obtain the solution for p, q, and c. Also Aloimonos and Herve proposed the uniqueness analysis [1].

But the algorithm can only be applied in this particular stereo system: parallel-axis stereo system. There is less overlap in the two retinas under parallel-axis stereo system than under general stereo system. Stereo vision is only effective for the 3D zone which is at the intersection of the two cameras' fields of view, we hope more overlaps to reconstruct the real 3-D world. We can also enlarge the overlap by decreasing the baseline, however, at the cost of less accuracy in depth computation. Thus, general stereo geometry should be applied to satisfy the two conditions at the same time, due to the two contradictive conditions under parallel-axis stereo geometry.

In the following chapter, a new correspondence-free stereo vision algorithm under the general stereo geometry will be presented.

CHAPTER THREE

CORRESPONDENCE-FREE STEREO VISION UNDER GENERAL STEREO SETUP

Correspondence establishment is about determining which pair of feature points in two input images are projected by the same 3D point in space. It is a key problem of stereo vision and has been studied for decades. A number of constraints like the epipolar geometry and correspondence uniqueness, and heuristics like the disparity continuity and restricted disparity range, have been proposed to solve the problem. Yet, extensive research (see surveys [4][11]) has shown that a perfect solution of the correspondence problem is generally difficult due to possible corruption of the image data by noise. Especially repetitive patterns in the scene cause ambiguity in matching. Most existing matching algorithms are still not robust enough to handle this case.

Aloimonos and Herve (1990) proposed a new algorithm [1] to recover planar surface under parallel-axis stereo geometry, which opened a new way for stereo vision. In their work, plane parameters can be achieved even without the process of correspondences establishment. Incited by their new viewpoint, we are working in the special direction to see how further we could probe.

This chapter is organized as follows. Section 1 gives our new method, including the algorithm in its original form and improved one. Then we apply our algorithm to some simple polyhedral objects, and test the method with experiments in section 2. Finally, in

section 3, we compare our correspondence-free based algorithm with the correspondence based method in some experiments.

3.1 Correspondence-free Stereo Vision Algorithm for Single Planar Surface Recovery under General Stereo Geometry

In the general stereo configuration, cameras do not necessarily have the image planes coplanar, nor do they have the optical axes parallel. Below we present a new algorithm that could handle such a case. The cameras need not have the same focal length, but we assume that they have been calibrated beforehand, i.e., the intrinsic and extrinsic parameters of the cameras have all been pre-estimated. On camera calibration there exist a number of algorithms in the literature to use (e.g., [10]).

Figure 5 shows a planar surface Π pictured by a stereo pair of cameras (Camera 1 and Camera 2) positioned in the general configuration. Let the rigid transformation between the cameras be described by rotation $\mathbf{R}$ and translation $\mathbf{t}$. Suppose a number of features of surface Π are visible as $\{(u_i,v_i)\}$ in one image (Image 1) and $\{(u_i',v_i')\}$ in the other image (Image 2), where $\{(u_i,v_i)\}$ and $\{(u_i',v_i')\}$ are image coordinates having been normalized by the focal lengths of the respective cameras.

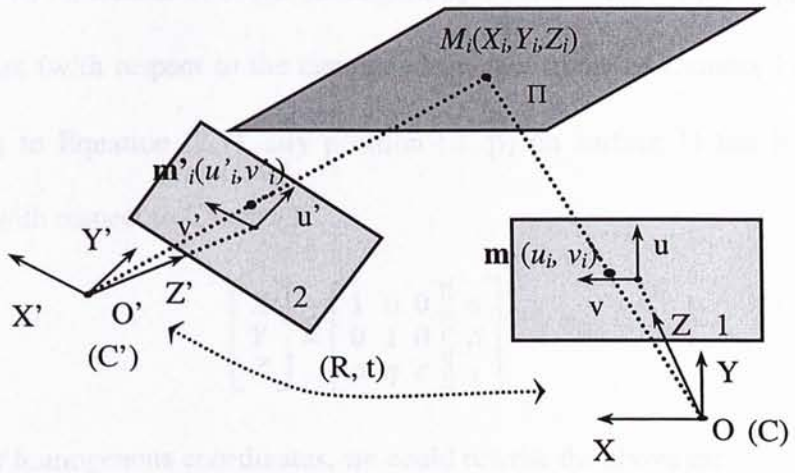

Figure 5: General stereo geometry

3.1.1 Algorithm in Its Basic Form

Suppose surface Π does not contain either of the two optical centers *C* and *C'* of the cameras, and let it be described as (2.7), for some parameters *p, q, c* with respect to the camera coordinate frame *X-Y-Z* of Camera 1. Our objective is to estimate *p, q, c* without the need of establishing precise correspondences across the stereo images. There is a family of planes -- planes that are orthogonal to image plane 1 (i.e., parallel to the Z-axis) -- that Equation (2.7) could not represent well. However, such planes will have their feature points visible almost all along a line (as opposed to over a spread-out region) in one of the stereo image pair, and are thus not quite recoverable from stereo vision anyway. Their presence is also easy to detect – some of the parameters *p, q, c* during the estimation process will have values of very large magnitude. In the following discussion we shall not consider the recovery of such planes.

Positions on surface Π could be indexed by 2 parameters (α, β). Suppose we use X,Y coordinates (with respect to the camera coordinate frame of Camera 1) for α, β. Then according to Equation (2.7), any position (α, β) on surface Π has the following 3D position with respect to Camera 1:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p & q & c \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{3.1}$$

Under homogenous coordinates, we could rewrite the above as:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p & q & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{3.2}$$

Note that the parameters (α, β) above are about position on surface Π not about the image of surface Π, so they are image-invariant.

Then for a set of features points $\{(X_i,Y_i,Z_i,1)^{\mathrm{T}} | i = 1,2,3,...,n\}$ on surface Π, their 3D positions $\mathbf{M}_i = (X_i \;\; Y_i \;\; Z_i \;\; 1)^{\mathrm{T}}$ are related to their position indices (α_i, β_i) on surface Π, by:

$$\mathbf{M}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p & q & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \\ 1 \end{bmatrix} \tag{3.3}$$

Camera maps the 3-D projective space $\boldsymbol{P}^3$ onto the 2-D projective plane $\boldsymbol{P}^2$ by perspective projection. The correspondence between point $\mathbf{M}_i$ and its normalized image $\mathbf{m}_i = (u_i \;\; v_i \;\; 1)^{\mathrm{T}}$ in image 1 can be expressed as:

$$\mathbf{m}_i \cong \mathbf{P}_E^r \mathbf{M}_i \tag{3.4}$$

where $\cong$ denotes equality up to a scale, $\mathbf{P}_E^r = \left[\begin{array}{c|c} \mathbf{I} & \begin{matrix}0\\0\\0\end{matrix} \end{array}\right]$, and $\mathbf{I}$ is 3×3 identity matrix.

Similarly, if $\mathbf{m}'_i = (u'_i\ v'_i\ 1)^{\mathrm{T}}$ is the normalized image coordinates in Image 2 of the ith feature point of surface Π, $\mathbf{m}'_i$ is related to $\mathbf{M}_i$ by

$$\mathbf{m}'_i \cong \mathbf{P}_E^l \mathbf{M}_i \tag{3.5}$$

where $\mathbf{P}_E^l = [\mathbf{R}|\mathbf{t}]$ which is related to the rigid transformation between the two cameras.

Then we have two matching sets of image-points: $\{\mathbf{m}_i | i = 1,2,3,...,n\}$ and $\{\mathbf{m}'_i | i = 1,2,3,...,n\}$, that are projected by the same set of feature points to Images 1 and 2 respectively. By combining Equations (3.3), (3.4), and (3.5), eliminating the vector $[\alpha_i\ \beta_i\ 1]^{\mathrm{T}}$, multiplying a scalar w_i to change an equality-up-to-scalar into an equality, and some simple algebraic manipulations, we have:

$$\begin{bmatrix} u'_i \\ t_x v'_i - t_y u'_i \\ t_x - t_z u'_i \end{bmatrix} = \begin{bmatrix} R_{11} + t_x(-\frac{p}{c}) & R_{12} + t_x(-\frac{q}{c}) & R_{13} + t_x(\frac{1}{c}) \\ t_x R_{21} - t_y R_{11} & t_x R_{22} - t_y R_{12} & t_x R_{23} - t_y R_{13} \\ t_x R_{31} - t_z R_{11} & t_x R_{32} - t_z R_{12} & t_x R_{33} - t_z R_{13} \end{bmatrix} \begin{bmatrix} w_i u_i \\ w_i v_i \\ w_i \end{bmatrix} \tag{3.6}$$

where $\mathbf{R} = [R_{ij}]$ and $\mathbf{t} = [t_x\ t_y\ t_z]^{\mathrm{T}}$.

Equation (3.6) corresponds to two equalities:

$$\begin{cases} U'_i = R_{11}U_i + t_x U_i(-\frac{p}{c}) + R_{12}V_i + t_x V_i(-\frac{q}{c}) + R_{13}F_i + t_x F_i(\frac{1}{c}) \\ d'_i = d_i \end{cases} \tag{3.7}$$

where $U'_i = \dfrac{u'_i}{t_x - t_z u'_i}$, $\qquad d'_i = \dfrac{t_x v'_i - t_y u'_i}{t_x - t_z u'_i}$

$$U_i = \frac{u_i}{(t_x R_{31} - t_z R_{11})u_i + (t_x R_{32} - t_z R_{12})v_i + (t_x R_{33} - t_z R_{13})},$$

$$V_i = \frac{v_i}{(t_x R_{31} - t_z R_{11})u_i + (t_x R_{32} - t_z R_{12})v_i + (t_x R_{33} - t_z R_{13})},$$

$$F_i = \frac{1}{(t_x R_{31} - t_z R_{11})u_i + (t_x R_{32} - t_z R_{12})v_i + (t_x R_{33} - t_z R_{13})},$$

$$d_i = \frac{(t_x R_{21} - t_y R_{11})u_i + (t_x R_{22} - t_y R_{12})v_i + (t_x R_{23} - t_y R_{13})}{(t_x R_{31} - t_z R_{11})u_i + (t_x R_{32} - t_z R_{12})v_i + (t_x R_{33} - t_z R_{13})}.$$

By powering both sides of the second equation of (3.7) with the same k, we have the equality $d'_i{}^k = d_i^k$. By multiplying the left side and right side of the first equation of (3.7) by d'^k_i and d_i^k respectively, we obtain

$$U'_i d_i'^k = [R_{11}U_i + t_x U_i(-\frac{p}{c}) + R_{12}V_i + t_x V_i(-\frac{q}{c}) + R_{13}F_i + t_x F_i(\frac{1}{c})]d_i^k \qquad (3.8)$$

which holds for the image projections $\mathbf{m}_i, \mathbf{m}'_i$ ($i=1,2,3...n$) of any ith feature point of surface Π.

Applying Σ to the above to get the summation of all n equations, we obtain:

$$\sum_{i=1}^{n} U'_i d'^k_i = \sum_{i=1}^{n} [R_{11}U_i + t_x U_i(-\frac{p}{c}) + R_{12}V_i + t_x V_i(-\frac{q}{c}) + R_{13}F_i + t_x F_i(\frac{1}{c})]\, d_i^k \qquad (3.9)$$

Note that the left and the right side of Equation (3.9) are both computable even without point-to-point correspondence established across the images. Let the power k assume the values of k_1, k_2, k_3 ($k_1 \neq k_2 \neq k_3$) respectively. Then a linear system of equations for the plane parameters p, q, and c is available as:

$$p\left[\sum_{i=1}^{n} t_x U_i d_i^{k_1}\right] + q\left[\sum_{i=1}^{n} t_x V_i d_i^{k_1}\right] + c\left[\left[\sum_{i=1}^{n} U'_i d'^{k_1}_i\right] - \left[\sum_{i=1}^{n} R_{11}U_i d_i^{k_1}\right] - \left[\sum_{i=1}^{n} R_{12}V_i d_i^{k_1}\right] - \left[\sum_{i=1}^{n} R_{13}F_i d_i^{k_1}\right]\right]$$

$$= \sum_{i=1}^{n} t_x F_i d_i^{k_1}$$

$$p\left[\sum_{i=1}^{n} t_x U_i d_i^{k_2}\right] + q\left[\sum_{i=1}^{n} t_x V_i d_i^{k_2}\right] + c\left[\left[\sum_{i=1}^{n} U'_i d'^{k_2}_i\right] - \left[\sum_{i=1}^{n} R_{11} U_i d_i^{k_2}\right] - \left[\sum_{i=1}^{n} R_{12} V_i d_i^{k_2}\right] - \left[\sum_{i=1}^{n} R_{13} F_i d_i^{k_2}\right]\right]$$

$$= \sum_{i=1}^{n} t_x F_i d_i^{k_2}$$

$$p\left[\sum_{i=1}^{n} t_x U_i d_i^{k_3}\right] + q\left[\sum_{i=1}^{n} t_x V_i d_i^{k_3}\right] + c\left[\left[\sum_{i=1}^{n} U'_i d'^{k_3}_i\right] - \left[\sum_{i=1}^{n} R_{11} U_i d_i^{k_3}\right] - \left[\sum_{i=1}^{n} R_{12} V_i d_i^{k_3}\right] - \left[\sum_{i=1}^{n} R_{13} F_i d_i^{k_3}\right]\right]$$

$$= \sum_{i=1}^{n} t_x F_i d_i^{k_3} \tag{3.10}$$

where $k=k_1, k_2, k_3$ ($k_1 \neq k_2 \neq k_3$)

Generally, this linear system of equations has unique solution for p, q, and c, which is about the 3D structure of the surface Π. In the case of parallel-axis stereo geometry, the rotation matrix $\mathbf{R}=\mathbf{I}$, and the translation vector $\mathbf{t}$ is only $[b\ 0\ 0]^T$ for some baseline b. . Under these conditions, Equation (3.10) will be reduced to the equation proposed in [1].

3.1.2 Algorithm Combined with Epipolar Constraint

From the previous section, it could be seen that a just-determined linear system of equations for the plane parameters p, q, c could be constructed by applying $d'^{k}_i = d_i^{k}$ ($k=k_1, k_2, k_3, k_1 \neq k_2 \neq k_3$) to Equation (3.7). However, experiments show if d_i or d'_i (which are analogous to the vertical image coordinates of the parallel-axis stereo configuration) has very large or small value, the solution is not stable. We use an approach that exploits the epipolar constraint instead. The approach, which is detailed below, allows more stable solution of p, q, c be attainable.

3.1.2.1 Algorithm Combined with Epipolar Constraint

The epipolar geometry of two cameras is captured by a 3×3 matrix **F** that is termed the Fundamental matrix. Fundamental matrix **F** is computable from knowledge of the extrinsic parameters (**R**,**t**) and intrinsic parameters (focal lengths, pixel dimensions etc.) of the two cameras (it is also directly computable from a few point correspondences over the images). The null spaces of **F** and $\mathbf{F}^T$ are two specific image points in the stereo images, namely epipoles **e** and **e'**, both expressed in terms of homogeneous coordinates. As depicted in Figure 6, the epipoles are in fact the projections of the optical centers of the respective cameras to the other image plane.

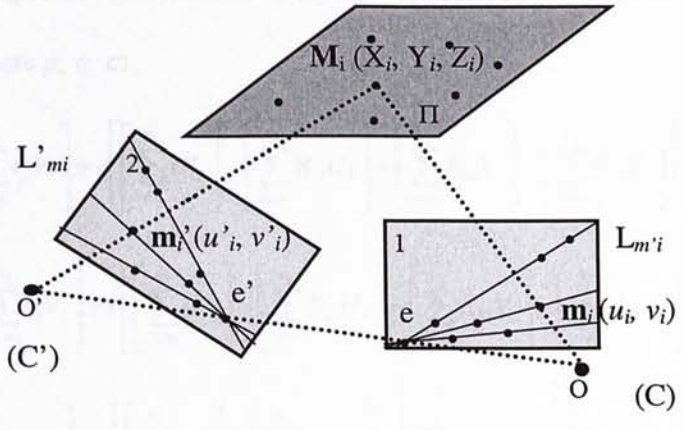

Figure 6: Epipolar constraint of stereo images.

The essence of **F** is, it allows the correspondence of any image point $\mathbf{m}_i$ or $\mathbf{m'}_i$ in one image to be located along the locus of a line in the other image. Given $\mathbf{m}_i$ in Image 1 or $\mathbf{m'}_i$ in Image 2, with knowledge of **F**, the corresponding image point in the other image must be on a line that could be determined from **F**. In fact, epipolar lines appear in conjugate pairs $\{(\mathbf{L}_i, \mathbf{L}'_i)\}$ over the stereo images, in the sense that for all points in line

$\mathbf{L}_i$ of Image 1, the corresponding image point must be in line $\mathbf{L}'_i$ of Image 2, and vice versa. Given any image point $\mathbf{m}_i$ (in homogeneous coordinates) of Image 1, the conjugate pair of epipolar lines that contain it are, in projective line format, $\mathbf{L}'_{m_i} = \mathbf{F} \cdot \mathbf{m}_i$ of Image 2 and $\mathbf{L}_{m_i} = \mathbf{m} \times \mathbf{e}$ of Image 1.

We could then group image points in the stereo images in accordance with whether they lie on the same conjugate pair of epipolar lines or not. Image points in the two images could thus be divided into a number of corresponding groups, with each group corresponding to image features that belong to same conjugate pair of epipolar lines and that have the same number in the two images. With three corresponding groups, say, and applying $k=0$ to Equation (3.9), we have a just-determined linear system of equations for the plane parameters p, q, c:

$$p\left[\sum_{i=m_1}^{n_1} t_x U_i\right] + q\left[\sum_{i=m_1}^{n_1} t_x V_i\right] + c\left[\left[\sum_{i=m_1}^{n_1} U'_i\right] - \left[\sum_{i=m_1}^{n_1} R_{11} U_i\right] - \left[\sum_{i=m_1}^{n_1} R_{12} V_i\right] - \left[\sum_{i=m_1}^{n_1} R_{13} F_i\right]\right] = \sum_{i=m_1}^{n_1} t_x F_i$$

$$p\left[\sum_{i=m_2}^{n_2} t_x U_i\right] + q\left[\sum_{i=m_2}^{n_2} t_x V_i\right] + c\left[\left[\sum_{i=m_2}^{n_2} U'_i\right] - \left[\sum_{i=m_2}^{n_2} R_{11} U_i\right] - \left[\sum_{i=m_2}^{n_2} R_{12} V_i\right] - \left[\sum_{i=m_2}^{n_2} R_{13} F_i\right]\right] = \sum_{i=m_2}^{n_2} t_x F_i \quad (3.11)$$

$$p\left[\sum_{i=m_3}^{n_3} t_x U_i\right] + q\left[\sum_{i=m_3}^{n_3} t_x V_i\right] + c\left[\left[\sum_{i=m_3}^{n_3} U'_i\right] - \left[\sum_{i=m_3}^{n_3} R_{11} U_i\right] - \left[\sum_{i=m_3}^{n_3} R_{12} V_i\right] - \left[\sum_{i=m_3}^{n_3} R_{13} F_i\right]\right] = \sum_{i=m_3}^{n_3} t_x F_i$$

where $(m_1 \sim n_1)$, $(m_2 \sim n_2)$, $(m_3 \sim n_3)$ represent three different corresponding groups.

In general, we have more than 3 conjugate pairs of epipolar lines available to us. We then have an over-determined system of equations to work with, of which we could adopt the least-squares error solution for the plane parameters.

3.1.2.2 Experimental Results

We have implemented the proposed algorithm and tested it with synthetic and real image data to investigate its performance. Below we show two sets of experimental results, both on real image data. In all the results shown, we used the algorithm described in [10] to estimate the intrinsic parameters and the extrinsic parameters (**R**,**t**) of the cameras, and used those parameters to calculate the Fundamental matrix **F** of the stereo setup. On the epipolar constraint, since there is inevitable error over the constraint, we give the constraint a tolerance of about 3 pixels, i.e., a feature point is considered to be on a particular epipolar line if its perpendicular distance to that line is no more than 3 pixels.

The first experiment was on planar surface and was aimed at investigating the 3D reconstruction accuracy of the described algorithm. The scene is about a poster hung on wall. We were to examine how accurate could we estimate the plane that contains the poster and the wall.

The two images of the scene are shown in Figure 7 and Figure 8 respectively. The baseline between the two viewpoints was about 0.5m wide. The images are both of resolution 1600×1200.

We extracted the top most textured feature points (i.e., image positions whose intensity sub-windows have the highest variance) from each image and grouped them according to the respective conjugate pair of epipolar lines they lie upon. There were a few feature points that have unique correspondence just based upon the epipolar constraint alone, but we threw such feature points away so as to construct a truly correspondence-free condition for the described algorithm. Among all the remaining

feature points we extracted from the images, 84 of them over each image were of the same number (at least two) on the corresponding conjugate pair of epipolar lines. The feature points spanned 17 conjugate pairs of epipolar lines, and they were accordingly divided into 17 matching groups. The least-squares-error solution for the plane parameters p, q, and c, of the system of equations expressed in Equation (3.11) is displayed in Table 1.

p	q	c
0.2	*0.3*	*1097.3*

Table 1: 3D reconstruction result of the correspondence-free method for the poster scene

To examine how accurate the plane parameters were estimated, we hand-picked 153 point-to-point correspondences across the images, reconstructed their 3D positions using triangulation, fitted a plane in 3D that best fitted such 3D positions under the least-squares-error criterion, and compared the parameters of such a plane with the results obtained using the correspondence-free algorithm. The plane parameters estimated using the correspondence-requiring method are displayed in Table. 2.

p	q	c
0.2	*0.3*	*1096.8*

Table 2: 3Dreconstruction result of correspondence requiring method for the poster scene.

It could be observed from Tables 1 and 2 that there is negligible difference between the reconstruction results of the two methods, which well illustrates the 3D reconstruction accuracy of the correspondence-free algorithm. This is particularly encouraging given that the two reconstructions are compared not on the same basis: one with point-to-point correspondences, and the other not.

To evaluate the reconstruction accuracy of the correspondence-free algorithm, we further chose 51 feature points from one image (Image 1) (Figure 7), recovered their 3-D information according to the plane parameters (p, q, and c) estimated from the correspondence-free algorithm, and acquired their reprojected positions in the other image (Image 2). Comparing the reprojected positions (shown as blue "crosses" in Figure 8) with the manually picked (true) positions (shown as red "crosses"), a distribution of the reprojection error (measured in terms of the Euclidean distance between the reprojected position and the true position) could be attained, as displayed in Figure 9. It was observed that approximately 96.08% of the reprojections have reprojection error less than 2 pixels. It could thus be concluded that the correspondence-free algorithm was indeed capable of reconstructing 3D accurately, and it could even be used for establishing point-to-point correspondences should correspondences be required.

The reason why we often used the reprojection error to evaluate the 3D reconstruction in this thesis is that the real position of a recovered 3D point with respect to one of the cameras' coordinates is quite difficult to be obtained by using some measuring methods except some methods of vision. We firstly choose some features in one image, and then calculate their 3D positions via the plane parameters estimated with

respect to one of the cameras' coordinates. After that, the projections of these recovered 3D points on another image frame with respect to another camera coordinates are obtained by perspective projection. We name the projections of these recovered 3D points on another image frame as reprojections. If the plane parameters are precisely estimated, the reprojections in the second image should be the correspondences of the features chosen in the first image. So the smaller reprojection error, the more precise the plane parameters are, if the error in choosing correspondence manually was ignored.

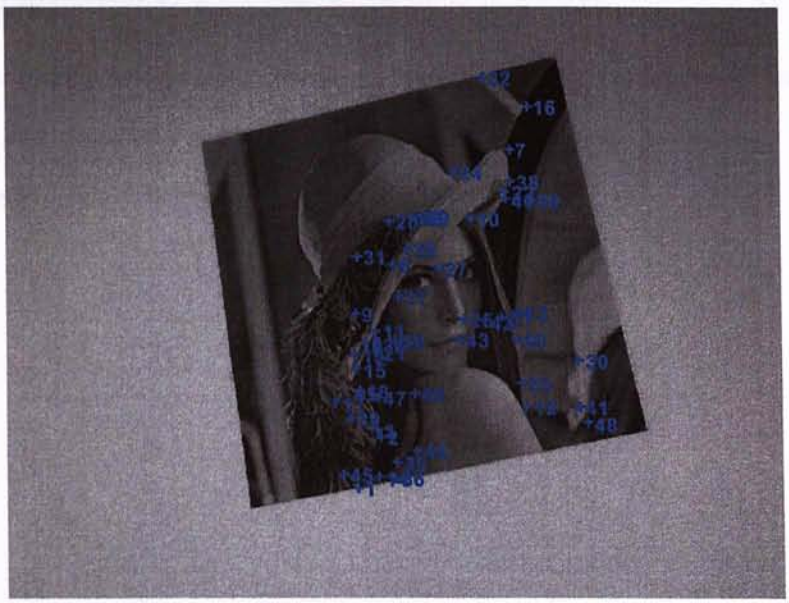

Figure 7: Image 1 of the poster scene, with 51 feature points marked as blue "crosses".

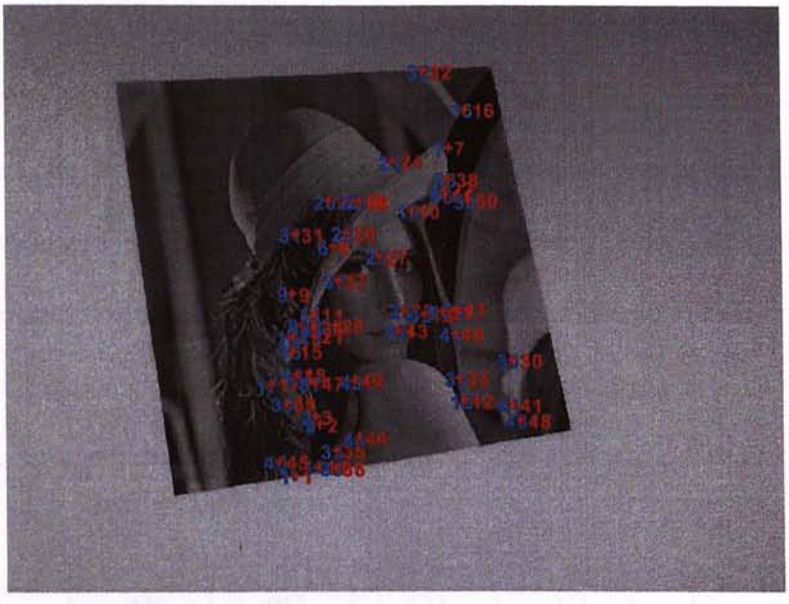

Figure 8: Image 2 of the poster scene, with the manually picked (true) correspondences of the 51 feature points marked as red "crosses", and the projected positions (from Image 1) of the correspondence-free algorithm marked as blue "crosses".

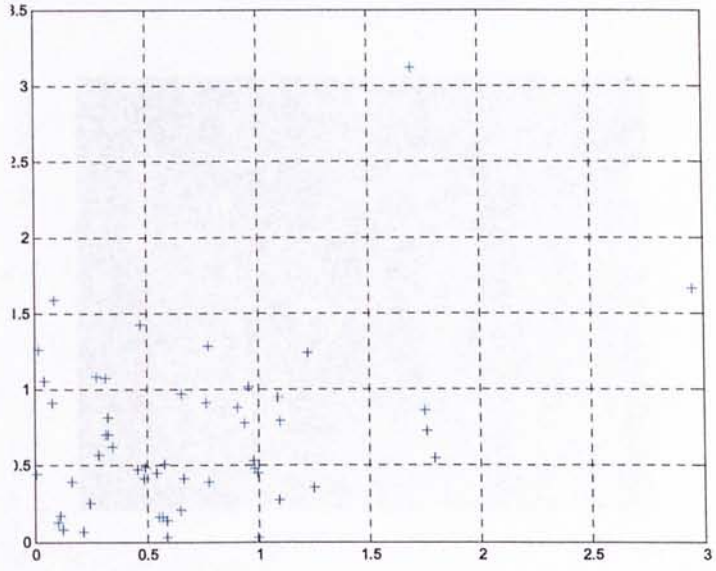

Figure 9: Reprojection errors (in the image domain, in pixel unit) of the correspondence-free algorithm for 51 feature points of the poster scene.

Another method to evaluate the accuracy of experimental result will be presented in the following part, and this method is more direct compared with the previous one. Four feature points are chosen in image 1 (Figure 10). The distance l_r between any feature points can be estimated by the recovered 3-D planar surface, according to the plane parameters (p, q, and c) estimated by the algorithm. However, the images is a poster on the wall, and the distance l_t between the same feature points in the real world can also be measured by us. Error will be defined as the absolute value of l_r and l_t. The error is show in Table 3.

	AB	*BC*	*CD*	*AC*	*BD*	*DA*
Truth	*413.2*	*415.0*	*413.5*	*587.0*	*585.5*	*416.7*
From (p, q, c)	*415.0*	*418.4*	*414.4*	*591.1*	*587.6*	*419.0*
Error	*1.8*	*3.4*	*0.9*	*2.1*	*2.1*	*2.3*

Table 3: Error (Unit: mm) : average relative error: 0.46%

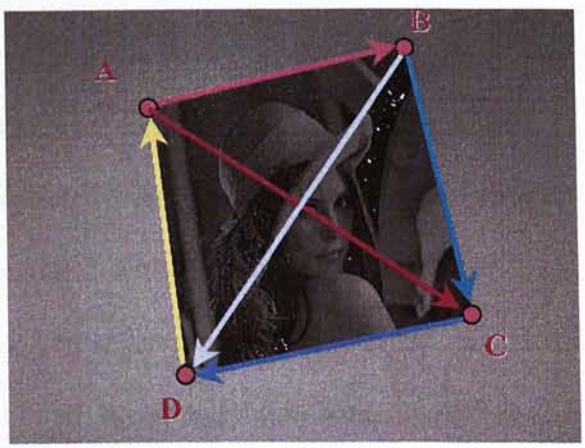

Figure 10: Four feature points are extracted, and the distance between the points can be estimate by the parameters *p*, *q*, and *c*

The proposed algorithm could also handle scenes that are pictured from significant distance (in comparison with their distance range) from the cameras, for they could be treated as planar scenes. In the following we show one set of experimental results on such scenes. The stereo images are displayed in Figure 11 and 12. The images were taken with a baseline of about 0.7m wide, and are of resolution 1600×1200. Using the same steps as those in the previous experiment, we obtained 56 feature points over 14 conjugate pairs of epipolar lines (each epipolar line containing at least two feature points). The plane parameters (p, q, and c) of the best plane that describes the scene was estimated using the correspondence-free algorithm, and the result is displayed in Table 4:

p	q	c
-0.0043	-0.2983	1.1537×10^5

Table 4: 3D reconstruction result of the correspondence-free method for the distant scene.

We also evaluated the accuracy of the result using the reprojection method mentioned in the previous subsection. 49 feature points (Figure. 11) were chosen in one image, and their reprojected positions (shown as blue "crosses" in Figure. 12) to the other image were compared with the manually picked (true) positions (shown as red "crosses" in Figure. 12). The reprojection errors are displayed in Figure. 13. Approximately 85.71% of the reprojection error is within 6 pixels. Also, it was observed that the reprojection errors are smaller at places where the scene is more distant from the cameras, which agrees with the expectation.

Figure 11: Image 1 of the distant scene, with 49 feature points marked as blue "crosses".

Figure 12: Image 2 of the distant scene, with the manually picked (true) correspondences of the 49 feature points marked as red "crosses", and the reprojected positions (from Image 1) of the correspondence-free algorithm marked as blue "crosses".

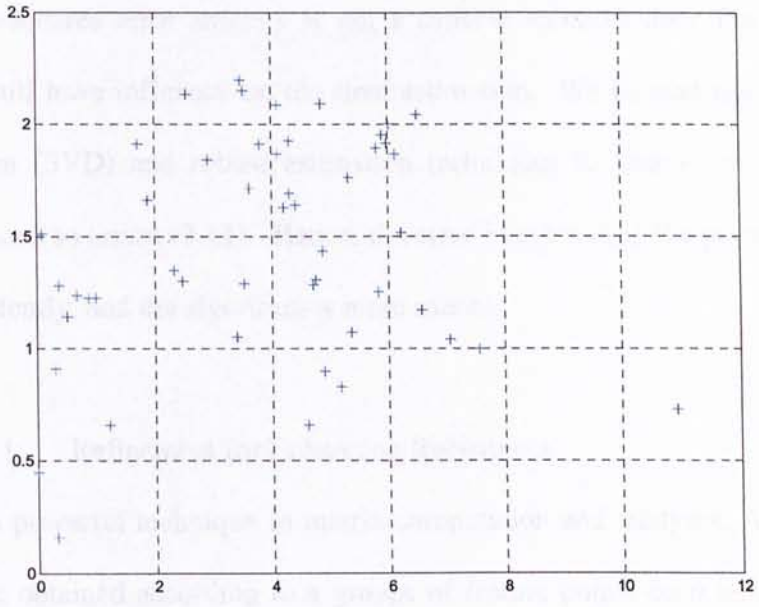

Figure 13: Reprojection errors (in the image domain, in pixel unit) of the correspondence-free algorithm for 49 feature points of the distant scene.

3.1.3 Algorithm Combined with SVD And Robust Estimation

In practice, however, some of the feature group pairs we use over the images could be faulty due to the different contrast and different fields of view of the images. Some of the pairs may contain a few feature points that are not really corresponding. Although in our previous algorithm (3.11) we refined feature by restricting the feature points to be those that are on one of conjugate epipolar lines, and are of the exactly same number. However, after this feature points refining, there are yet some wrong groups. That is, the numbers of feature points on a conjugate pair of epipolar lines across the two images are same, but these feature points are not exactly corresponding totally. Furthermore, the real world is far more complicated than a single planar surface, and there are usually some obstacles in the scene, which is the more general case we hope to handle.

The least-squares error strategy is not a suitable solution since the faulty feature group pairs still have influence on the final estimation. We instead use singular value decomposition (SVD) and robust estimation techniques to remove the influence the wrong equations so arisen (3.11). Hence, the error in estimating the plane parameters is reduced efficiently, and the algorithm is more robust.

3.1.3.1 Refinement for Enhancing Robustness

SVD is a powerful technique in matrix computation and analyses. Assume n linear equations are obtained according to n groups of feature points on n conjugate epipolar lines, then we rewrite (3.11) as:

$$\mathbf{AX} = \mathbf{0} \tag{3.12}$$

where $\mathbf{A}$ is $n\times4$ matrix of coefficients of the linear system of equations (3.11) , and $\mathbf{X}=[p\ q\ c\ 1]^{\mathrm{T}}$. Some rows of $\mathbf{A}$ are wrong, due to some wrong groups with feature points that are not exactly corresponding totally. Thus, each column $\mathbf{C}^i$ of $\mathbf{A}$ consists of a signal component $\mathbf{C}_s$ and a noise component $\mathbf{C}_n$:

$$\mathbf{A} = (\mathbf{C}^1 \quad \mathbf{C}^2 \cdots \mathbf{C}^N) \tag{3.13}$$

where each $\mathbf{C}^i = \mathbf{C}_s^{\,i} + \mathbf{C}_n^{\,i}$. The vector $\mathbf{C}_s$ representing the signal is known to lie in a rank k subspace. The work of decomposing $\mathbf{A}$ into signal component and noise component is reduced to obtain the optimal k rank of approximation $\mathbf{A}_k$ to $\mathbf{A}$.

Via singular value decomposition:

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}} \tag{3.14}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{\Sigma}$ is a diagonal matrix, in which the nonnegative singular values s_{ij} appear along its diagonal in descending numerical order.

Using the SVD, the original matrix $\mathbf{A}$ can be decomposed into the orthogonal components $\mathbf{A}_s = \mathbf{U}\,\mathbf{\Sigma}_k\,\mathbf{V}^{\mathrm{T}}$, which is the rank k subspace corresponding to the signal subspace [21]. In our linear system, k=3. Therefore planar parameters $\mathbf{X}=[p\ q\ c\ 1]^{\mathrm{T}}$ will be obtained by solving the new linear system:

$$(\mathbf{U}\,\mathbf{\Sigma}_k\,\mathbf{V}^{\mathrm{T}})\,\mathbf{X} = \mathbf{0} \tag{3.15}$$

Assume $\hat{\mathbf{X}}$ is the least-squares error solution for (3.15). However, the wrong equations of the system still have an effect on the solution. Therefore, robust estimation is adopted to eliminate the outliers [2]. In practice, we erase only one row vector with maximum residual in $\mathbf{A}$ to create a new matrix of coefficients $\tilde{\mathbf{A}}$. We then have a new linear system of equations:

$$\tilde{\mathbf{A}}\,\mathbf{X} = \mathbf{0}. \tag{3.16}$$

Similarly, we apply SVD to obtain the optimal k rank of approximation $\tilde{\mathbf{A}}_k$ to $\tilde{\mathbf{A}}$, and then construct a new linear system to substitute (3.12). We do the whole process repeatedly, until the maximum residual is less than a predetermined threshold during the robust estimation. The improved algorithm combined with SVD and robust estimation is not only more applicable to single planar surface recovery and robust enough to avoid disturbance of small obstacles in the scene, but also of great help in multiple planar surfaces reconstruction.

3.1.3.2 Experimental Results

The experiment is about a planar surface disturbed by some small obstacles. We aimed at inspecting the robustness of the improved algorithm combined with SVD and robust estimation.

The two input images are shown in Figure.14 and Figure.15 respectively. In the images we notice there are some small boxes and bottles in front of the poster that is to be recovered. The baseline between the two viewpoints is about 0.4m wide. The resolution of the both images is1280×960. 200 features were extracted using Harris' corner detector [8] in the two images respectively first, and then from each image we grouped them according to the respective conjugate pair of epipolar lines they lie on. Only 41 of them were of the same number (more than one) on 16 corresponding conjugate pairs of epipolar lines. We dropped the one-to-one correspondence after the feature points grouping, for the sake of testing our correspondence-free algorithm. The result of feature points grouping according to conjugates of epipolar lines are shown in Figure.14 and Figure.15. The crosses locate the features' positions, and the number represents different conjugate of epipolar lines. In fact, after examining the groups of features, we found 9 of the total 16 groups of features are obviously not totally corresponding with our eyes, even including group 9 and 14, which are on the obstacles, instead of the planar surface. We applied the original correspondence-free algorithm (only with least-squares error strategy) without SVD and robust estimation, and obtained the planar parameter in Table 5:

p	q	c
0.4973	0.7323	928.2366

Table 5: 3D reconstruction result of the correspondence-free method with only least-squares error strategy

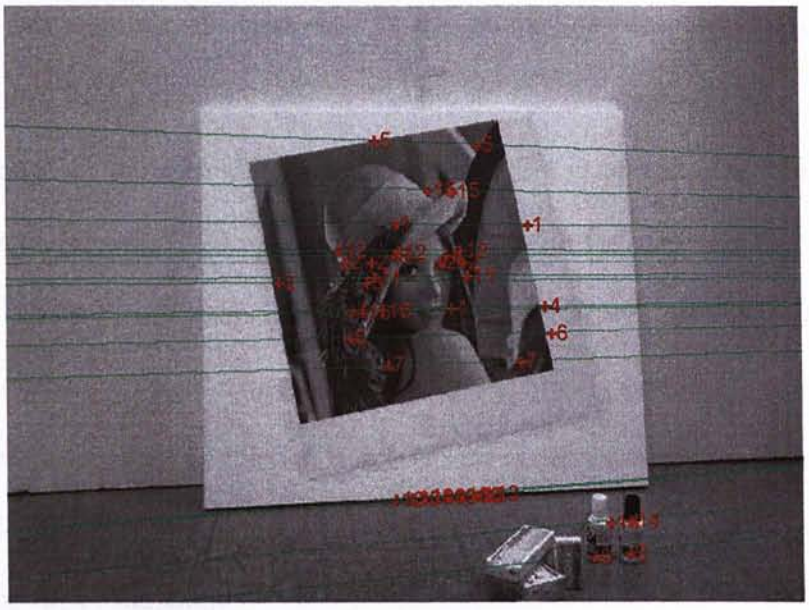

Figure 14: Image 2 of the scene, with 41 feature points marked as red "crosses" on 16 conjugates of epipolar lines. 9 of the total 16 groups of features are not totally corresponding.

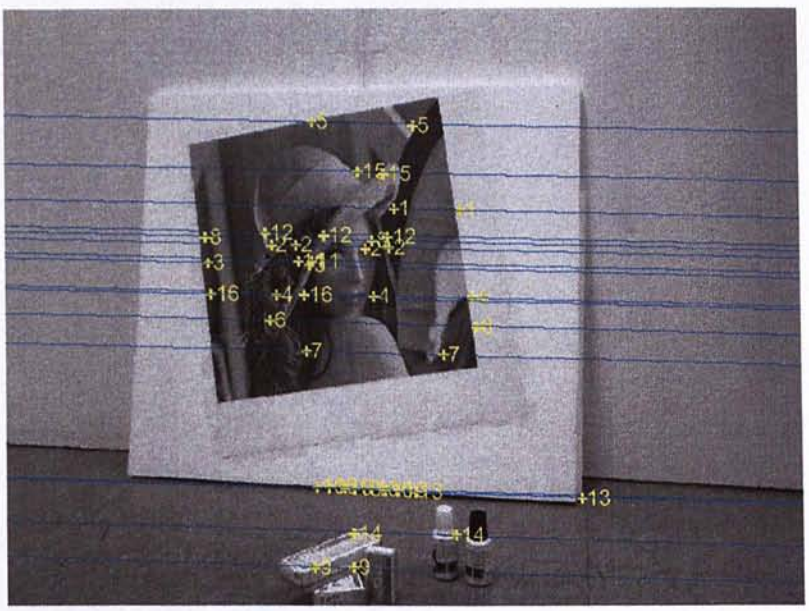

Figure 15: Image 1 of the scene, with 41 feature points marked as yellow "crosses" on 16 conjugates of epipolar lines. 9 of the total 16 groups of features are not totally corresponding.

Similarly, we evaluated the accuracy of the result using the reprojection error method mentioned in the previous section. Totally 40 feature points (Figure. 16) were extracted in image 1, and their reprojected positions were shown as blue "crosses" in image 2 shown as Figure 17. Compared blue "crosses" with the manually picked (true) positions (red "crosses" in Figure. 17), the reprojection error was displayed in Figure. 18. The mean of the reprojection error along x-axis direction is around 33 pixels, which shows the planar parameters calculation is failed. That is, the least-squares error strategy cannot provide a suitable solution since the faulty feature group pairs still have influence on the final estimation.

Figure 16: Image 1 of the scene, with 40 feature points marked as blue "crosses".

Figure 17: Image 2 of the scene, with the manually picked (true) correspondences of the 40 feature points marked as red "crosses", and the projected positions (from Image 1) of the correspondence-free algorithm marked as blue "crosses". Considerable reprojection error is shown in this figure.

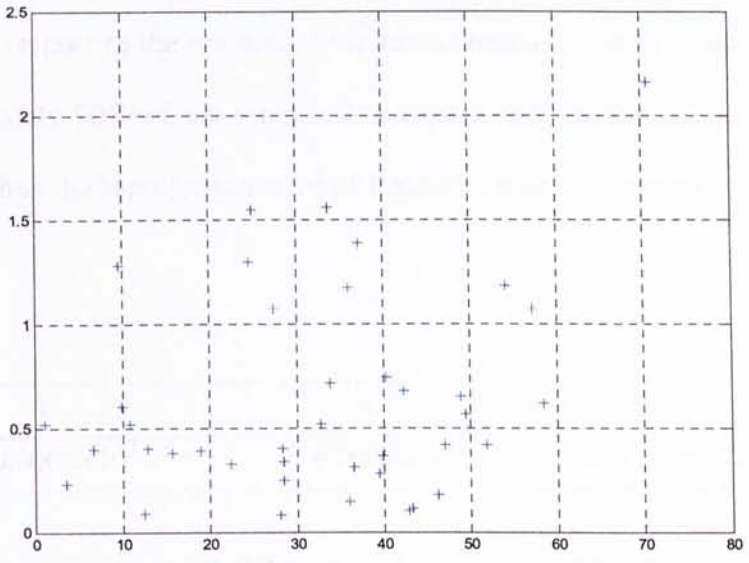

Figure 18: Reprojection errors (in the image domain, in pixel unit) of the correspondence-free algorithm for 40 feature points of the distant scene. The mean of the reprojection error along x-axis direction is around 33 pixels, which shows the planar parameters calculation is failed.

Below we adopted the improved algorithm combined with SVD and robust estimation to obtain the recovered planar surface. After the process of SVD and robust estimation, most of the wrong groups were kicked out. In Figure 19 and Figure20, the epipolar lines without crosses and number represent the faulty groups of features that have been discovered and discarded. We noticed 7 of the 9 obvious faulty groups of features were discarded. Group 4 seemed to be totally corresponding, but it was also cut out. Finally, only 8 groups of features on 8 conjugates of epipolar lines are used to calculate the planar parameters, including two faulty groups, group 13 and 16, which happened to satisfy the linear system and could not be kicked out. The result of plane parameters is shown in Table 6. In despite of two faulty groups (group 13 and 16) of features' presence, the accuracy of the result is still satisfying. We also appraise the experimental results using reprojection error method. The reprojection error was computed with respect to the previous 40 features chosen in image 1, shown as Figure. 21. Approximately 90% of the reprojection error is only within 1.5 pixels, which is much smaller than the reprojection error of the result using the method of least-squares error.

p	q	c
0.0002×10^3	0.0002×10^3	1.0842×10^3

Table 6: 3D reconstruction result of the improved correspondence-free method with SVD and robust estimation.

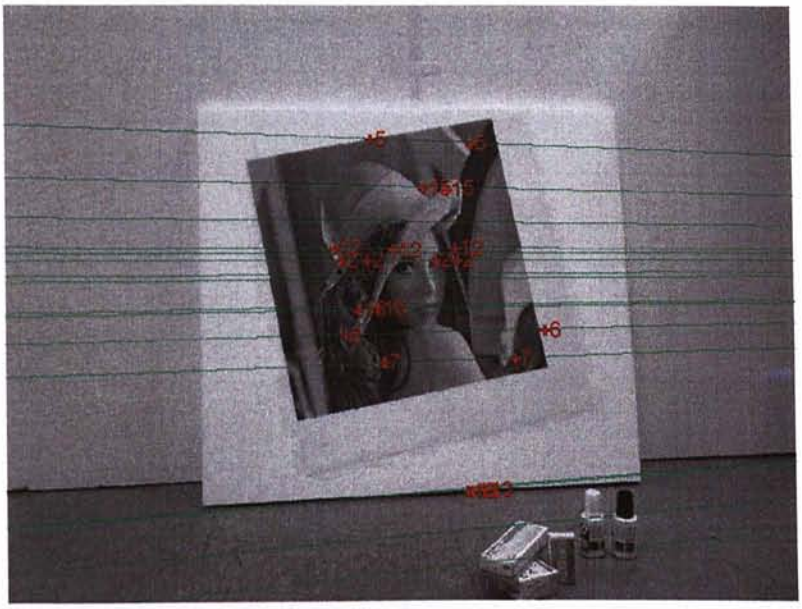

Figure 19: Image 2 of the scene, with 20 feature points marked as red "crosses" on 8 conjugates of epipolar lines.

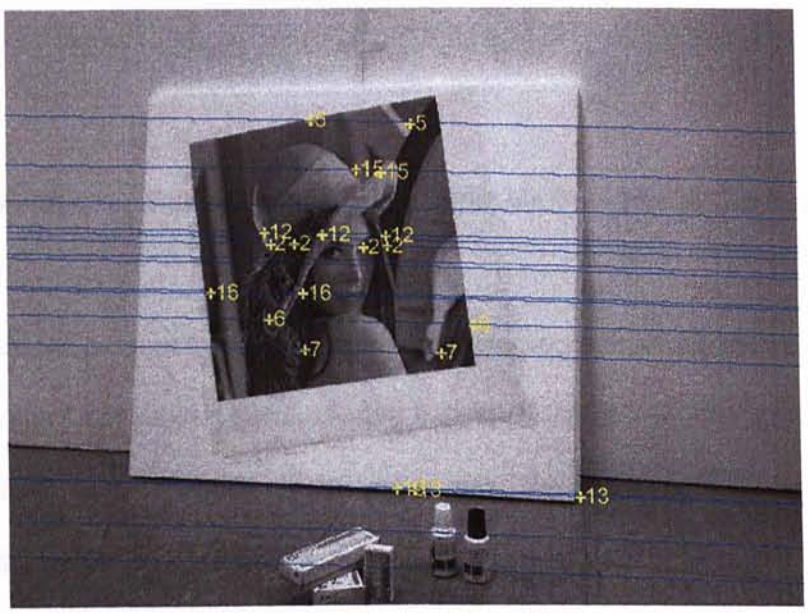

Figure 20: Image 1 of the scene, with 20 feature points marked as yellow "crosses" on 8 conjugates of epipolar lines. Initially, 9 of the total 16 groups of features are obviously not totally corresponding. 7 of the 9 obvious faulty groups of features were discarded using SVD and robust estimation. Group 4 seemed to be totally corresponding, but it was also cut out. Finally, only 8 groups of features on 8 conjugates of epipolar lines are used to calculate the planar parameters, including two faulty groups, group 13 and 16, which happened to satisfy the linear system and could not be kicked out.

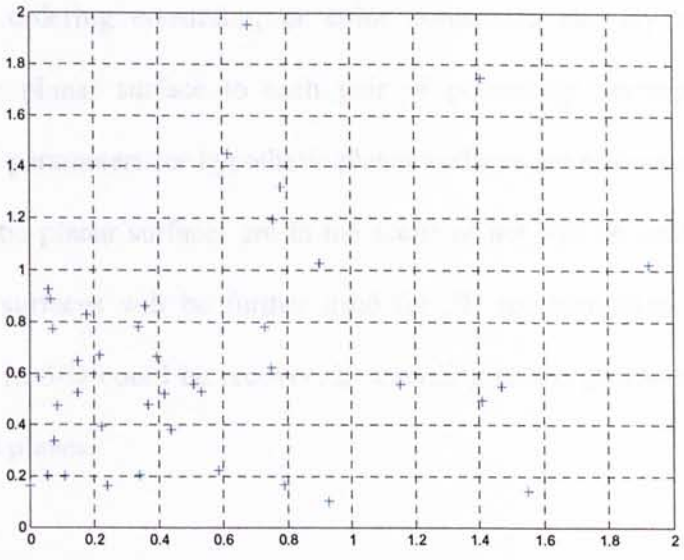

Figure 21: Reprojection errors (in the image domain, in pixel unit) of the correspondence-free algorithm for 40 feature points of the scene.

3.2 Correspondence-free Stereo Vision Algorithm for Multiple Planar Surface Recovery

In this section we propose an algorithm for reconstructing scene with multiple planar surfaces, which does not require point-to-point correspondences over the input images to be established. Experimental results show that the algorithm is capable of reconstructing the polyhedral part of scenes.

The algorithm takes the form of hypothesis-and-confirmation framework. In details, the main idea of multiple planar surfaces recovery without feature-to-feature correspondence is to divide the whole images into some small patches with enough features, such that some of the small patches do not cover more than one surface. Then some potentially corresponding areas across the two images will be established by

epipolar constraint, ordering constraint, or color constraint, etc. By applying the algorithm for single planar surface to each pair of potentially corresponding area candidates, the plane parameters for hypothetic planar surfaces are estimated. After that, whether the hypothetic planar surfaces are in the scene or not will be confirmed. Only the certified planar surfaces will be further used for 3D reconstruction. Finally, the polyhedral structure in 3-D could be recovered, according to the geometrical relations among the recovered planes.

3.2.1 Plane Hypothesis

We adopt the mean shift procedure-based image segmentation [3] to separate the stereo pair of images into some patches. Some smaller area of regions will be eliminated due to fewer features within these regions. Assume there are m regions of interest $R_r=\{I_r^1, I_r^2,\ldots, I_r^m\}$ in image 1, and n regions $R_l=\{I_l^1, I_l^2,\ldots,I_l^n\}$ in image 2.

These regions of interest R_r and R_l will then be separated into several potentially corresponding candidate groups according to similarity of the average gray level (or color) within each region. We denote these candidate groups as: $\{C_r^i, C_l^i\}$, where superscript i represents the ith group of corresponding candidates and $C_r^i \subset R_r, C_l^i \subset R_l$. Suppose $C_r^i = \{I_r^f,\cdots,I_r^h\}$ and $C_l^i = \{I_l^d,\cdots,I_l^k\}$, in which the average intensities (or colors) are similar. The rest regions in R_r (or R_l), whose average intensities (or colors) are quite different from any other region in R_l (or R_r), will be discarded.

The color constraint often cannot establish the one-to-one potentially corresponding regions. Hence, epipolar geometry would be an assistant method in constructing the one-to-one corresponding regions in each candidate group.

Feature points are extracted within any region in a candidate group, for instance, C_r^i and C_l^i. We draw the conjugate epipolar lines across these regions according to their feature points. And we define potentially corresponding ratio as the ratio of the number of conjugate epipolar lines with the same number of feature points on, to the total number of the conjugate epipolar lines across one region in C_r^i and another region in C_l^i. The higher potentially corresponding ratio is, the higher possibility of the one-to-one corresponding regions is. The regions with lower ratio will not be used to estimate planes parameters.

The algorithm for single plane recovery is now applied to each one-to-one potentially corresponding pair of regions, to estimate the plane parameters for the potentially existent planar surfaces.

There are usually some wrong pairs among the potentially corresponding pair of regions. For instance, some regions cover more than one planar surface, or some pairs of regions are completely not corresponding, and also some pairs are only partly matching. However, our improved algorithm for single planar surface recovery is able to handle these cases. SVD and robust estimation method will be of great help in refining the linear systems established with respect to these potentially matching pairs. Moreover, feature points that are near or on the boundary of regions are cast off, since image segmentation cannot provide perfect matching regions, and the boundaries of regions are often roughly located.

3.2.2 *Plane Confirmation And 3D Reconstruction*

The issue is now how to evaluate these planes parameters estimated from each corresponding pair of regions. We use the textureness cross correlation to figure out whether the potentially existent planes appear in the scene or not.

Cross-correlation is one of the standard statistical techniques used for target matching. Suppose one of the planes parameters is (p^h, q^h, c^h) calculated from the matching regions $\{I_r^q, I_l^p\}$, where I_r^q is a region in image 1 and I_l^p is in image 2. Let the set of feature points $m_{rq} = \{\mathbf{m}_{rq}^{\ 1}, \cdots, \mathbf{m}_{rq}^{\ i}, \cdots, \mathbf{m}_{rq}^{\ n}\}$ be within region I_r^q, where $\mathbf{m}_{rq}^{\ i}$ is the ith feature point coordinate with respect to the image frame. Via homography $\mathbf{H}^h$, which is only effective within the surface (p^h, q^h, c^h), the corresponding feature point $\mathbf{m'}_{rq}^{\ i}$ on image 2 should be:

$$\begin{bmatrix} \mathbf{m'}_{rq}^{\ i} \\ 1 \end{bmatrix} \cong \mathbf{H}^h \begin{bmatrix} \mathbf{m}_{rq}^{\ i} \\ 1 \end{bmatrix} \tag{3.17}$$

$$\text{where } \mathbf{H}^h = \begin{bmatrix} R_{11} + t_x(-\frac{p^h}{c^h}) & R_{12} + t_x(-\frac{q^h}{c^h}) & R_{13} + t_x(\frac{1}{c^h}) \\ R_{21} + t_y(-\frac{p^h}{c^h}) & R_{22} + t_y(-\frac{q^h}{c^h}) & R_{23} + t_y(\frac{1}{c^h}) \\ R_{31} + t_z(-\frac{p^h}{c^h}) & R_{32} + t_z(-\frac{q^h}{c^h}) & R_{33} + t_z(\frac{1}{c^h}) \end{bmatrix}.$$

A template $\mathbf{T}_{rq}^{\ i}$ is chosen as a $k \times k$ window, centered by $\mathbf{m}_{rq}^{\ i}$ on image 1. We then obtain a new template $\mathbf{T'}_{rq}^{\ i}$ by warping $\mathbf{T}_{rq}^{\ i}$ according to (3.17). Some elements in $\mathbf{T'}_{rq}^{\ i}$ are perhaps vacant due to the image stretching. The vacant pixels will be filled by interpolation. Finally a $j \times j$ ($j<k$) template $\mathbf{B}_{rq}^{\ i}$ chosen in $\mathbf{T'}_{rq}^{\ i}$ with the same center is used to calculate cross correlation.

Centered by $\mathbf{m'}_{rq}^{\ i}$, we select $l \times l$ base window ($l>j$) $\mathbf{B'}_{rq}^{\ i}$ on image 2. We have to search in $\mathbf{B'}_{rq}^{\ i}$ to find out the highest similarity between $\mathbf{B}_{rq}^{\ i}$ and $\mathbf{B'}_{rq}^{\ i}$, because some errors in (p^h, q^h, c^h) make $\mathbf{m}_{rq}^{\ i}$ and $\mathbf{m'}_{rq}^{\ i}$ not the real matching pairs. Thus, we use the maximum value r^i to stand for the normalized cross correlation coefficient matrix $\mathbf{r}_c^{\ i}$:

$$r^i = \max(\mathbf{r}_c^{\ i}) = \max(f_{Cross-correlation}(\mathbf{B}_{rq}^{\ i}, \mathbf{B'}_{rq}^{\ i})) \tag{3.18}$$

For each feature point in m_{rq}, we will calculate its maximum normalized cross correlation coefficient: $\{r^1, \cdots, r^i, \cdots, r^n\}$, and have the average value r_{avg} to represent how correctness the plane parameter (p^h, q^h, c^h) is, since the higher cross correlation coefficient r_{avg}, the more accurate parameters (p^h, q^h, c^h) are. Furthermore, we hope that the feature points with strong texture have more effect on the cross correlation coefficient. Therefore, we introduce the concept of 'textureness' to make the criteria more convincible. We define that the textureness is related to the standard deviation of $\mathbf{B'}_{rq}^{\ i}$ defined as $\mathbf{T'}_{std}^{\ i}$, and the standard deviation of $\mathbf{B}_{rq}^{\ i}$ defined as $\mathbf{T}_{std}^{\ i}$ for the feature point $\mathbf{m}_{rq}^{\ i}$. Then weighted correlation coefficient for the set of feature points m_{rq} within region I_r^q is:

$$C_r^{\ h} = \sum_{i=1}^{n} T'_{std}{}^{i} \cdot r^i \Big/ \sum_{i=1}^{n} T_{std}{}^{i} \tag{3.19}$$

The plane (p^h, q^h, c^h) is accurately estimated and really exists in the scene, if the value of $C_r^{\ h}$ is high enough.

If there are also some planes parameters that are similar in value, then any one of them can be selected to stand for other similar ones.

On the work of recovering the 3D structure of the planar surfaces, we did not turn to the boundary detection, since we are now concentrated on some simple polyhedral structure recovery. Thus our reconstruction work is based on the geometrical relations among the recovered planes combined with visible-invisible segment of each plane according to the new viewpoint.

3.2.3 Experimental Results

We have tested the proposed algorithm for multiple planar surfaces recovery with real image data to investigate its performance. Two sets of experimental results will be shown below. In all the results shown, we used the algorithm described in [10] to estimate the intrinsic parameters (**A**) and the extrinsic parameters (**R**, **t**) of the cameras, and used (**A**, **R**, **T**) to calculate **F**. We give the epipolar constraint a tolerance of about 2 pixels, since there is inevitable small error over the constraint. We adopt Harris' corner detection [8] to extract features, and apply mean shift procedure-based image segmentation [3] to the input images.

3.2.3.1 Experiment on a Box Model

This experiment was aimed at investigating the 3D reconstruction accuracy of the described algorithm for multiple planar surfaces. The images were taken with a baseline of about 0.5m wide, and are of resolution 1280×960, which are shown in Fig. 28 and Fig. 29 respectively. Totally 6 planar surfaces are in the scene. For the sake of better description we name the surfaces as surface 1 to 6, which are shown in Fig. 26. We applied mean shift procedure-based image segmentation [3] and set the same

parameters to the input stereo image pair. Segmentation results of the two images are different in some regions, as shown in Fig.22 and Fig.23.

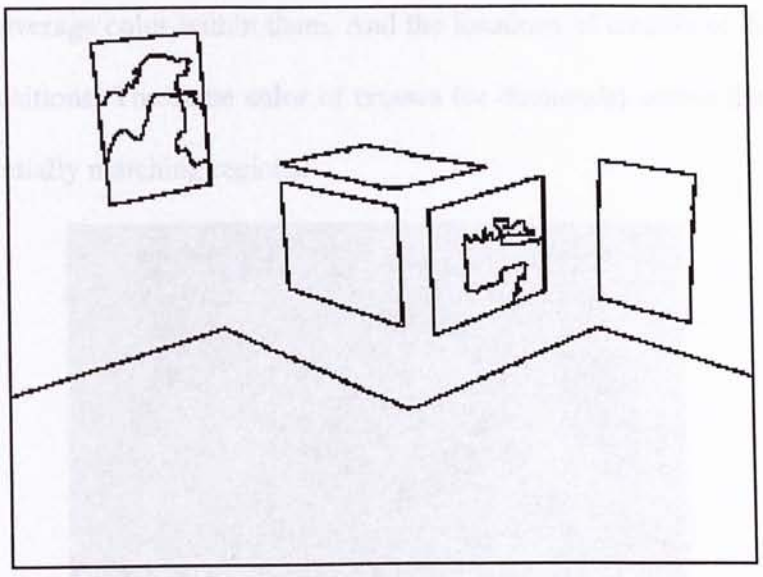

Figure 22 :Segmentation result of Image 1. The image is divided into 9 regions.

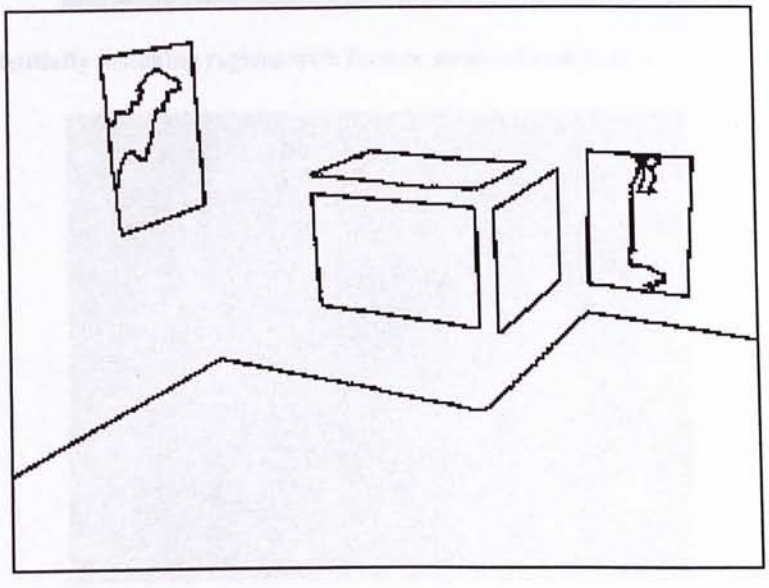

Figure 23 : Segmentation result of Image 2. Totally 9 regions are within the image.

750 features were extracted on the images by Harris' corner detector [8], respectively. Then these features are grouped according to the different regions. We cast off the feature points that are near or on the boundary of these regions, in the case of

rough boundaries of these regions. We applied epipolar constraint and color constraint to establish the potentially matching regions. In Fig. 24 and Fig. 25, different regions are filled with the average color within them. And the locations of crosses or diamonds are the features' positions. The same color of crosses (or diamonds) across the image pair represents potentially matching regions.

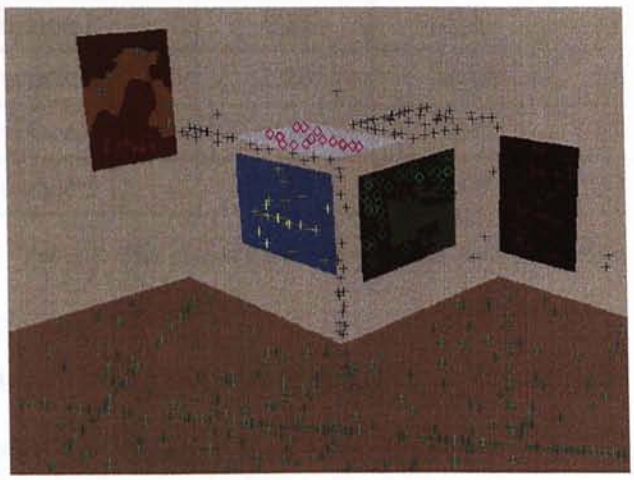

Figure 24: Potentially matching regions with feature points of Image 1.

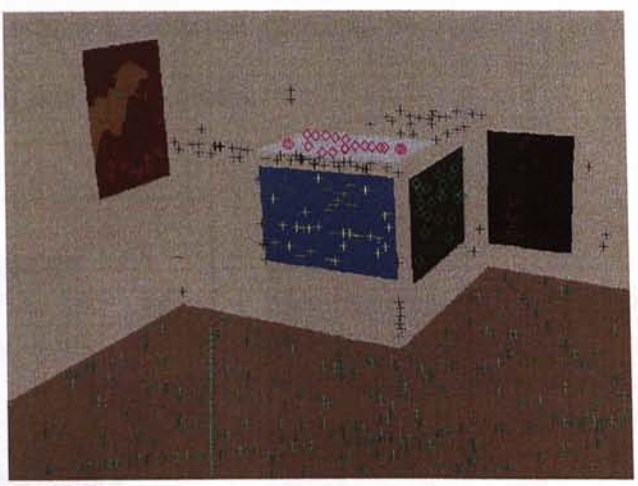

Figure 25: Potentially matching regions with feature points of Image 2. The same color of crosses (or diamonds) across Image 1 and Image 2 represents potentially matching region pairs.

Totally 8 planes were obtained by our potentially existent planes estimation. However, 3 of the 8 ones were proved to be wrong by the planes refining method. The planes parameters (p, q, c) estimated by the algorithm are shown in Table 7:

	p	q	c
Surface 1	1.57	-0.46	2237.39
Surface 2	-0.88	-0.26	1328.18
Surface 3	1.20	-0.33	1403.99
Surface 4	-0.86	-0.09	1904.94
Surface 5	0.10	3.00	2052.48

Table 7: The planes parameters for the multiple surfaces.

We warped image 1 to the viewpoint of image 2, via the estimated planes parameters. In Figure 23 we notice that surface 6 is in a wrong position, due to its wrong recovered parameters, which had been discovered and eliminated by the plane Confirmation method. Surface 6 was not estimated correctly because it is very close and almost parallel to the optical axis of the cameras.

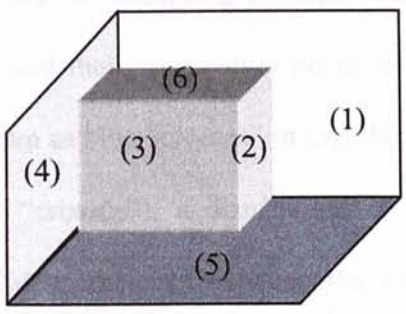

Figure 26: Box Model

Figure 27: Image reprojection

Some surfaces of the box are approximately orthogonal in 3D, and some ones are approximately parallel to each other. Therefore, we tested the estimated plane parameters by calculating the angle between any recovered surfaces. The results are shown in Table 8:

	Surface 1	Surface 2	Surface 3	Surface 4	Surface 5
Surface 1		84.2014°	7.2923 °	83.2233 °	87.9390°
Surface 2	84.2014°		89.3035 °	6.9183 °	88.1281°
Surface 3	7.2923°	89.3035°		89.9311°	88.5402°
Surface 4	83.2233°	6.9183 °	89.9311°		81.2563°
Surface 5	87.9390°	88.1281°	88.5402°	81.2563°	

Table 8: The angles between recovered surfaces of the polyhedral scene.

To achieve a further evaluation of the reconstruction accuracy, we manually chose 43 features on the recovered 5 surfaces from Image 1 (Fig. 28), recovered their 3-D information according to their planes parameters estimated from the algorithm, and acquired their reprojected positions in Image 2. Comparing the reprojected positions (shown as blue "crosses" in Fig. 29) with the manually picked (true) positions (shown as pink "crosses"), a distribution of the reprojection error (measured in terms of the Euclidean distance between the reprojected and the true position) could be attained (Fig.30). It was observed that 97.67% of the reprojections have reprojection error less than 3 pixels.

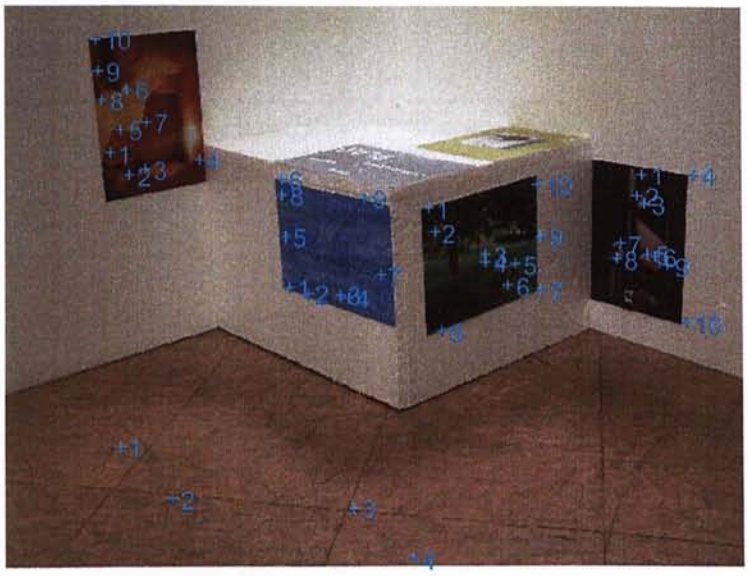

Figure 28: Image 1 of the polyhedral scene, with 45 feature points marked as blue "crosses".

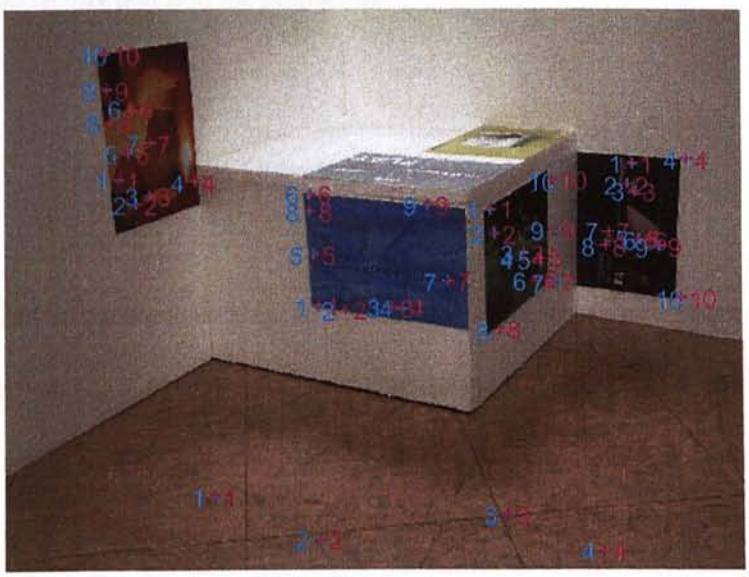

Figure 29: Image 2 of the polyhedral scene, with the manually picked (true) correspondences of the 45 features marked as pink "crosses", and the projected positions (from Image 1) of the correspondence-free algorithm marked as blue "crosses".

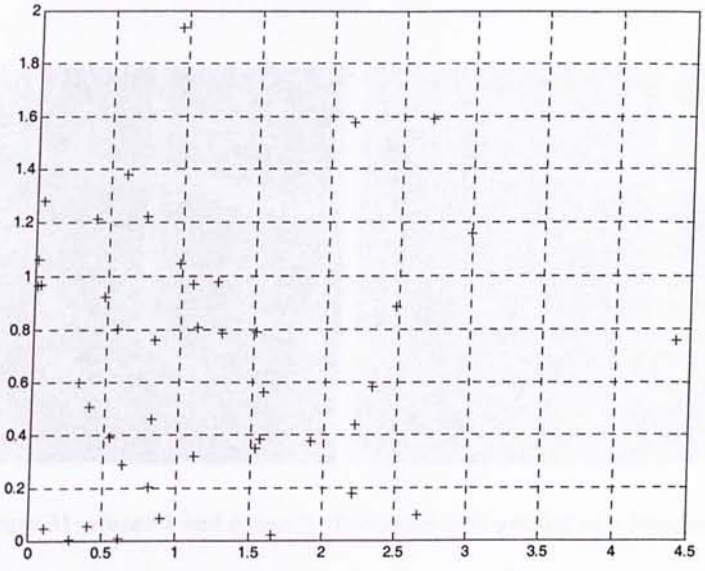

Figure 30: Reprojection errors (in the image domain, in pixel unit) of the correspondence-free algorithm for 45 feature points of the polyhedral.

3.2.3.2 Experiment on Outside Scene

The proposed algorithm is able to handle the case in which there are repetitive patterns in the image data. Usually, establishment of the convincible point-to-point correspondence is rather difficult in this case. Even using software *image-matching* of Z.Y. Zhang [27] or the software of P. H. S. Torr [24], a large portion of correspondences could not be estimated correctly. Here we show our experiment result on these scenes. The scene is about approximately orthogonal walls standing on the ground of a building. The two images of the scene are shown in Figure 31.

Figure 31: Image 1 and image 2 of the walls and ground of a building

The baseline between the two viewpoints was about 0.7 m. the images are both of resolution 1280×960. 450 feature points on each image were detected for estimating planes parameters. We denote that the ground is surface 1, the wall on the right is surface 2 and the left wall is surface 3.The planes parameters (p, q, c) estimated by our algorithm are shown in Table. 9:

	p	q	c
Surface 1	0.02	2.62	3949.08
Surface 2	1.05	-0.35	3037.17
Surface 3	-1.13	-0.21	3784.59

Table 9: The planes parameters for 3 surfaces of the building.

The angles between the three recovered surfaces are shown in Table 10. (Angle <1~2> represents the angle between surface 1 and surface 2.)

Angle <1~2>	Angle <1~3>	Angle <1~3>
89°	84°	87°

Table 10: The angles between three recovered surfaces of the building.

We recovered the scene according to the estimated planar parameters and relations among the planes. The structure under new viewpoints is shown in Figure 32.

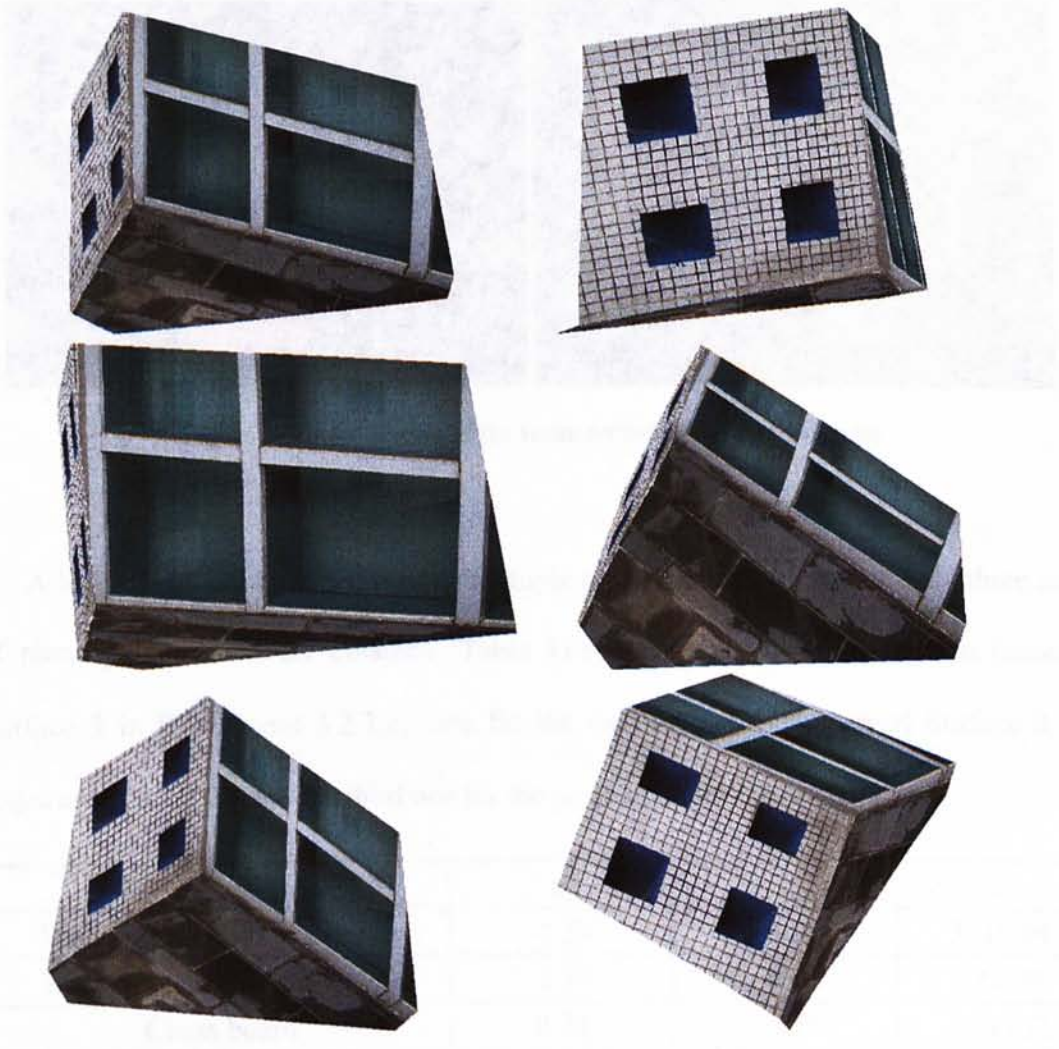

Figure 32: Reconstruction of the walls of the building under new viewpoints.

3.2.3.3 Experiment on Occlusion

This experiment aims at testing the robustness of our system. The scene is same as the previous one (Experiment 3.2.3.2), except a chair and a piece of board in front of the walls. In this case, a lot of parts of the two walls and the ground are occluded by the

obstacles. The input images are shown in Figure 33. 750 features (shown as red crosses) are detected on the two images.

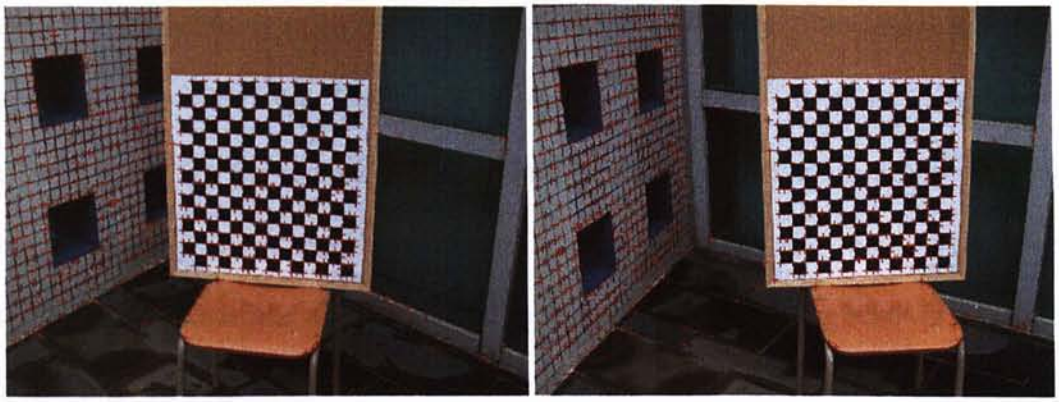

Figure 33: Input images of the scene occluded by some obstacles

After we applied the algorithm for multiple planar surfaces recovery, only three sets of plane parameters were obtained (Table 11): one for the wall on the left (named Surface 3 in Experiment 3.2.3.2), one for the wall on the right (named Surface 2 in Experiment 3.2.3.2), and the third one for the piece of board (chessboard):

	p	*q*	*c*
Wall on the left (Surface 3)	-1.59	-0.37	3715.94
Wall on the right (Surface 2)	-1.80	-1.90	2002.98
Chess board	0.72	-0.33	2790.52

Table 11: Plane parameters for the scene.

We notice that there was no result for the ground in the scene, and the parameters for the wall on the right (Surface 2) was wrong, moreover large error was in the parameters for the wall on the left (Surface 3). The following is the analysis of the reason why this experiment is failed.

Uneven distribution of the feature points on the input images is the major reason due to the unequal distribution of the texture within the images. In Fig.33, we noticed most of the features were detected within the areas of the wall on the left and the chessboard. However there were a few features distributed in the rest of parts of the image because of less texture. To the regions full of feature points, it is difficult to group the feature points correctly according to different conjugate of epipolar lines, because the distribution of the features is too dense. Thus, faulty feature groups result in great error. To the rest of regions, where there was lack of the candidates for the parameters estimation due to the few features within these areas, the plane parameter could not be calculated. Therefore, if the texture were distributed equally throughout the whole image, our system would be robust enough to work well.

3.3 Experimental Results on Correspondence-free Vs. Correspondence Based Methods

In this subsection, we will compare our correspondence-free based method with the traditional correspondence based algorithm. We applied Z. Y. Zhang's *Image Matching* software to some stereo image pairs to generate corresponding features. Zhang's technique represents state of the art in stereo vision, and has been widely tested and works well.

As it is said in the previous chapter, the advantage of our correspondence-free method is to deal with the image pair with repetitive patterns. Fig. 31 shows our input image pair. Table 12 is the main parameters we set in the *Image Matching* software:

Size of Correlation Window:	7	Max Number of Points:	800
Max Disparity in X &Y Direction:	200	Max Number of Iteration:	20
Threshold in Correl. Score:	0.8	Radius of Neighborhood:	64

Table 12: Parameters set for software *Image Matching*

Figure 34 shows the matching result by Zhang's algorithm. Totally 212 correspondences were established. Among the 212 correspondences, we notice that only one matching pair (No. 189) of feature points on the ground was established, besides 3 pairs (No. 200, 204, 211) on the boundary of the ground. It was not reliable to make use this kind of four correspondences to achieve the plane parameter for the ground. Moreover, many wrong matches appeared on the images of the left wall, because of the plentiful repetitive patterns and dense features extracted by Harris' corner detector. Therefore, large error would occur in estimating the planes parameters for these two surfaces.

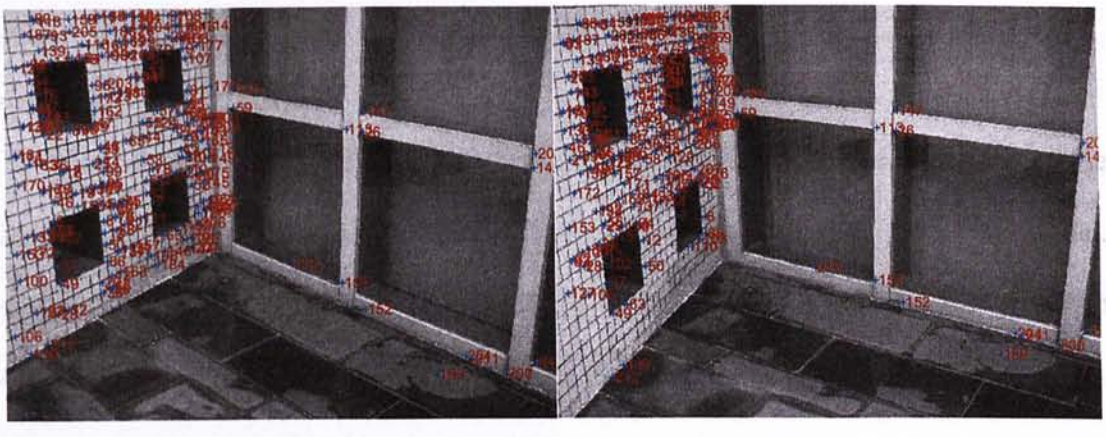

Figure 34 : Correspondences estimated by Z. Y. Zhang's *Image Matching* software

However our correspondence-free stereo algorithm is relatively powerful to handle the same image pair. Result of our correspondence-free method is shown in Table 9 and Table 10 in the previous chapter.

Such cases as Figure 31 are a little special. The next experiment on comparing correspondence and correspondence-free method is about a more general scene. Figure 35 shows the two input images. The scene at significant distance could be treated as planar scenes, as we have mentioned before.

Figure 35: Input image pair. 28 pairs of matches were selected for the use of calculation of reprojection error.

We still adopted Zhang's method here to establish the correspondences across the input images. Then these correspondences were used to estimate the plane parameters, although 14 of entire124 correspondences are wrong by inspection. Also our proposed correspondence-free algorithm is capable to estimate the plane parameters. Similarly, the plane parameters calculated by the two different methods were evaluated by the value of reprojection error, which was described amply in Section 3.1.2. Totally 28 pairs of matches shown in Figure 35 were selected for the use of calculatio mn of reprojection

The diagrams of reprojection error standing for the accuracy of the results calculated by the two different methods are shown in Figure 36.

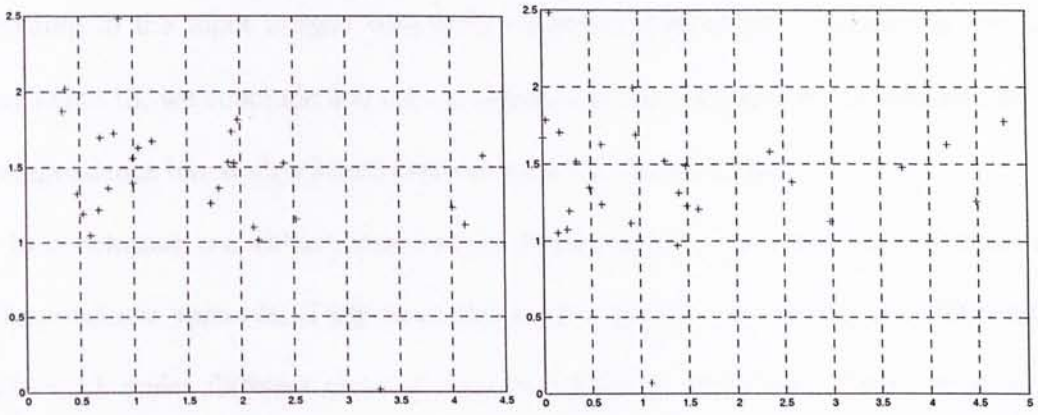

Figure 36: Diagrams of reprojection error. The left one is the reprojection error for evaluating the plane parameters estimated by the 124 pairs of correspondences, while the right diagram is for our proposed correspondence-free algorithm.

Only a little differences could be observed from the above two diagrams. Hence we use means and variances to value the reprojection error, and consequentially evaluate the plane parameters. The statistic means and variances of the reprojection error are shown in Table 13(unit: pixel):

	Correspondence	Correspondence-free
Mean of error in x-direction	1.6668	1.4538
Mean of error in y-direction	1.4319	1.4093
Variance of error in x-direction	1.1551	1.4134
Variance of error in y-direction	0.4301	0.4194

Table 13: Statistic means and variance of the reprojection error

The mean of reprojection error is merely around 1.5 pixels in both x-direction and y-direction for either correspondence algorithm or our correspondence-free method, according to the input images with both resolution 1600×1200. Comparing the data from Table 13, we conclude that our correspondence-free algorithm is as accurate as the correspondence based algorithm in estimating the plane parameters.

In conclusion, our correspondence-free algorithm does not contradict the traditional correspondence methods. They have the same objective (recovering the 3D world), while work under different point of view and different conditions. Therefore, it is not appropriate to judge which method is better or not. The two methods could be applied alternatively to deal with the planar surfaces if no limitation of conditions, such as abundant repetitive patterns in the images. Even in some cases, we hope the combination of both of the algorithms, for the sake of accuracy and efficiency of the 3D recovering. Reconstructing the 3D world by all means is our ultimate target, after all.

CHAPTER FOUR

CONCLUSION AND FUTURE WORK

The establishment of correspondences, i.e., pairing up feature points in two images such that each pair of points are projected by the same3D point in space, is a key and difficult problem of stereo vision. Aloimonos and Herve (1990) presented an algorithm that could recover 3D structure of a single planar surface without requiring point-to-point correspondences to be established. The algorithm is however applicable only to images captured under the parallel-axis stereo imaging geometry.

In this thesis, we present a novel algorithm for recovering scene consisting of a single planar surface under general stereo geometry. The algorithm is capable to deal with the planar surface disturbed by some small obstacles for its robustness. Then we go one step further and propose a method that could recover even scene consisting of multiple planar surfaces without establishing the one-to-one correspondences across the input images. And no specific restriction on the stereo imaging geometry is needed. Experimental results show that the algorithm is effective in recovering some 3D structure of polyhedral scene. Especially, our method shows advantages when dealing with repetitive textures and bold patterns in the image pairs, which is rather difficult to the traditional correspondence method.

However, our correspondence-free method for multiple planar surfaces recovery still has large room to make improvements. Now our proposed algorithm works using feature points. Feature points' extractions are often different in the input image pairs, which

results in the difficulties for the following steps. Thus, we will refine our present algorithm by using of lines together with points, not just points in our future work. On reconstructing 3D structure, we should work out a more effective strategy to handle of textureless regions, where plan confirmation is absent during the 3D reconstruction. Our reconstruction now is only based on the planes parameters combined with visible-invisible segment of each plane according to the new viewpoint. In our future work, plane region determination should be from full information about planes and partial information about which regions they might cover.

Therefore, the work of structure reconstruction is still a challenge although the 3D planes parameters are obtained, if handling scenes with occlusions.

APPENDIX

6.1 Homography

The concept of homography is the foundation of the algorithms we developed in thesis. In this chapter, we will give a brief introduction of it.

3D points $\mathbf{M}_0$, $\mathbf{M}_1$ and $\mathbf{M}_2$ define a plane $\prod$. Then for any point $\mathbf{M}_i$ on the same plane $\prod$ will be described as:

$$\mathbf{M}_i = [(\mathbf{M}_1 - \mathbf{M}_0) \quad (\mathbf{M}_2 - \mathbf{M}_0) \quad \mathbf{M}_0] \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{6.1}$$

where α and β are any scalar.

We know that by projective engine, which maps the three-dimensional projective space P^3 onto the two-dimensional projective plane P^2 by perspective projection, the projected image points $\mathbf{m}_i$ and $\mathbf{m}'_i$ of $\mathbf{M}_i$ on the two retinas respectively will be:

$$\begin{bmatrix} \mathbf{m}_i \\ 1 \end{bmatrix} \cong \mathbf{P}^1{}_E \begin{bmatrix} \mathbf{M}_i \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{m}'_i \\ 1 \end{bmatrix} \cong \mathbf{P}^2{}_E \begin{bmatrix} \mathbf{M}_i \\ 1 \end{bmatrix}$$

Substitute term $\mathbf{M}_i$ by equation 1, then we have:

$$\begin{bmatrix} \mathbf{m}_i \\ 1 \end{bmatrix} \cong \mathbf{P}^1{}_E \left[\left(\begin{bmatrix} \mathbf{M}_1 \\ 1 \end{bmatrix} - \begin{bmatrix} \mathbf{M}_0 \\ 1 \end{bmatrix} \right) \quad \left(\begin{bmatrix} \mathbf{M}_2 \\ 1 \end{bmatrix} - \begin{bmatrix} \mathbf{M}_0 \\ 1 \end{bmatrix} \right) \quad \begin{bmatrix} \mathbf{M}_0 \\ 1 \end{bmatrix} \right] \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix}$$

which can be rewritten as:

$$\begin{bmatrix} \mathbf{m}_i \\ 1 \end{bmatrix} \cong \left[\left(\omega^1{}_1 \begin{bmatrix} \mathbf{m}_1 \\ 1 \end{bmatrix} - \omega^1{}_0 \begin{bmatrix} \mathbf{m}_0 \\ 1 \end{bmatrix} \right) \quad \left(\omega^1{}_2 \begin{bmatrix} \mathbf{m}_2 \\ 1 \end{bmatrix} - \omega^1{}_0 \begin{bmatrix} \mathbf{m}_0 \\ 1 \end{bmatrix} \right) \quad \omega^1{}_0 \begin{bmatrix} \mathbf{m}_0 \\ 1 \end{bmatrix} \right] \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{6.2}$$

where $\omega^1{}_0$, $\omega^1{}_1$ and $\omega^1{}_2$ are any nonzero scalar, and $\mathbf{m}_0$, $\mathbf{m}_1$, $\mathbf{m}_2$ are the projections of the 3D points on image plane 1.

We use symbol ${\mathbf{P}^1}_B$ to simplify equation (6.2):

$$\begin{bmatrix} \mathbf{m}_i \\ 1 \end{bmatrix} \cong {\mathbf{P}^1}_B \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{6.3}$$

Similarly, the projections of the 3D points on the image plane 2 can be expressed as:

$$\begin{bmatrix} \mathbf{m'}_i \\ 1 \end{bmatrix} \cong {\mathbf{P}^2}_B \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{6.4}$$

Combine Equation (6.3) with (6.4), and eliminate the term $[\alpha\ \beta\ 1]^{T.}$ Therefore, the homography, which occurs on the 2D image planes due to plane $\prod$ will be defined as:

$$\begin{bmatrix} \mathbf{m'}_i \\ 1 \end{bmatrix} \cong \mathbf{H} \begin{bmatrix} \mathbf{m}_i \\ 1 \end{bmatrix}$$

where $\mathbf{H} = {\mathbf{P}^2}_B ({\mathbf{P}^1}_B)^{-1}$.

That is, given 2 images of a planar surface $\prod$ (or 2 images of a very distant scene, or 2 images of a general scene that is pictured under only a pure rotation of camera), every pair of corresponding image positions (that are projected by the same point of $\prod$) in the two images, $\mathbf{m}=(u,v)$ and $\mathbf{m'}=(u',v')$, are related by

$$[u',v',1]^T \cong \mathbf{H}[u,v,1]^T \tag{6.5}$$

The 3×3 matrix **H** represents the mapping named homography induced by surface $\prod$ to the two images [5]. In other words, once **H** is known, every position in one image could be warped to the image domain of the other image using Equation (6.5). It could be observed from Equation (6.5) that **H** could be estimated from as few as four point correspondences over the images.

BIBLIOGRAPHY

[1]: J. Aloimonos and J. Y. Herve. "Correspondence-free stereo and motion: planar structure. " *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 12, No. 5, pp. 504-510 (1990).

[2]: V. Barnett and T. Lewis. Outliers in Statistical Data, pp. 328-340, Wiley, N. Y,1994.

[3]: D. Comanicu, P. Meer: "Mean shift: A robust approach toward feature space analysis". IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, pp.603-619, May 2002.

[4]: U.R. Dhond and J.K. Aggarwal. "Structure from stereo- A review," IEEE Transactions on Systems, man and Cybernetics, 19(6): 1489-1510, November /December 1989.

[5]: O. Faugeras, "Stratification of three-dimensional vision: projective, affine, and metric representations", in Journal of the Optical Society of America: A, March 1995, Vol. 12, No. 3, pp. 465-484.

[6]: O. Faugeras and Q. T. Luong, The Geometry of Multiple Images, The MIT Press, 2001.

[7]: W. Grimson, "Computational experiments with a feature based stereo algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 7, pp.17-34, 1985.

[8]: C. Harris and M. Stephens: "A combined corner and edge detector. " In Proc. Alvey Conf., pp 189-192, 1987.

[9]: R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.

[10]: J. Heikkilä. "Geometric camera calibration using circular control points", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, pp. 1066-1077, October 2000.

[11]: G.A. Jones. "Constraint, Optimization, and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features. " Computer Vision and Image Understanding, 65(1): 57-78, January 1997.

[12]: T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment", in Proceeding of the IEEE International Conference on Robotics and Automation, pp.1088-1095, Califirnia, April 1991.

[13]: T. J. Keating, P. R. Wolf and F. L. Scarpace, "An Improved Method of Digital Image Correlation", Photogrammetric Engineering and Remote Sensing, Vol. 41, No. 8 (August), pp.993-1002 (1975).

[14]: R. Klette, K. Schluns and A. Koschan, Computer Vision Three-Dimensional Data from Images, Springer-Verlag Press, Singapore, 1998.

[15]: A. Koschan, "A frame work for area-based and feature-based stereo vision", Machine Graphics and Vision, Vol. 2, No. 4, pp.285-308, 1993.

[16]: M. D. Levine, D. A. O'Handley and G. M. Yagi, " Computer determineation of depth maps", Computer Graphics and Image Processing, Vol. 2, pp.131-150, 1973.

[17]: D. Marr and T. Poggio. "Cooperative Computation of Stereo Disparity. " Science, 194:283-287, October 1976.

[18]: K. I. Mori, M. Kidode and H. Asada, "An interactive prediction and correction method for automatic stereo comparision. " Computer Graphics And Image Processing, 2:393-401, 1973.

[19]: V. S. Nalwa, A Guided Tour of Computer Vision, Addison-Wesley Publishing Company, 1993.

[20]: T. Poggio, V. Torre and C. Koch, "Computational vision and regularization theory, " Nature, Vol. 317, pp. 314-319, September 1985.

[21]: L. L. Scharf, "The SVD and reduced-rank signal processing, " in SVD and Signal Processing II: Algorithms, Analysis and Applications (R. J. Vaccaro, ed.), pp.3-31, Elsevier Science Publishers, 1991..

[22]: D. Scharstein, "A Taxonomy and Evaluation of Dense Two-Frame StereoCorrespondence Algorithms", International Journal of Computer Vision, Vol.47, pp.7-42, May 2002.

[23]: Y. Shirai, Three-Dimensional Computer vision, Springer-Verlag Press, Berlin Heidelberg, 1987.

[24]: http://research.microsoft.com/~philtorr/

[25]: O. Veksler, "Stereo correspondence with compact windows via minimum ratio cycle", in Proceeding of the Intenational Conference on Computer Vision, pp. 540-547, 2001.

[26]: M. Yachida, Y, Kitamura and M. Kimachi, " Trinocular Vision: new approach for correspondence problem ", in Proceeding of 8th International Conference on Pattern Recognition, pp.1041-1044, France, 1986.

[27]: Z.Y. Zhang: http://www-sop.inria.fr/robotvis/personnel/zzhang/softwares.html