Proceedings of the 1996 IEEE
International Conference on Robotics and Automation
Minneapolis, Minnesota - April 1996

# A Stereo Matching Algorithm Based on Shape Similarity for Indoor Environment Model Building

Xueyin Lin, Zhigang Zhu, Wen Deng
Department of Computer Science and Technology, Tsinghua University
Beijing 100084, China

## Abstract

*A novel stereo matching algorithm has been proposed in this paper based on the observation that there are plenty of horizontal features in indoor environments. In this algorithm, the shape identity of all horizontal feature correspondences is achieved by means of a special image transformation —reprojection transformation. By using edge string as an intermediate structure, a generalized Hough transformation strategy fuses a variety of criteria, such as shape similarity, disparity continuity and pixel attribution similarity, etc., together. In this way it copes with the correspondence of horizontal and non-horizontal features for 3D environment model building. First the principle of reprojection transformation using calibrated or uncalibrated images are discussed, then the generalized Hough transform strategy is addressed. Preliminary experimental results indicate that the algorithm is an effective and efficient method for indoor environment model building by using stereo methodology.*

## 1. Introduction

Stereo vision has been an important problem in computer vision for decades. Nowadays it has been applied increasingly to create models of both indoor and outdoor environments, such as terrain and other natural environments for various kinds of use in flight simulation, virtual reality applications, and human-computer interactions [4][5]. Recently research on stereo has been extended to incorporate more than two views in a projective framework and has made remarkable progress [2][3][7].

One of the crucial problem to be solved in stereo vision has been identified as finding the correct correspondences of pair-wise related image points, which represent a single point in the physical scene, and it has been the focus of activities within the stereo vision research area for many years. The majority of the stereo matching approaches can be roughly classified as the area-based and feature-based methods. Area-based methods choose point as matching primitive and judge matching based on attribution similarity of the neighborhood of the points. It is supposed to be more flexible than feature-based methods and is more suitable for natural scene reconstruction. On the other hand, feature-based methods use feature as matching primitive. As features are extracted and described based on information gathered in a rather large region, they are less ambiguous than point primitives and hence make the stereo matching more reliable than the former. The problem is, however, that as features are usually extracted within a rather large region their shapes might probably be deformed by the effect of perspective projection drastically. As a result, feature correspondence should be based on high level invariant such as perception organization, cross-ratio, etc. Since high level invariant is usually abstracted from global-wise information, they are usually sensitive to the performance of segmentation and suffer from occlusion. Besides feature-based method is less flexible than the area-based one.

Recently we have developed a novel method of stereo matching for indoor environment model building. The novelty of our method stems from introducing image geometric transformation into stereo vision. The method is motivated by the observation that if the locations of all the horizontal surface patches in a room have been determined the layout of that room will be roughly determined. Images captured by stereo cameras are transformed by special designed image transformation, which we call reprojection transformation, so that features in both images representing the same horizontal figure are with similar or even identical shape and hence shape similarity can be used to ease the establishing of feature correspondence. This method that favors horizontal feature correspondence, however, should not compromise non-horizontal figure correspondence. Based on the consideration a generalized Hough transform strategy is used for feature matching. In this strategy edge string is used as an intermediate data structure, and a variety of criteria, such as shape similarity, attribution similarity of its constituent feature point, are applied to establish feature correspondence. In this way the advantages of both area-based and feature-based methods are taken to make feature matching reliable and flexible.

0-7803-2988-4/96 $4.00 © 1996 IEEE

The organization of this paper is as follows. The principle of the reprojection transformation we use and its advantages in solving stereo matching is discussed first. The outline of our stereo matching algorithm and some of its important issues are described next. Some of our experimental results are presented afterwards, followed by a discussion in the last section.

## 2. Reprojection transformation

### 2.1 Principle of reprojection transformation

An image transformation is purposively designed and applied to the stereo pair. It makes the pair-wise features in the stereo pair, corresponding to figures lying on horizontal planes, render the same shapes. The possibility and realization of designing such an image transformation are discussed through the proof of the following theorems. The fact, that a pair of pictorial figures to be similar, is equivalent to that the angle between the straight line pair, passing through any corresponding point pair in both figures, is identical. In the following theorems only the latter is addressed.

**Theorem** 1: The reprojection transformation can be developed to any pair of stereo images so that pair-wise related image lines that represent a single straight line lying on a parallel plane set PP in scene, can share the same direction.

**Theorem** 2: The sufficient and necessary condition for the pair-wise related image line segments, mentioned above, in the transformed stereo image pair, to be with equal length and the same direction is that the connection line of both optical centers of the cameras be parallel to the parallel plane set.

The proof will be discussed after some notations and definitions have been addressed.

**Notations and definitions:**

a. The stereo image pair $\psi_1$ and $\psi_2$ are captured by camera $O$ and $O'$ respectively.

b. Without losing the generality a world coordinate system is established with its origin at the optical center of the first camera $O$, and a 3 × 1 vector $X$ is used to indicate coordinates of any point in space and $x$ its homogeneous coordinate vector in $\psi_1$. The coordinate of optical center of the second camera is expressed as vector $O'$.

c. A second coordinate system is established by shifting the world coordinate system to the optical center $O'$, and $X'$ is used to indicate the respective coordinates of any space point. Therefore the relation of coordinates of any space point in both systems is

$$X' = X - O' \qquad (1)$$

d. The normal vector with unity length of PP is n. One of the plane $\pi \in$ PP is arbitrarily chosen as its representative and its equations respective to the two systems are

$$n^T X = d_\pi, \quad \text{and}$$

$$n^T X' = d_\pi - n^T O' = d'_\pi \qquad (2)$$

where T is a transpose symbol, $d_\pi$ and $d'_\pi$ are two scalar numbers indicating the distances between π with $O$ and $O'$ respectively.

e. On plane π a pair of projective points are defined for any point $\notin$ π. Each of the projective points is the intersection point of plane π with an optical ray, from the respective camera center to that space point. Therefore the respective projective points of a space point $X$, $X_1$ and $X_2$, can be calculated as

$$X_1 = \frac{d_\pi}{d} X \text{ and } X_2 = \frac{d'_\pi}{d'} X' + O' \qquad (3)$$

where $d = n^T X$ and $d' = n^T X' = n^T(X - O') = d - n^T O'$ respectively. Projective line of a space line is similarly defined and can be described by its two ends. For example, if a line segment on one plane of PP other than π is described by its two end points $X^1$ and $X^2$, as seen in Fig. 1, from (3) we have

$$X_1^2 - X_1^1 = \frac{d_\pi}{d}(X^2 - X^1) \text{ and}$$

$$\begin{aligned} X_2^2 - X_2^1 &= \frac{d'_\pi}{d'} X^{2'} - \frac{d'_\pi}{d'} X^{1'} \\ &= \frac{d'_\pi}{d'}(X^2 - X^1) \end{aligned} \qquad (4)$$

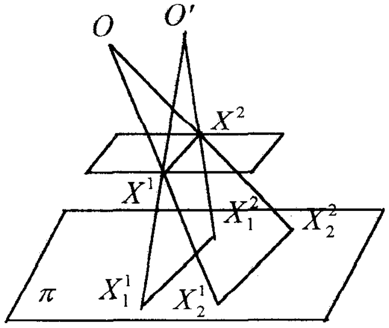


Fig. 1

It is evident, that both projective lines are parallel to each other. The advantage of introducing projective point and projective line is that images $\psi_1$ and $\psi_2$ can be imagined to be two images of respective projective points and lines on the same plane π captured from two different view points, and the effect of any image transformation can be explained based on the relation of corresponding projective point pair and projective line pair.

Considering all the notations and definitions mentioned above we are ready to prove Theorem 1 and 2. As only the figures on PP are concerned, in the following only the projective line related to PP is addressed.

**Proof of Theorem** 1: Construct a homography transformation $A_1$, based on four arbitrarily selected real point correspondences on plane π, and use it to transfer image $\psi_2$ to a new image $\psi'_2$, so that the image point of any real point of π in $\psi'_2$ be in the same position with its correspondence in image $\psi_1$. The image of any corresponding projective point pair on π, however, will not be registered with each other in both images, and projective line correspondences in $\psi_1$ and $\psi'_2$ are usually not parallel to each other either. Since each projective line correspondence is a pair of parallel line on plane π, their intersection points in images will be the vanishing point while $\psi_1$ and $\psi'_2$ are registered together. Therefore a vanishing line can be determined by such two different vanishing points. By means of the vanishing line another homography transform $A_2$ can be designed to transform both images $\psi_1$ and $\psi'_2$ to image $\psi'_1$ and $\psi''_2$ respectively, so that every projective line in one image will be parallel to its correspondence in the other image. One implementation method of the second homography transformation can be described as follows:

Choose one point P in space out of the image plane as the projection center; construct a plane PL parallel to a plane determined by the point P and the vanishing line; project image $\psi_1$ and image $\psi'_2$ separately into that plane PL with the rays emanated from the projection center P, and get new images $\psi'_1$ and $\psi''_2$ respectively.
□

From the above proof it can be seen that as any space position out of image plane can be chosen as the projection center, the reprojection transform is not unique. If it is required that the shapes of all the horizontal figures be recovered in reprojected images, at least one camera ( in our case, camera 1 ) should be calibrated, and the reprojection parameters of $A_2$ can be determined by the calibration procedure.

**Proof of Theorem** 2: It is obvious that the prerequisite condition for any pair-wise related straight line features, representing the same straight line segment in PP, to be with equal length in the reprojected stereo image pair is that its corresponding projective lines should be with the same length. If a straight line segment is represented by its two end points $X^1$ and $X^2$, it can be seen that from equation (4) that.

$$\begin{aligned}\left(X_1^1-X_1^2\right)-\left(X_2^1-X_2^2\right)&=\left(\frac{d_\pi}{d}-\frac{d'_\pi}{d'}\right)\left(X^1-X^2\right)\\&=\left(\frac{d_\pi}{d}-\frac{d_\pi-n^TO'}{d-n^TO'}\right)\left(X^1-X^2\right)\end{aligned}\qquad(5)$$

Evidently the necessary and sufficient condition should be $n^TO'=0$ □

## 2.2 The characteristics of reprojected stereo image

Since the purpose of introducing a reprojection transform into the stereo matching algorithm is that all the features in one image, representing horizontal figures in our case, present the same shape with their correspondences in the other one, from now on, only such kind of situation is addressed. In that case, the characteristics of transformed stereo images are:

1) All the feature correspondences, representing figures on the same horizontal physical plane, have the same disparity value. Therefore, disparity equality can be used to fulfill horizontal features' matching.

2) It is easy to see that from equation (3), (1) and $n^TO'=0$ that the distance of projective point pair of any point X is

$$X_2-X_1=\frac{d_\pi}{d}X'+O'-\frac{d_\pi}{d}X=\frac{d_\pi-d}{d}O'\qquad(6)$$

Therefore the epipolar lines in both reprojected images are parallel to each other. If calibrated images are used in the reprojection transformation, their direction is determined by the vector $O'$.

# 3. Shape similarity based stereo algorithm

## 3.1 Outline of the algorithm

The purpose of introducing shape similarity criterion into stereo matching algorithm is to ease the establishing of the correspondence of horizontal features. As figures in scenes, however, are usually the mixture of horizontal and non-horizontal ones, algorithm should not only take care of features with shape similarity but shape non-similar situation as well. Due to occlusion of objects, only the common visible part of a horizontal figure in both images presents similar shape, and therefore algorithm should concern the effect of occlusion as well. Considering all these observations a shape similarity based stereo matching algorithm is developed and its outline is as follows:

a. Taking stereo image pair and transforming them by a reprojection image transformation. In our case model building in Euclidean space is required and reprojection transformation based on calibrated image is used. Therefore the shape of every horizontal figure in scene is recovered in both images.

b. Extracting edge points in either pictures and then linking them to form two sets of edge strings with single pixel width.

c. Using edge string as an intermediate data structure in a generalized Hough transform strategy for feature matching which is based on criteria such as shape similarity, pixel local attribution similarity and etc.

d. Propagating matching to texture-free area.

In the following only terms b and c is addressed hereafter. Term d is not discussed in the paper.

### 3.2 Edge string — intermediate data structure

Edge points are extracted by using Sobel operation for convenience and followed by a peak value detection and linking procedure. The output of edge extraction and linking procedure will be single pixel-wide edge strings. The advantages of edge string are manifold. First, edge string is one kind of model-free data structure. In comparison with any analytical description model it can flexibly and faithfully deal with curves with various shape complexities. Second and more important is that edge string is used as an intermediate data structure. It means each edge string generated in the preprocessing stage does not necessarily correspond exactly to a physical figure in scene, and that it can be re-organized in the matching process to deal with a variety of situations. For example, if a horizontal figure can be seen in one image entirely, but partially seen in the other, as a result, only corresponding part of one string can be matched with the other. In this way matching horizontal and non-horizontal feature and effect of occlusion can be dealt with in a single matching process. Last but not the least is that, as edge string is just used as an intermediate structure, the matching results will not be sensitive to the performance of string segmentation.

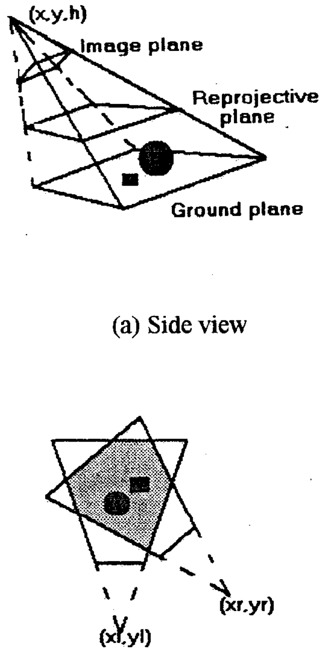
### 3.3 Generalized Hough transform strategy

The basic requirements for the matching strategy are

a. As each edge string may split to several substrings, or merge with other string during the matching process, split or merge function should be embedded in the matching strategy.

b. In the matching process several different criteria should be combined together to cope with respective different situations. Disparity equality test, for example, is used to perform matching between horizontal features pixel, and pixel attribution similarity and disparity continuity are used for non-horizontal feature to match its correspondence.

Keep above requirements in mind a generalized Hough transform strategy is adopted. In the strategy each string in one set is tested one after another by using a pixel voting methodology with a two-dimensional table. During testing the *i*th column of the table is used to store the candidate matching of *i*th pixel and the *j*th bin in the column will store a non-zero weight if a candidate matching with disparity value of *j* is found for the *i*th pixel. Each pixel in the string under testing searches for its candidate correspondence in other image within the area restricted by epipolar constraint and maximum disparity value. If the string corresponds to a horizontal figure in scene, most of its constituent pixels will share the same parallax value with their correspondences. By



(a) Side view

(b) Top view

Fig 2 Stereo setup

accumulating the number of contiguous bins with non-zero weight along each row, the row with large accumulated value will indicate that a horizontal feature correspondence has been established and its disparity value determined with the corresponding row number.

If disparity equality test fails, attribution similarity criterion is used for searching pixel-wise matching and disparity continuity criterion is used to enhance correct matching. Uniqueness and order consistence, usually used in stereo matching, are also used afterwards.

## 4. Experiments

The stereo setup we use is a pair of cameras as shown in Fig. 2. Each of the cameras is looking down at the ground with a proper title angle as shown in Fig. 2(a). Both of the cameras may take the same or different title angle, but with the same height H. The view from the top of the setup is shown in Fig. 2(b). In this figure $(X_l, Y_l, H)$ and $(X_r, Y_r, H)$ are the perspective centers of the corresponding cameras with respectively and two trapezia are drawn to indicate the sensible regions of the cameras.

The preliminary experiments have been conducted for the environment in our lab. Two sets of our experimental processes are shown in Fig. 3 and Fig. 4 respectively. In Fig. 3 a stack of two books, a cup and a piece of broken flagstone with a free-form contour are on the table. Fig. 3(a) and (b) are the original image pair. Fig. 3(c) and (d) are their corresponding reprojected images. Fig. 3(e) and (f) show the extracted edge segments after thinning and linking. It can be seen that most of the contours are extracted, but the cup brim and the book cover are not complete. Most of the contour line segments have been successfully matched to their correspondences in the initial matching process and some errors are refined in the refining procedure afterwards. Based on the parallax value estimates the reconstruction of the scene is shown in Fig. 3(f).

Fig. 4 is another scene with a tea can, a box, a 5" floppy diskette ,and two books. All the horizontal contours are correctly matched except some missing boundary lines. As the camera pair are positioned so far from each other that partial contour of the books can only be seen in one of the cameras and hence can not find their correspondences. In our experiments all of the significant horizontal segments have been properly matched, with precise disparity estimation in initial matching, only a few sub-lines have been missed (about 10%), all the errors can be refined in the refine process. For the purpose of comparison, we have also done the experiment on these pictures by using area-based point feature matching method, our method is about 50 times faster than the later with almost the same quality.

## 5. Discussion

In this paper a novel stereo vision approach has been introduced and discussed. In this approach a special designed image transformation, reprojection transformation, is adopted to recover figures' shapes on the horizontal planes so that the horizontal feature correspondence can be easily solved by taking advantages of shape similarity. By using edge point string, as an intermediate data structure, a generalized Hough transform strategy, can fuse criteria, such as shape similarity (disparity equality), disparity continuity and pixel attribution similarity and etc., to cope with correspondence of horizontal and non-horizontal features under the condition of object occlusion.

From the history of stereopsis research and our experience it can be seen that the applications of stereopsis methodology and its similar ones such as structure from motion is still a difficult topic in computer vision. The image preprocessing is still a bottleneck problem and the disparity propagation to the featureless area is also difficult if full automation is required. Recently various techniques have been suggested by using more than one techniques in combination to solve the problem. Structural stereopsis, for example, has been suggested[1] and a structural description are extracted from both images and used for feature matching afterwards. Model-based and single view understanding based paradigms have also been suggested[6][8]. In viewing of using model based stereopsis we argue that our method is specially suitable for model building in the indoor environment. Whenever the horizontal structures have been determined, the CAD models of the furniture and articles can be introduced immediately. From the state-of-art it seems to be an efficient way for such environmental model building.

**Acknowledge** The authors thank prof. R. Mohr of LIFIA-INRIA for his suggestion of using uncalibrated image in reprojection transformation.

**References**

[1] K.L. Boyer, A.C. Kak, "Structural Stereopsis for 3- D Vision ", *IEEE Trans. PAMI* Vol. 10, No. 2, pp. 144-166, 1988.

[2] O. D. Faugeras, "What can be Seen in Three Dimensions With an Uncalibrated Stereo Rig?" *Proc. of the European Conf. on Computer Vision*, pp. 321-334, Santa Margherrta Ligure. Italy, June, 1992.

[3] R. Hartley, "Projective Reconstruction and Invariants from Multiple Images" *IEEE Trans. PAMI*, Vol. 16, No. 10, pp. 1036-1040, 1994.

[4] R. Koch, "Automatic Reconstruction of Buildings from

Stereoscopic Image Sequences," *Computer Graphics Forum*, Vol. 12, No. 3, pp. C-340 - 350, 1993.

[5] W. S. Kim, “Graphical Operator Interface for Space Telerobotics”, *Proc. ICRA*, pp. 761-768, 1993.

[6]H. Maitre, W. Luo, "Using Models to Improve Stereo Reconstruction", *IEEE Trans. PAMI*, Vol. 14, No. 2, pp. 269-277, 1992.

[7]A. Shushua, “Projective Structure from Uncalibrated Images: Structure from Motion and Recognition”, *IEEE Trans. PAMI*, Vol. 16, No. 8, pp. 778-790, 1994.

[8] A. Ude, H. Brodw, R. Dillmamn, “Object Localization Using Perceptual Organization and Structural Stereopsis”, the *Proc. of the 3rd International Conf. on Automation, Robotics, and Computer Vision*, pp. 197-201, 1994.

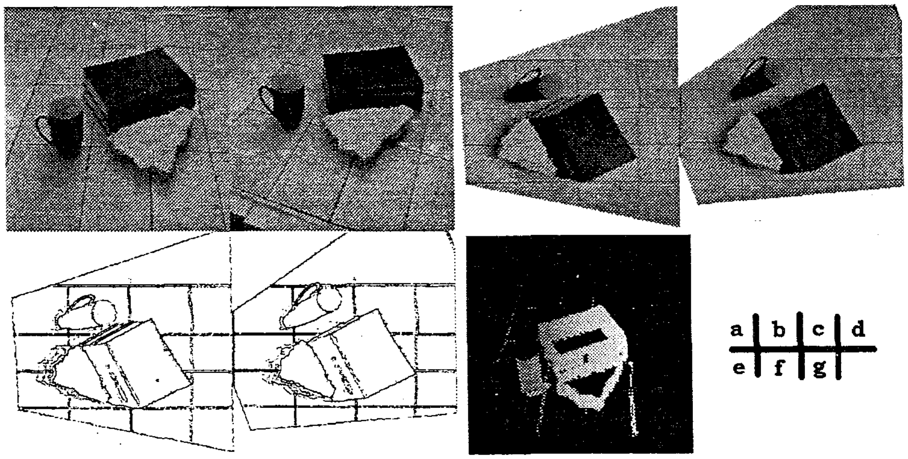


Fig. 3 Experiment 1

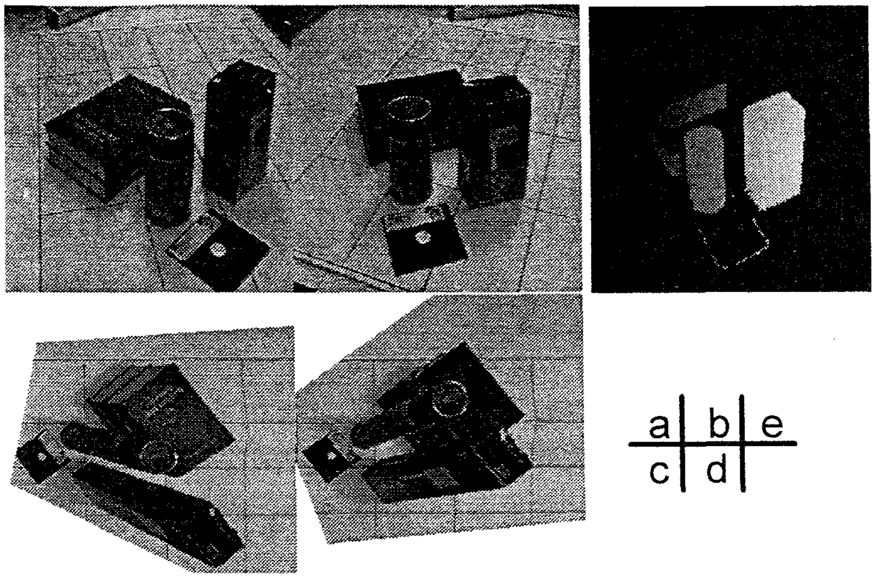


Fig. 4 Experiment 2