Computer Science Department of The University of Auckland
CITR at Tamaki Campus (http://www.citr.auckland.ac.nz)

CITR-TR-77 November 2000

Camera Calibration of Rectangular Textures

Jacky Baltes[1]

Abstract

This paper describes a practical method for the camera calibration given a single image of a regular texture. This paper uses the calibration of images of skyscrapers as an example. The paper introduces two algorithms for the assignment of realworld coordinates to feature points. The first algorithm selects five closely connected feature points and determines the orientation of the rectangular pattern. The second algorithm iteratively sorts the feature points and assigns real world coordinates to them. Lastly, the Tsai camera calibration algorithm is used to compute the camera parameters.

[1] CITR, Department of Computer Science,Tamaki Campus, University of Auckland, Private Bag 92019, Auckland, New Zealand

Robot Localisation Using an Omnidirectional Colour Image

David C.K. Yuen* and Bruce A. MacDonald

The Department of Electrical and Electronic Engineering,
The University of Auckland, Private Bag 92019, Auckland, New Zealand.
d.yuen@auckland.ac.nz,b.macdonald@auckland.ac.nz

Abstract. We describe a vision-based indoor mobile robot localisation algorithm that does not require historical position estimates. The method assumes the presence of an *a priori* map and a reference omnidirectional view of the workspace. The current omnidirectional image of the environment is captured whenever the robot needs to relocalise. A modified hue profile is generated for each of the incoming images and compared with that of the reference image to find the correspondence. The current position of the robot can then be determined using triangulation as both the reference position and the map of the workspace are available. The method was tested by mounting the camera system at a number of random positions positions in a 11.0m × 8.5 m room. The average localisation error was 0.45 m. No mismatch of features between the reference and incoming image was found amongst the testing cases.

1 Introduction

Under the traditional deliberative motion control architecture, a robot needs to know its own position in the environment before making a navigation plan. If the robot is first switched on or wants to re-position itself after getting lost, no reliable previous position estimates will be available for the localisation stage. Many common localisation methods, notably dead-reckoning using extended Kalman filtering [4], cannot cope with such a condition.

In this paper, we describe a passive, vision-based localisation technique that does not involve the use of historical position estimates, and takes advantage of the richer information in an image. An omnidirectional imaging system is introduced to provide colour and textual information to the system. The distinctive features from an incoming image are extracted using a region segmentation method. The extracted features are then matched with those from a reference image to generate matched landmarks. The placement of artificial landmarks in the environment is unnecessary.

In section 2, we review previous work in vision-based localisation methods that do not require historical position estimates. Section 3 outlines our localisation approach. It also describes the image segmentation and triangulation

* This work was supported in part by the Foundation for Research, Science and Technology, New Zealand, with a Top Achiever Doctoral Scholarship.

techniques adopted in the system. The test results are discussed in section 4 before summarising the paper in section 5.

2 Maps and Landmarks

Map matching can usually be carried out without the use of an image. A local map is first generated for the area around the robot, using the measurements from a laser or ultrasonic range finder [2, 7]. The local map is then matched against different regions of a global map, at different orientations. Since the map matching uses a local distance map, the localisation process can be confounded if objects with similar shapes are present in the environment. Also, the correlation operation requires considerable computation.

Many industrial robots are guided by bar codes [5], reflective tape [3, p313–317], ceiling light patterns [3, p472–477] or other artificial landmarks. A global positioning system (GPS) is a notable example of an artificial emitter in an outdoor navigation. While the landmark recognition step is usually quite simple, the cost of laying out and maintaining the well calibrated landmarks can be very expensive, and impractical in some environments.

Visual images usually have high spatial resolution and can provide details such as the colour and texture of the object being observed. With the extra information provided by visual sensors, the robot can have a better understanding of the complex surroundings. In many cases, natural landmarks can be extracted from the incoming images.

Using the concept of a "view field" [1], tiny visual features may be extracted from an image together with their relative spatial relations, to form a landmark. The memory requirement for the storage of typical indoor scenes is thus reduced to about 16000 bytes per m^2. Both Lin and Zhang [6, 8] process the sparsely sampled omnidirectional image with neural networks to extract landmarks for localisation, in which 120 and 1600 bytes were retained respectively for each image frame. While the storage of these landmarks requires only modest amount of memory, the image capturing stage involves a lot of preparative work and makes the localisation system quite inflexible.

3 Vision-Based Localisation System

Algorithm 1 shows the overall process of localisation. Our method assumes an *a priori* map for the environment. An omnidirectional image is used to simplify camera motion; panning control is not required.

To locate the robot, a vertically central strip of an omnidirectional image is segmented into regions by analysing the horizontal hue profile, then matched against region boundaries in a reference image, and triangulation is used to calculate the new robot position.

The imaging system comprises two Sony EVI-D31 cameras and two OMT SEQ-P1S frame grabber cards with a Pentium based controller, to be mounted on

Algorithm 1 Localise

On first invocation, call Initialise()
CurrentImage = ObtainImage()
Create all tokens of 3 consecutive region MHI median values for ReferenceImage
Create all tokens of 3 consecutive region MHI median values for CurrentImage
Find longest token match between ReferenceImage and CurrentImage
for each of the first, middle and last matching boundary pairs: **do**
    Triangulate position from the map position of the boundary pair
end for
return the average of the three position estimates

ObtainImage:

Take 8 images at 45^o increments, link them together to one image
Extract the 30 pixel high central strip
Calculate the MHI for each pixel in the strip
for each 10-pixel wide band **do**
    Calculate the band MHI median
end for
Find region boundaries by differentiating the band median sequence
for each region between boundaries **do**
    calculate the region MHI median
end for
return the sequence of region MHI values

Initialise:

ReferenceImage = ObtainImage()
Load the environment map
Calculate the map positions of boundaries in ReferenceImage

our mobile robot as a multi-purpose flexible vision system. To ensure controllable images for testing the current development stage, a single camera is mounted on a tripod. The images captured for this study have a resolution of 320×240 pixels and a colour depth of 24 bits. To facilitate comparing results, the zoom control of the camera was adjusted for a view angle of 45^o (horizontal) $\times$ 34^o (vertical) at 84cm above the floor. At each location, 8 images were taken in 45^o increments. At present the camera head should face the same direction when taking the first image amongst each series; the purpose is to discover the robot position and later we expect to remove this constraint and also discover the orientation. The 8 images were linked together to form a panoramic view of the environment, shown in Figure 1. A horizontal strip of 2560×30 pixels is then cut from the center of the omnidirectional image and used for the rest of the processing.

The representation of the image may be further simplified by extracting the hue channel of an HSV model. For humans, colour discontinuity often represents separation between objects. While the hue channel is relatively immune to variations in illumination, some hue values have little meaning and are sensitive to minor changes, notable values near white, gray and black. The modified hue index (MHI) is then defined:

Fig. 1. Omnidirectional view of the workspace: a) the original panoramic image. b) The horizontal strip cut from original view, which is marked by the white box shown in image a). (The view shown in b) has been stretched vertically for better display.)

$$MHI = \begin{cases} -2/3 * \pi & S >= 0.15 \text{ and } V >= 10 & \text{(black)} \\ -1/4 * \pi & S < 0.15 \quad \text{and } V >= 90 & \text{(gray)} \\ -1/3 * \pi & S < 0.15 \quad \text{and } V > 90 & \text{(white)} \\ H & \text{otherwise} & \text{(other colours)} \end{cases} \tag{1}$$

where H,S,V represents the hue $[0, 2\pi)$, saturation $[0, 1]$ and value $[0, 100]$.

The image is divided into 10-pixel wide vertical bands and the median MHI is computed for each band. Most of the smaller uncharted objects, e.g. network cable ducts, electric switches etc, are removed by band median filtering.

When viewing a large object, we may find regions with relatively constant values in the MHI band median profile, as illustrated in Figure 2. The regional boundaries may represent object edges or distinctive changes in the surface features of objects. We can locate potential regional boundary lines by thresholding the differentiated MHI band median profile. To facilitate the later matching operation, a "region median" is calculated for each detected region by calculating the median MHI of all the bands within the region boundaries.

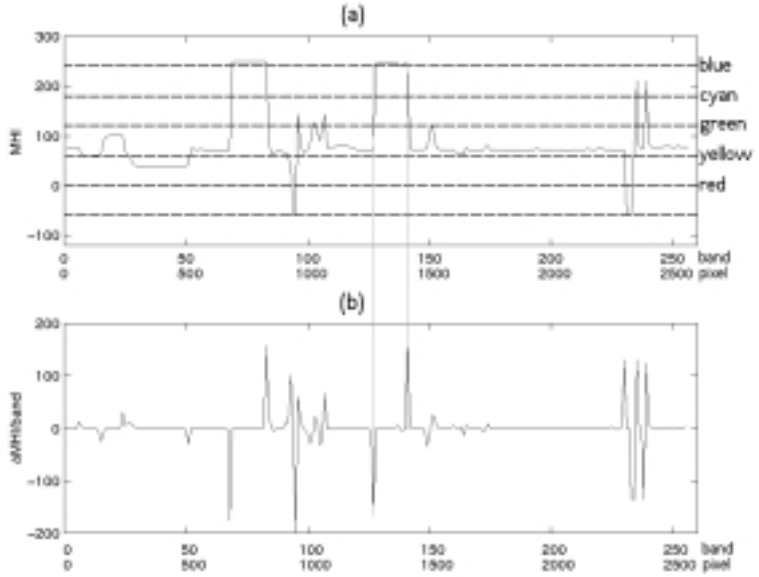


Fig. 2. (a) The modified hue index profile. (b) the differentiation of the MHI profile.

3.1 Preparations for the Map and Reference Image

Since the band median filtering method removes minor features, the level of detail required in the map is not high, and maps should not be difficult to maintain. The complexity of the environment determines the minimum number of reference images that needs to be taken. If the visibility of different parts of the workspace to the reference point is blocked, more reference points are required. In this study, a simpler environment was considered where only one reference point was sufficient. The exact position of the reference point was determined by surveying before taking the first image.

The viewing angle from the reference point to the edges of the large objects can be calculated from the coordinates of the regional boundaries on the omnidirectional image. The map position of these objects can then be estimated by extending the line-of-sight at the given viewing angle until an intersection is formed on the map, as depicted in Figure 3.

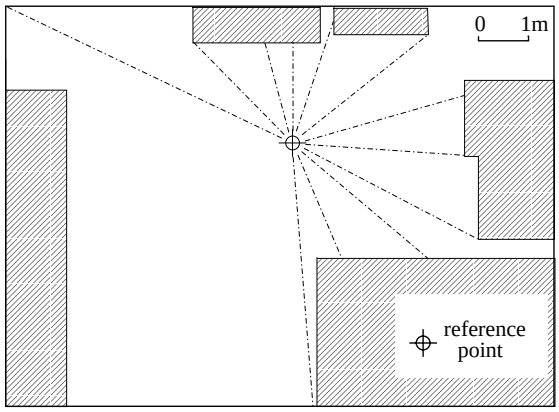


Fig. 3. Mapping of the observed feature for the reference image. The map position of an observed feature can be found by extending the line-of-sight at the given viewing angle until an intersection is formed.

3.2 Localisation System

An omnidirectional snapshot of the environment is taken whenever the robot needs to re-locate itself, and the MHI is calculated to identify regions. Since the positions of large objects are known, the current position of the robot can be identified using triangulation once enough matches have been established between boundary lines in the reference and current images, that represent features in the map.

The feature matching process is crucial to the performance of the localisation stage. When the robot moves to different parts of the room, the relative size of the regions on the MHI profile may change. Some features may become too small

and be left unaccounted for. Due to the presence of uncharted objects, some unexpected features may appear while some expected ones may be occluded. Also changes in reflectance of object surfaces may appear as features after MHI processing. The proposed matching algorithm should be tolerant to these defects.

Omnidirectional images have the important property that the *sequence* of modified hue regions remains the same, providing all the objects are still visible to the observer. A sequence of triples is formed for the reference image by grouping the region median values of three consecutive regions (that is for regions $\{(1,2,3),(2,3,4),(3,4,5),\ldots\}$) into "tokens." The list of region median values for the current image is then searched to locate the possible matches for each of the reference tokens. A match is declared if the region medians for each of the three consecutive regions of current image are within a certain tolerance from the respected regions of the reference token. The tolerance level was set to $\frac{5}{36}\pi$ radians in this study. Ideally, we can obtain a token sequence match from the incoming image that contains as many regions as the reference. In practice the longest token is taken as the best match.

The location and orientation of the robot (x, y, ϕ) can be found by solving the following non-linear simultaneous equations:

$$\tan(2 * \pi - \phi - \theta_i) = \frac{y_i - y}{x_i - x} \tag{2}$$

where x_i, y_i, represent the x, y coordinates of the i^{th} object edge on the map, and θ_i represents the observed angle of the i^{th} object edge from the robot. See Figure 4 for further explanation.

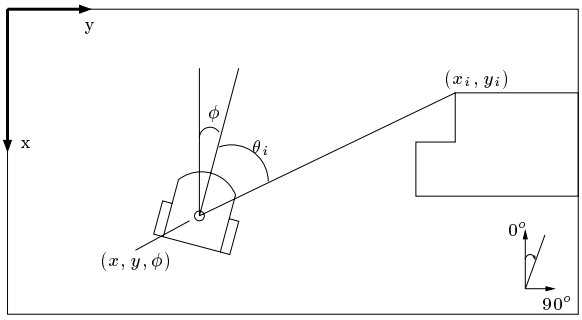


Fig. 4. Geometric conventions.

In this study, the camera head was aligned to a fixed direction before taking the first image. The localisation module thus needs to solve for only the two position variables (x, y), So a minimum of two matched features are required.

As an initial investigation, the average is taken of three sets of position estimates, which are generated by taking the observed angles of the first, last and the middle regional boundaries of the longest token match from equation 2.

4 Results and Discussions

The vision-based localisation method was tested in an 11.0m × 8.5m laboratory. As shown in Figure 5, nine random testing positions were generated. The test results are shown in Table 1. The average localisation is 0.45 m with a standard deviation of 0.22 m. No mismatch was found between the reference and current image when examined the longest token match for each testing case.

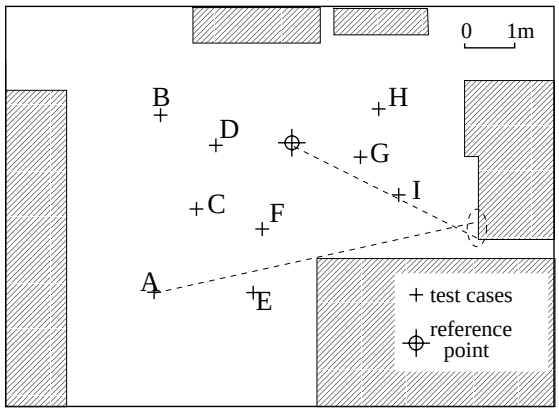


Fig. 5. The testing environment for the localisation algorithm. The influence of partial occlusion is demonstrated. The dotted lines from location A and the reference point represent their line-of-sight when supposedly viewing the same edge of an object. Due to partial occlusion, the robot at location A is not really the true edge and thus leads to a large localisation error.

Table 1. Localisation error of the testing cases

Position	A	B	C	D	E	F	G	H	I
x-coordinate (m)	5.57	2.11	3.93	2.70	5.58	4.53	2.95	1.98	3.69
y-coordinate (m)	2.98	3.10	3.82	4.14	4.91	5.13	7.02	7.43	7.79
localisation error (m)	0.66	0.91	0.50	0.38	0.25	0.45	0.23	0.27	0.44

Although the proposed method may not be accurate enough for the use in a standalone localisation system, that does not poise a serious problem. In this study, we intend to develop a vision-based localisation system that does not depend on the historical position estimates. In this way, the relative rough position estimates can be refined using more established localisation methods, such as extended Kalman filtering.

The test samples that give large localisation error are located far away from the reference point. The view can be quite different from that captured at the reference point. For example, only a fraction of the partition can be visualised at location A. As a result, the observed boundary at location A is not really the true edge of the partition (circled with dots in Figure 5) and thus leads to a large error.

In the current system, the robot position was calculated using only three of the matched features with the rest being discarded. These other matches could potentially be used to improve the accuracy and robustness of the technique. In addition, range sensors can be introduced to the system to reduce the ambiguities arisen during various stage of the operation.

5 Conclusion

A vision-based robot localisation system is proposed that does not involve the use of historical position estimates. A modified hue profile is generated for each of the incoming omnidirectional images. The extracted hue regions are matched with that of the reference image to find corresponding region boundaries. As the reference image, exact location of the reference point and the map of the workspace are available, the current position of the robot can be determined by triangulation.

The method was tested by placing the camera set-up at a number of different random positions in a 11.0m $\times$ 8.5m room. The average localisation error was 0.45 m. No mismatch of features between the reference and incoming image was found. While the proposed localisation method may not be sufficiently accurate if used alone, it provides a good initial position estimate for the use of other more established localisation methods, such as extended Kalman filtering.

References

1. Christian Balkenius. Spatial learning with perceptually grounded representations. *Robotics and Autonomous Systems*, 25:165–175, 1998.
2. Alberto Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):249–265, June 1987.
3. H.R. Everett. *Sensors for mobile robots: theory and application.* A.K. Peters Ltd., 1995.
4. Leopoldo Jetto, Sauro Longhi, and Giuseppe Venturini. Development and experimental validation of an adaptive extended kalman filter for the localization of mobile robots. *IEEE Transactions on Robotics and Automation*, 15(2):219–229, 1999.
5. John J. Leonard and Hugh F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(3):376–382, 1991.
6. Long-Ji Lin, Thomas R. Hancock, and J. Stephen Judd. A robust landmark-based system for vehicle location using low-bandwidth vision. *Robotics and Autonomous Systems*, 25:19–32, 1998.
7. Brian Yamauchi. Mobile robot localization in dynamic environment using dead reckoning and evidence grids. In *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, pages 1401–1406, Minneapolis, Minnesota, Apr 1996. IEEE.
8. Jianwei Zhang, Alois Knoll, and Volkmar Schwert. Situated neuro-fuzzy control for vision-based robot localisation. *Robotics and Autonomous Systems*, 28:71–82, 1999.

Binocular Stereo by Maximising the Likelihood Ratio Relative to a Random Terrain

Georgy Gimel'farb

CITR, Department of Computer Science
Tamaki Campus, University of Auckland
Private Bag 92019, Auckland, New Zealand
g.gimelfarb@auckland.ac.nz
WWW home page: http://www.tcs.auckland.ac.nz/~georgy/

Abstract. A novel approach to computational binocular stereo based on the Neyman–Pearson criterion for discriminating between statistical hypotheses is proposed. An epipolar terrain profile is reconstructed by maximising its likelihood ratio with respect to a purely random profile. A simple generative Markov-chain model of an image-driven profile that extends the model of a random profile is introduced. The extended model relates transition probabilities for binocularly and monocularly visible points along the profile to grey level differences between corresponding pixels in mutually adapted stereo images. This allows for regularising the ill-posed stereo problem with respect to partial occlusions.

1 Introduction

Computational binocular stereo that reconstructs 3D terrains from stereo pairs of images is an ill-posed inverse photometric problem because a rich variety of different optical surfaces can produce the same stereo pair [11, 12]. The ill-posedness is caused mainly by partial occlusions hindering stereo observation of some terrain points and by uniform or repetitive colouring of the surface. To partially regularise the problem, stereo images have to be matched wuth due regard to binocular and monocular visibility of terrain points.

Most of the known stereo matching algorithms (see, for instance, [1, 2, 8, 9, 13]) state and solve the stereo problem as a statistical problem of estimating a hidden Markov model of an epipolar terrain profile. The prior profile model is combined with the conditional model of stereo images, given the profile, to derive the posterior model and use it for measuring similarity between the stereo images for each possible profile. Then the reconstruction is conducted by maximising the similarity between the images. In many cases the similarity is measured with no explicit account of possible partial occlusions.

Symmetric Dynamic Programming Stereo (SDPS) discussed in [3, 6] follows the same scheme but allows for discriminating between the binocularly (BVP) and only monocularly visible points (MVP) along the profile during the reconstruction. All variants of an epipolar profile are represented by continuous paths in a specific graph of profile variants (GPV). Each GPV-node has three states

specifying whether it represents the BVP yielding two corresponding pixels in a stereo pair or the MVP depicted only in the left or in the right stereo image. The allowable transitions between the successive GPV-nodes along the profile depend on their visibility states. In this case the similarity for the BVPs is obtained by comparing the corresponding pixels but some heuristic weights for the MVPs have to be involved to search for a profile yielding the best similarity between the images [4].

This paper proposes another approach that is based on the Neyman–Pearson criterion [10] and involves explicit Markov models of an epipolar profile: the reconstructed profile has to maximise the likelihood ratio with respect to a purely random one. The Markov-chain model of a random epipolar profile introduced in [4,5] is extended below by relating the transition probabilities for each GPV-node to grey values in the corresponding pixels of a stereo pair. Transitions to the GPV-nodes representing BVPs depend on grey level deviations between the mutually adapted stereo images, the adaptation tending to reduce relative photometric distortions of the binocularly visible corresponding parts of the images. Each BVP-transition specifies also the probabilities of transitions to the adjacent nodes representing MVPs so that all the transition probabilities in the GPV are related to the images. Thus the proposed model allows for a partial probabilistic regularisation of terrain reconstruction such that the transition probabilities for purely random profiles constitute the regularising parameters [7].

The paper is organised as follows. Section 2 considers the profile reconstruction based on maximising the likelihood ratio. The extended probabilistic model of an epipolar profile is presented in Section 3. Experimental results and conclusions are given in Section 4.

2 Profile Reconstruction Using Log-Likelihood Ratio

Let x denote the x-coordinate of the GPV-node, p be the integer x-parallax, or disparity between the corresponding pixels with integer coordinates x_{L} and x_{R} in stereo images, and $s \in \{\mathrm{B}, \mathrm{ML}, \mathrm{MR}\}$ be the visibility state indicating the BVP or MVP visible only in the left or right image, respectively. It holds for the symmetric stereo geometry [3,6] that $x = (x_{\mathrm{L}}+x_{\mathrm{R}})/2$, $p = x_{\mathrm{L}}-x_{\mathrm{R}}$, $x_{\mathrm{L}} = x+p/2$, and $x_{\mathrm{R}} = x - p/2$.

Figure 1 shows a fragment of the GPV, each GPV-node (x, p, s) having seven admissible transitions to the next three nodes $(x+0.5, p-1, s')$, $(x+0.5, p+1, s')$, and $(x+1, p, s')$. According to the explicit generative Markov model of the profile variants, every epipolar profile $\mathbf{p} = [(x_i, p_i) : \ i = 1, \ldots, n]$ with n GPV-nodes is generated as a Markov chain of $n-1$ successive admissible transitions from each current GPV-node (x_i, p_i, s_i) to the next one $(x_{i+1}, p_{i+1}, s_{i+1})$; $i = 1, \ldots, n-1$.

Let g_{L} and g_{R} be the left and right images of a stereo pair. Let $\Pr(\mathbf{p}|g_{\mathrm{L}}, g_{\mathrm{R}})$ and $\Pr(\mathbf{p})$ specify the probability distributions of profiles in the GPV under two simple statistical hypotheses: an "image-driven" or a purely random profile, respectively. Thus the profile reconstruction can be based on the Neyman–Pearson criterion [10] of choosing the first hypothesis by comparing to a par-

ticular threshold Θ the likelihood ratio or, what is the same, the log-likelihood ratio $L(\mathbf{p}|g_{\mathrm{L}}, g_{\mathrm{R}})$ of an image-driven profile with respect to a random profile:

$$L(\mathbf{p}|g_{\mathrm{L}}, g_{\mathrm{R}}) = \log \Pr(\mathbf{p}|g_{\mathrm{L}}, g_{\mathrm{R}}) - \log \Pr(\mathbf{p}) \geq \Theta. \tag{1}$$

The most adequate profile in the GPV can be chosen by maximising the log-likelihood ratio in Eq. (1):

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} L(\mathbf{p}|g_{\mathrm{L}}, g_{\mathrm{R}}). \tag{2}$$

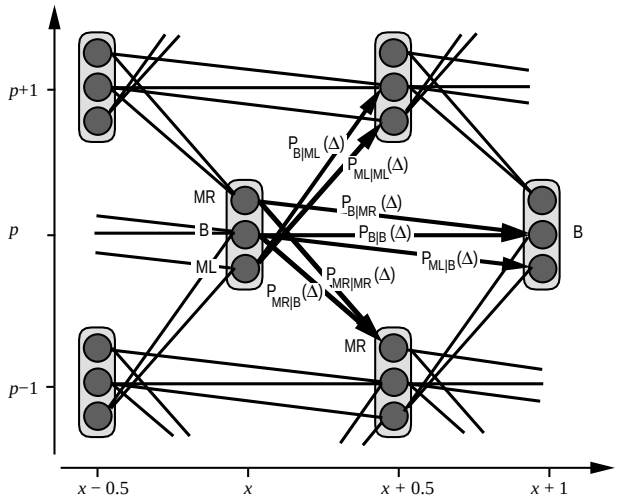


Fig. 1. Transition probabilities for the GPV.

Because the similarity between stereo images is given in terms of the likelihood ratio, this approach allows for comparing profiles of different length and solving Eq. (2) with sequential decision rules to accelerate the reconstruction. In this latter case some profile variants can be rejected after comparing their initial parts of length $m < n$ if the likelihood ratio becomes lower than a particular threshold that depends generally on m.

The SDPS algorithm in [3, 6] is easily modified to implement the reconstruction process of Eq. (2) if both the probability distributions $\Pr(\mathbf{p}|g_{\mathrm{L}}, g_{\mathrm{R}})$ and $\Pr(\mathbf{p})$ in Eq. (1) are represented by the products of transition probabilities specified by particular explicit Markov-chain models.

3 Markov-Chain Models of a Profile

A purely random profile is described in [4, 5] as a stationary Markov chain of the GPV-nodes with the transition probabilities $P_{s'|s}$ that depend only on the visibility states. Both the monocular cases ML and MR are equivalent by symmetry

and can be denoted as M:

$$\begin{array}{ll} P_{\mathrm{ML|ML}} = P_{\mathrm{MR|MR}} & \equiv P_{\mathrm{M|M}}; \\ P_{\mathrm{B|ML}} = P_{\mathrm{B|MR}} & \equiv P_{\mathrm{B|M}} = 1 - P_{\mathrm{M|M}}; \\ P_{\mathrm{ML|B}} = P_{\mathrm{MR|B}} & \equiv P_{\mathrm{M|B}} = 0.5\,(1 - P_{\mathrm{B|B}})\,. \end{array} \tag{3}$$

Thus the stationary Markov chain producing purely random profiles is specified by the two transition probabilities $P_{\mathrm{B|B}}$ and $P_{\mathrm{M|M}}$.

For the image-driven Markov chain generating the actual terrain profiles, stereo images are assumed to be mutually adapted along each profile. The adaptation is performed within a given range $E = [2 - e_{\max}, e_{\max}]$; $1 \le e_{\max} < 2$, of admissible ratios between grey level increments for the corresponding successive BVPs in both stereo images (see [3,6] for more detail). It has a goal of excluding or reducing relative photometric image distortions. Then the probability of transition from a current GPV-node (x, p, s) to the next node representing a BVP (x', p', B) can be related to the residual grey level difference $\Delta_{x',p'}$ between the corresponding points x'_{L} and x'_{R} in the adapted images for each profile under consideration.

The transition probabilities $\Pr_{s'|s}(\Delta_{x',p'}) \equiv \Pr_{s'|s}(\Delta)$ satisfy the obvious conditions of Eq. (3), the two last MVP-transitions being equalised from the same considerations of symmetry:

$$\begin{array}{ll} \Pr_{\mathrm{ML|ML}}(\Delta_{x+0.5,p+1}) & = 1 - \Pr_{\mathrm{B|ML}}(\Delta_{x+0.5,p+1}); \\ \Pr_{\mathrm{MR|MR}}(\Delta_{x+0.5,p-1}) & = 1 - \Pr_{\mathrm{B|MR}}(\Delta_{x+1,p}); \\ \Pr_{\mathrm{MR|B}}(\Delta_{x+0.5,p-1}) & = \Pr_{\mathrm{ML|B}}(\Delta_{x+1,p}) \\ & = 0.5 \cdot \left(1 - \Pr_{\mathrm{B|B}}(\Delta_{x+1,p})\right)\,. \end{array} \tag{4}$$

The log-likelihood ratio $l(x_i, p_i, s_i; x_{i-1}, p_{i-1}, s_{i-1} | g_{\mathrm{L}}, g_{\mathrm{R}})$ for the i-th transition along a profile $\mathbf{p}$ combines the transition probabilities of Eq. (4) depending on a given stereo pair $(g_{\mathrm{L}}, g_{\mathrm{R}})$ with the probabilities of Eq. (3) that specify a purely random profile:

$$\begin{array}{l} l(x_i, p_i, s_i; x_{i-1}, p_{i-1}, s_{i-1} | g_{\mathrm{L}}, g_{\mathrm{R}}) = \\ \qquad \log \Pr_{s_i|s_{i-1}}(\Delta_{x_i,p_i}) - \log P_{s_i|s_{i-1}}\,. \end{array} \tag{5}$$

Thus the log-likelihood ratio for the total profile $\mathbf{p}$ is an additive functional with respect to $\mathbf{p}$

$$L(\mathbf{p} | g_{\mathrm{L}}, g_{\mathrm{R}}) = \sum_{i=1}^{n} l(x_i, p_i, s_i; x_{i-1}, p_{i-1}, s_{i-1} | g_{\mathrm{L}}, g_{\mathrm{R}}) \tag{6}$$

that can be maximised by dynamic programming techniques.

Experiments in Section 4 show that the probabilities $P_{\mathrm{M|M}}$ and $P_{\mathrm{B|B}}$ of Eq. (3) can be considered in Eq. (2) as regularising parameters that define smoothness and visual quality of the reconstructed terrain. The final reconstruction results depend also on the image adaptation range E and on the chosen probability function $\Pr_{s'|s}(\Delta)$ in Eq. (4).

4 Experimental Results and Concluding Remarks

Several digital x-parallax maps (DPM) consisting of the epipolar terrain profiles reconstructed from the artificial stereo pair "Corridor" and the natural stereo pair "Pentagon" in Figure 2 are presented in Figure 3. The "ground truth" with the total disparity range [0,10] in Figure 2, (c), allows for checking the actual quality of reconstruction of the "Corridor" scene.

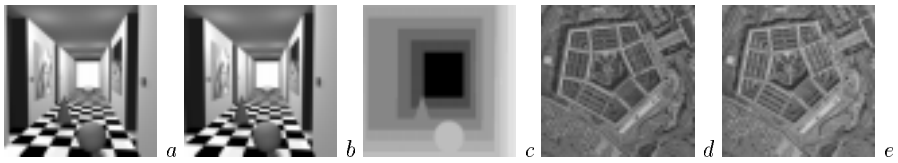


Fig. 2. Artificial stereo pair 256×256 "Corridor" (a, b) with the known "ground truth" in terms of ideal integer disparities shown in the range image (c) and the natural stereo pair 512×512 "Pentagon" (d, e).

These and other experiments (see, for instance, [7]) show that the reconstructed DPMs are in close agreement with visual perception within a sufficiently wide "triangular" domain of the regularising parameter space. But it is worth noting that the orthoimages of these scenes, formed from the stereo pairs in line with the DPMs, are quite similar over almost the total parameter space. Thus even visually unacceptable solutions yield close similarity between the images as could be expected from the ill-posedness of the problem.

The reconstruction results are rather similar for the following three different probability functions with $\sigma = 5, \ldots, 15$:

$$\Pr_{\mathrm{B}|s}(\Delta) = 0.998 \exp\left(-\frac{\Delta^2}{\sigma^2}\right) + 0.001; \tag{7}$$

$$\Pr_{\mathrm{B}|s}(\Delta) = 0.998 \exp\left(-\frac{|\Delta|}{\sigma}\right) + 0.001; \tag{8}$$

$$\Pr_{\mathrm{B}|s}(\Delta) = 0.998 \exp\left(-\sqrt{\frac{|\Delta|}{\sigma}}\right) + 0.001, \tag{9}$$

and for the different adaptation ranges $E = [0.8, 1.2] \ldots [0.4, 1.6]$ although the reconstructed terrains become smoother for the larger values of σ and larger adaptation ranges. The "Corridor" and "Pentagon" scenes in Figure 3 are reconstructed using the same image adaptation range $E = [0.8, 1.2]$ but different transition probabilities of Eq. (8) and Eq. (7), respectively, with $\sigma = 10$.

In principle, the functions $\Pr_{\mathrm{B}|s}(\Delta)$ can be empirically estimated using training samples of the epipolar stereo pairs with known terrain models. But the experiments show that reconstruction results depend much more on the regularising parameters.

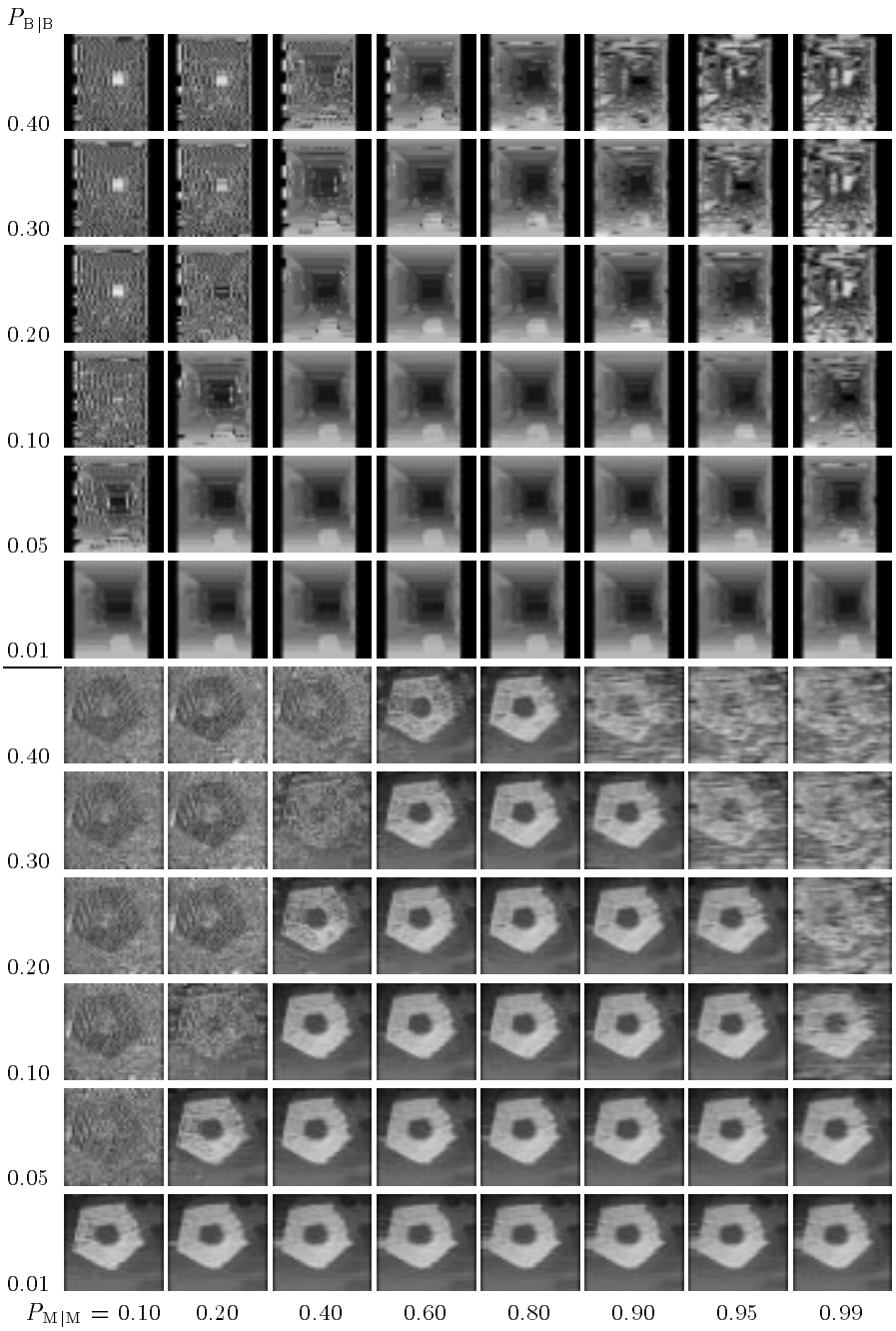


Fig. 3. Range images of the reconstructed scenes "Corridor" and "Pentagon".

Table 1. Mean absolute differences (m.a.d.) and standard deviations (st.d.) of the reconstructed "Corridor" from the ground truth in Figure 2, (c).

Transition probabilities of Eq. (8) with $\sigma = 5$									
	$P_{\mathrm{M\,|M}}$	0.10	0.20	0.40	0.60	0.80	0.90	0.95	0.99
$P_{\mathrm{B\,|B}} = 0.40$	m.a.d.	2.99	2.93	2.07	1.26	1.03	1.41	1.91	1.97
	st.d.	1.96	2.05	2.12	1.74	1.35	1.49	1.64	1.61
0.30	m.a.d.	2.97	2.89	1.50	0.92	0.86	0.97	1.48	1.95
	st.d.	1.99	2.08	1.94	1.47	1.29	1.28	1.51	1.61
0.20	m.a.d.	2.96	2.54	0.97	0.68	0.65	0.76	0.91	1.88
	st.d.	2.01	2.13	1.60	1.25	1.10	1.14	1.22	1.61
0.10	m.a.d.	2.87	1.27	0.56	0.46	0.44	0.48	0.56	1.08
	st.d.	2.08	1.84	1.06	0.85	0.44	0.48	0.56	1.08
0.05	m.a.d.	1.88	0.65	0.43	0.38	0.38	0.37	0.40	0.70
	st.d.	2.09	1.19	0.79	0.67	0.65	0.62	0.66	1.02
0.01	m.a.d.	0.48	0.37	0.35	0.35	0.34	0.34	0.34	0.37
	st.d.	0.90	0.62	0.54	0.55	0.53	0.54	0.53	0.55

Transition probabilities of Eq. (8) with $\sigma = 10$									
	$P_{\mathrm{M\,|M}}$	0.10	0.20	0.40	0.60	0.80	0.90	0.95	0.99
$P_{\mathrm{B\,|B}} = 0.40$	m.a.d.	3.02	2.98	1.79	0.80	0.74	1.42	2.08	2.22
	st.d.	2.05	2.09	2.04	1.39	1.08	1.51	1.65	1.71
0.30	m.a.d.	3.02	2.87	0.94	0.48	0.54	0.75	1.52	2.20
	st.d.	2.05	2.12	1.58	0.91	0.87	1.06	1.53	1.71
0.20	m.a.d.	3.01	2.42	0.51	0.36	0.39	0.49	0.71	2.16
	st.d.	2.08	2.12	1.00	0.63	0.66	0.76	1.01	1.72
0.10	m.a.d.	2.83	0.75	0.37	0.36	0.35	0.36	0.41	0.98
	st.d.	2.12	1.40	0.62	0.55	0.57	0.57	0.62	1.23
0.05	m.a.d.	1.50	0.39	0.35	0.35	0.35	0.34	0.35	0.51
	st.d.	1.90	0.66	0.54	0.54	0.54	0.53	0.55	0.78
0.01	m.a.d.	0.35	0.35	0.34	0.34	0.34	0.33	0.34	0.35
	st.d.	0.54	0.53	0.54	0.54	0.53	0.53	0.53	0.53

Table 1 shows the mean absolute differences and standard deviations between the reconstructed digital x-parallax models of the "Corridor" scene and the ground truth of Figure 2,c for the transition probabilities of Eq. (8) with $\sigma = 5$ and $\sigma = 10$. The latter reconstruction results are shown in Figure 3. This scene has the dominant uniform or almost uniform colouring, and the best results with the mean absolute error 0.34 and standard deviation 0.53 – 0.54 in the total disparity range $[0, 10]$ are obtained for $P_{\mathrm{B|B}} = 0.01$ and $P_{\mathrm{M|M}} = 0.80 \ldots 0.95$ ($\sigma = 5$) or $0.40 \ldots 0.95$ ($\sigma = 10$). The "Pentagon" scene with more textured colouring gives the apparently best visual results for larger values of $P_{\mathrm{B|B}}$, e.g., $P_{\mathrm{B|B}} = 0.10 \ldots 0.20$ and $P_{\mathrm{M|M}} = 0.80$ in Figure 3, and the like results are even better for $\sigma = 5$ as shown in [7].

These experiments suggest that the proposed approach offers advantages over conventional reconstruction techniques based on stereo matching that allows for

partial occlusions. Also, it shows promise of taking account of image colouring uniformity if this latter could be properly related to the regularising parameters.

Acknowledgements

This work was supported in part by the University of Auckland Research Committee research grant XXXX/9343/3414108.

References

1. Baker, H. H.: Surfaces from mono and stereo images. Photogrammetria **39** (1984) 217–237.
2. Förstner, W.: Image matching. In: R. M. Haralick and L. G. Shapiro: *Computer and Robot Vision*. Vol. 2, chapter 16. Reading, Addison-Wesley (1993) 289–378.
3. Gimel'farb, G. L.: Intensity-based computer binocular stereo vision: signal models and algorithms. Int. J. of Imaging Systems and Technology **3** (1991) 189–200.
4. Gimel'farb, G. L.: Regularization of low-level binocular stereo vision considering surface smoothness and dissimilarity of superimposed stereo images. In: C. Arcelli, L. P. Cordella, G. Sanniti di Baja (Eds.): *Aspects of Visual Form Processing*. Singapore, World Scientific (1994) 231–240.
5. Gimelfarb, G. L.: Symmetric bi- and trinocular stereo: tradeoffs between theoretical foundations and heuristics. Computing Supplement **11** (1996) 53–72.
6. Gimel'farb, G.: Stereo terrain reconstruction by dynamic programming. In: B. Jaehne, H. Haussecker, P. Geisser (Eds.): *Handbook of Computer Vision and Applications* **2**: *Signal Processing and Pattern Recognition*. San Diego, Academic Press (1999) 505–530.
7. Gimel'farb, G., Li, H.: Probabilistic regularisation in symmetric dynamic programming stereo. In: *Proc. of the Image and Vision Computing New Zealand'2000 Conf.*, November 2000, Hamilton, New Zealand. (2000) [In print].
8. Hannah, M. J.: Digital stereo image matching techniques. Int. Archives on Photogrammetry and Remote Sensing **27** (1988) 280–293.
9. Kanade, T., Okutomi, M.: A stereo matching algorithm with an adaptive window: theory and experiment. IEEE Trans. on Pattern Analysis and Machine Intelligence **16** (1994) 920–932.
10. Kendall, M. G., Stuart, A.: *The Advanced Theory of Statistics*. **2**: *Inference and Relationship*. London, Charles Griffin (1967).
11. Kyreitov, V. R.: *Inverse Problems of Photometry*. Novosibirsk, Computing Center of the Siberian Branch of the Academy of Sciences of the USSR (1983) [In Russian].
12. Poggio, T., Torre, V., Koch, C.: Computational vision and regularization theory. Nature **317** (1985) 317–319.
13. Wei, G.-Q., Brauer, W., Hirzinger, G.: Intensity- and gradient-based stereo matching using hierarchical Gaussian basis functions. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 1143–1160.

Camera Calibration of Rectangular Textures

Jacky Baltes
Centre for Imaging Technology and Robotics
University of Auckland
Auckland, New Zealand
j.baltes@auckland.ac.nz

No Institute Given

Abstract. This paper describes a practical method for the camera calibration given a single image of a regular texture. This paper uses the calibration of images of skyscrapers as an example. The paper introduces two algorithms for the assignment of real world coordinates to feature points. The first algorithm selects five closely connected feature points and determines the orientation of the rectangular pattern. The second algorithm iteratively sorts the feature points and assigns real world coordinates to them. Lastly, the Tsai camera calibration algorithm is used to compute the camera parameters.

1 Introduction

This paper describes an application of our camera calibration method, which was initially developed for RoboCup.

RoboCup is an international competition of fully autonomous robots playing soccer [4]. The first competiton was organized in 1997, and it has rapidly increased in popularity.

Apart from the obvious challenges in robotics, control theory, path planning, artificial intelligence, and machine learning, RoboCup also presents an interesting domain for real-time computer vison. In the small league, robots are identified using a global vision system. To achieve adequate control, a vision system must track ten robots, a ball and compute their position, orientation, velocity with a cycle time of less than 20ms. More details about the All Botz videoserver can be found in [1].

This paper shows how the camera calibration method which was originally developed for the RoboCup domain can be used to compute the calibrate images of any regular textures or co-planar feature points using a single image.

Section 2 is an introduction to camera calibration in general and the challenges of calibration using regular textures. The extraction of calibration points is shown in section 3. The Tsai camera calibration used as back end to compute the final calibration parameters is described in 4. The paper concludes with section 5.

2 Camera Calibration

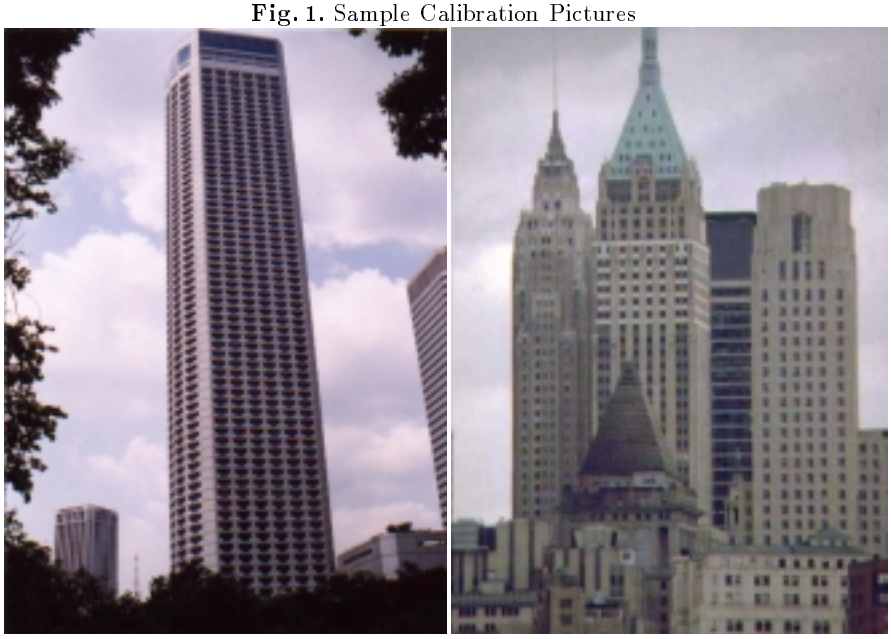
Accurate camera calibration is an essential ingredient in any computer vision system. Therefore, it has been a very active and productive research area. A number of different camera calibration methods have been developed [5]. Most of these methods are based on the pin-hole camera model.

The input to camera calibration methods is a set of image features and their associated real-world attributes. For example, the well known Tsai calibration uses a set of image points with known real world coordinates as input. Haralick proposes a calibration method that use the image coordinates of parallel lines[3].

Most camera calibration methods use calibration objects, for example cubes with colour patches. These calibration objects allow accurate control over their feature points and allow therefore accurate calibration.

Calibration of pictures of natural scenes requires a different approach because the calibration objects are too small. For example, assume that we need to calibrate the geometry of the images shown in Fig. 1. It is clearly impractical to build a calibration cube of this dimension.

Fig. 1. Sample Calibration Pictures



Skyscraper (Singapore) Wall Street (New York)

As can be seen though, both images contain regular feature points that can be used for calibration. For example, if the distance between floors and between

windows is known, then the position of the centres of the windows can be calculated. This is the basic method in our approach. However, the problem with real world scenes is that the features are hard to extract from the image. Furthermore, there are many feature points and to assign them manually would be time consuming and error prone. Lastly, not all feature points can be extracted from the image and some of them will be missing. The following section discusses a system to automatically deal with these problems.

3 Extraction of Calibration Points

This section describes the algorithm for sorting the feature points and assigning real world coordinates to them. Our system uses a semi-automatic approach. First, a manual preprocessing step is used to segment the image (Subsection 3.1). Secondly, an automatic routine is used to compute the centres of the feature points, to sort them, to correct for missing features, and to assign real world coordinates to the points.

3.1 Image Preprocessing

The first step in the camera calibration routine is a manual preprocessing step to clean up the image, to remove unwanted artefacts, and to select suitable parameter settings for the colour segmentation routines. Figure 2 shows the output of the pre-processing step for the Singapore skyscraper (see Fig. 1).

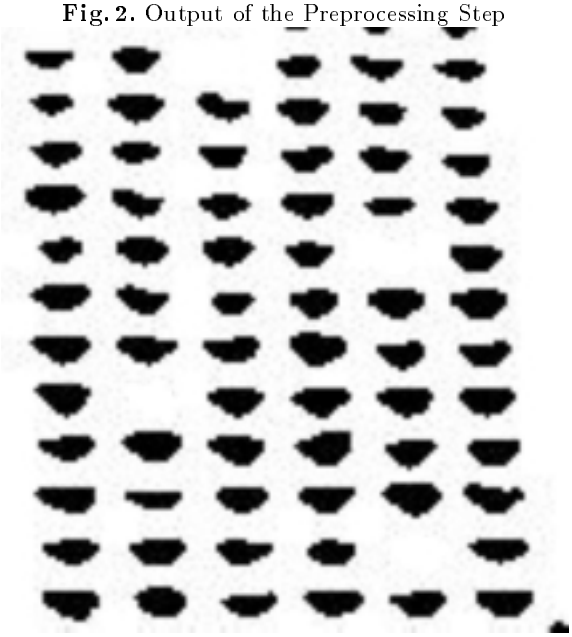
As can be seen, some of the features were not distinct enough to be recognized and are missing in the preprocessed image.

3.2 Sorting of Feature Points

This section describes the heart of the calibration routine — the algorithm to assign real world coordinates to the feature points that were extracted in the pre-processing step. The algorithm consists of two main parts: (a) algorithm 1 computes an initial guess of the translation matrix. (b) algorithm 2 is an iterative algorithm which assigns real world coordinates to the feature points and updates the guess of the transformation matrix.

In line 1 of algorithm 1, the centres of the features are extracted. Then in line 2, a number of seed points are selected. These seed points are used to compute the initial transformation matrix. Five seed points are selected to form a cross in the two-dimensional plane. To guarantee a good estimate of the initial transformation matrix, the five centres that form a cross closest to the centre of the image are selected. The routine `computeTransformationMatrix` called in line 3 uses the Boyer-Moore pseudo inverse method to compute a least means square (LMS) approximation for the 3×4 transformation matrix for the perspective projection. Since our system assumes that the distances between squares are

Fig. 2. Output of the Preprocessing Step



Algorithm 1 Algorithm to Assign Real World Coordinates

1: $S = \text{extractCentres}(Image)$
2: $P = \text{selectSeedPoints}(S)$
3: $M_{XY} = \text{computeTransformationMatrix}(S, \text{realWorldCoords}(S))$
4: $M_{YX} = \text{computeTransformationMatrix}(S, \text{swapCoord}(\text{realWorldCoords}(S)))$
5: **if** $\text{error}(M_{XY}) < \text{error}(M_{YX})$ **then**
6: $\quad M_{Initial} = M_{XY}$
7: **else**
8: $\quad M_{Initial} = M_{YX}$
9: $\quad \text{swapCoordinates}(S)$
10: **end if**
11: $C_S = \text{assignCentres}(S, M_{Initial})$

not identical, the algorithm checks both possible orientations for the width and the height of the features and selects the best match (Line 4).

Algorithm 2 describes the method for the assignment of real world coordinates to the feature points. This algorithm iteratively selects the unassigned feature points that are closest to an already assigned point. In line 8, the variables b_x and b_y contain the closest number of blocks given the current estimate M of the transformation matrix. This estimate of the number of blocks allows the algorithm to compensate for missing feature points. Once the real world coordinates have been assigned to the new point, a new transformation matrix is computed (line 11). This process is repeated until all points have been assigned.

Algorithm 2 Algorithm to Assign Real World Coordinates($S, M_{Initial}$)

1: $M = M_{Initial}$
2: $N =$
3: **while** $S \neq$ **do**
4: **for all** $s \in S$ **do**
5: $s_{North}, s_{East}, s_{South}, s_{West} =$ findNearestNeighbors(s, M)
6: **end for**
7: **for all** $n \in \{s_{North}, s_{East}, s_{South}, s_{West}\}$ **do**
8: $b_x = dist((s-n)/width, M), b_y = dist((s-n)/height, M)$
9: $n_x, n_y = s_x + bx * width, s_y = by * height$
10: $N = N + n, S = S - n$
11: $M =$ computeTransformationMatrix$(N,$realWorldCoords$(S))$
12: **end for**
13: **end while**
14: return N

The output of the system is shown in Fig. 3 shows the output of our system. All feature points are correctly aligned and the system correctly corrects for the missing features. The initial five seed points are shown in the gray shaded region.

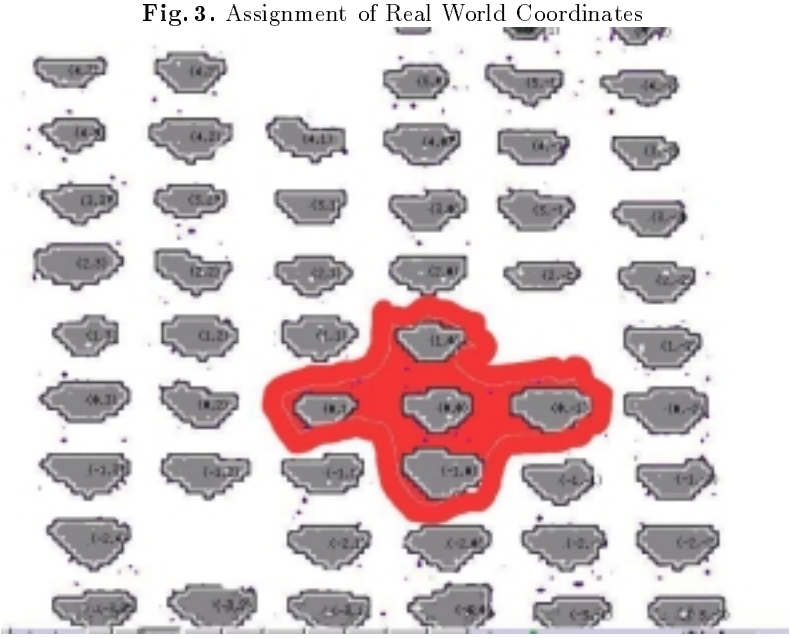
4 Tsai Camera Calibration

After the computation of the matching points, we use a public domain implementation of Tsai's camera calibration to compute the extrinsic and intrinsic parameters of the camera model.

The Tsai calibration method uses a four step process to compute the parameters of a pin hole camera with radial lens distortion.

Firstly, the position (X_T, Y_T, Z_T) and the orientation (R_X, R_Y, R_Z) of the camera with respect to the world coordinate system is computed. This involves solving a simple system of linear equations. This step translates the 3D World coordinates into 3D camera coordinates and computes the six extrinsic parameters of the camera model.

Fig. 3. Assignment of Real World Coordinates



In Step 2, the perspective distortion of a pin hole camera is compensated for. This step is a non-linear approximation and computes the focal length f of the camera. The output of this step are the ideal undistorted image coordinates.

Thirdly, the radial lens distortion parameters (κ_1, κ_2) are computed. These parameters compensate for the pin cushion effect of video cameras, that is straight lines along the edges of the camera are rounded. The output of step 3 are the distorted image coordinates.

Lastly, the image coordinates are discretized into the real image coordinates by taking the number of pixels in each row and column of an image into consideration.

The last three steps compute five intrinsic parameters of the camera model (focal length, lens distortion, scale factor for the rows, and the origin in the image plane).

The Tsai method is a very efficient, accurate, and versatile camera calibration method and is therefore very popular in computer vision. Nevertheless, one of the shortcomings of Tsai's method is that it is not able to compute the uncertainty factor S_X from only co-planar calibration points. In practice, this does not present a big problem, since most natural 3D textures consist of several planar textures. The calibration points that are extracted using our method can be concatenated into a large set of 3D calibration points.

Furthermore, during the sorting of the feature vectors, our extraction algorithms use a simple pin hole camera model to compute the perspective projec-

tion parameters. Therefore, our method is independent of the exact calibration method used.

5 Conclusion

This paper presents a practical system for the accurate calibration of real world scenes that have a regular texture. Our system is currently limited to rectangular textures, but this limitation is due to the current implementation. In theory, any regular texture is suitable.

The Tsai calibration does not provide an efficient method for calculation of the uncertainty factor S_x given co-planar calibration points. We are investigating other suitable methods as the one described in [2].

References

1. Jacky Baltes. Practical camera and colour calibration for large scale rooms. In *Proceedings of the IJCAI Workshop on RoboCup*, Stockholm, Sweden, July 1999.
2. J. Batista, H. Araujo, and A.T. Almeida. Monoplanar camera calibration: Iterative multistep approach. In *BMVC93*, page xx, 1993.
3. R. Haralik. Determining camera parameters from the perspecive projection of a rectangle. *Pattern Recognition*, 22(3), 1989.
4. Hiroaki Kitano, editor. *RoboCup-97: Robot Soccer World Cup I.* Springer Verlag, 1998.
5. Roger Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, Miami Beach, FL, 1986.