Robust 3-D Landmark Tracking using Trinocular Vision

John Mallon, Ovidiu Ghita, Paul F. Whelan
Vision Systems Laboratory, School of Electronic Engineering
Dublin City University, Dublin 9, Ireland
Ph: +353 1 7005869, Fax: +353 1 7005508
{john.mallon,ghitao,paul.whelan}@eeng.dcu.ie

Abstract

Position determination and verification of a mobile robot is a central theme in robotics research. Several methods have been proposed for this problem, including the use of visual feedback information. These vision systems typically aim to extract known or tracked landmarks from the environment to localise the robot. Detection and matching these landmarks is often the most computationally expensive and error prone component of the system. This paper presents a real-time system for robustly matching landmarks in complex scenes, with subsequent tracking. The vision system comprises of a trinocular head, from which corner points are extracted. These are then matched with respect to robustness constraints in addition to the trinocular constraints. Finally, the resulting robustly extracted corners are tracked from frame to frame to determine the robot's rotational deviations.

Keywords: Trinocular vision, Landmark detection, Landmark tracking, Robotic navigation

1. Introduction

Determining the location of a robot is an important problem in navigation of an autonomous vehicle in an unstructured environment. A problem arises as there are always errors associated with the robot's motion. The actual position and orientation of a mobile robot is traditionally estimated using odometry, derived from wheel encoders and the robot's kinematic model. It has inherent advantages in that it relies on simple geometric equations [1,2] and sensors, but the fundamental disadvantage lies in the accumulation nature of measurement errors. Errors are introduced from many sources including terrain conditions and sensor error. These inaccuracies force a restriction on the use of odometry for short periods of time

Several methods have been proposed in the literature to compensate for these errors and localise the robot with respect to its environment. Most significantly, visual information is used for this task, and as a result it has been heavily investigated. Many of these techniques require a means of detecting features (also referred to as landmarks) in sensor data that may correspond to structures in the environment. Three main methods are used for localising the robot using these landmarks. First, the observed landmarks are geometrically fitted to an *a-priori* map of the robot's environment [3,4]. A second approach assumes prior knowledge of location, and a model of robot motion can be used to predict the appearance and/or location of landmarks. Many authors including [5,6] describe systems based on this approach, using an extended Kalman filter to reduce positional errors. Finally, a third approach assumes that the features themselves can be tracked over time in the image domain [7,8]. Detection and matching of landmarks is common to these approaches and is often the most computationally expensive and error prone component of the system. This can be simplified by using active emitters or artificial landmarks [9,10] to facilitate easier landmark extraction and registration. However these simplifications restrict the robot's adaptability considerably. Consequently, methods for robustly extracting and matching landmarks from natural scenes have been investigated. Despite the quantity of work, there are several problems that remain to be solved. The successful operation of all vision based methods relies on robust landmark detection and matching. However, in complex and dynamic environments, most systems still lack robustness, reactivity and flexibility. As a result there is no widespread acceptance or convergence towards any vision based localisation method.

In this paper we describe a novel, but more importantly, robust strategy for natural landmark correspondence with subsequent tracking, to detect and correct trajectory deviations for a mobile robot in real time. The approach can readily be used with any vision based landmark localisation technique, indoor or outdoor, and offers the much lacked robustness required for visual servoing applications [11]. The approach takes advantage of a trinocular system from which corner points are robustly matched, using multiple constraints including the trinocular constraint. These points are then tracked from frame to frame to establish the ego-motion of the images and hence the robot's. The paper is organised as follows: Section 2 briefly outlines the drawbacks of odometry and the assumptions made about positional errors in our mobile robot system. Section 3 describes the trinocular vision system used in this implementation. Landmark extraction and matching is detailed in Section 4, while motion estimation is dealt with in Section 5. Section 6 presents some experimental results.

2. Odometry

It has been accepted that odometry data in mobile robotics cannot be relied upon for extended periods of time [12]. In a two dimensional space, the location of the robot can be represented by the triplet *(X Y θ)* [1,2], where *X*, *Y* and *θ* represent the position and orientation of the robot's co-ordinate frame relative to a base or environment frame. Errors are introduced from many elements including uneven terrain surfaces and encoder inadequacies at low velocities, such as those encountered when accelerating around the stationary mode. These errors are considered random and can be represented as shown in Fig. 1, where the ellipses represent the relative uncertainty of the robot's position. Considering the mobile robot, MOBIUS [13], with two centrally situated, diametrically opposed drive wheels, these errors are introduced to the system in the form of rotational errors in *θ*, as the robot cannot translate perpendicular to the direction of travel (*θ*). These rotational deviations skew the robot's trajectory, resulting in position ambiguity. Many authors including [6,7] have proposed to use a sequence of visual information for ego-motion estimation and hence determine positional deviations. Harris's 3-D vision system DROID [14] uses the visual motion of image corner features, but is very dependent on correctly matching corners from the stereo pair and the subsequent frame to frame matching.

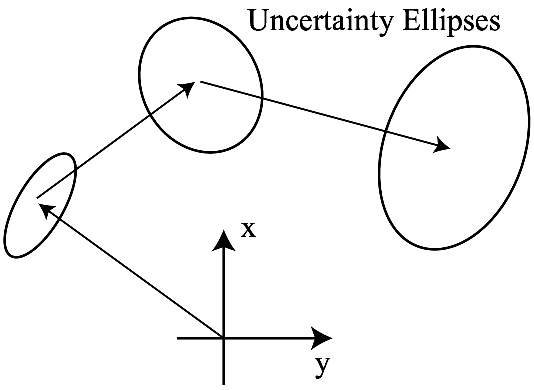


Fig. 1. A typical sequence of mobile robot translations, with associated positional uncertainties [12].

3. Trinocular vision

Our system uses a linear configuration of three equally spaced cameras, calibrated to ensure the epipolar lines lie along the scan lines. This allows us to exploit the advantages of long and short baselines [15]. Corner features are extracted from each image. These points are matched in the short baseline pair and the long baseline pair, using the right outermost camera as the reference, as described in Section 4.2. In addition to the epipolar constraint, the trinocular constraint is also applied. This states that for correctly matched points using linearly arranged and equally spaced cameras, the disparity in the short baseline must be twice that of the long baseline for each match with respect to the reference. This constraint has the effect of removing matching ambiguity as illustrated in Fig. 2, and dramatically improves matching robustness. Also, matching across three perspective views tends to retain only feature points that are more invariant to perspective changes, thus increasing the possibility of correctly extracting and matching them from frame to frame.

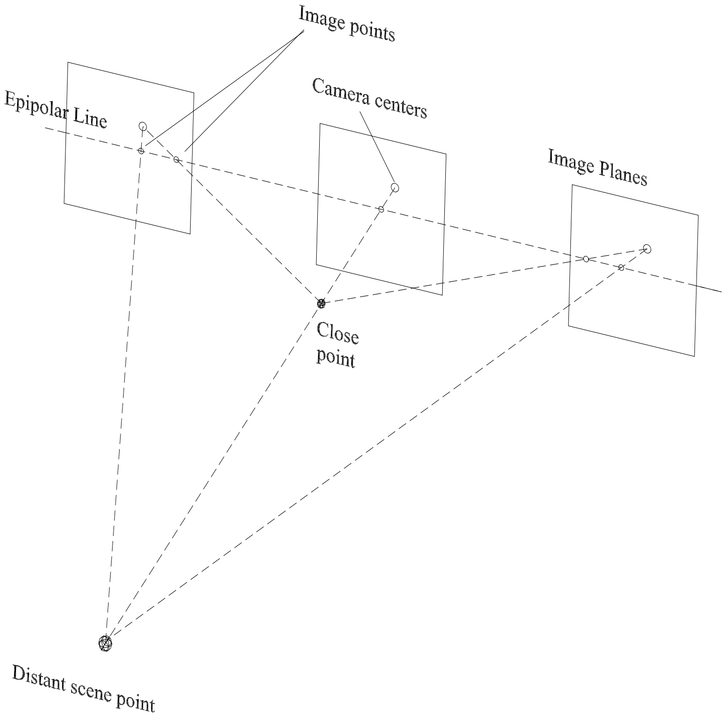


Fig. 2. Trinocular configuration, shows the equally spaced linear arrangement. Long baseline (two outside images) dimension is 30cm. The systems ability to deal with matching ambiguities and occlusions that cannot be dealt with successfully by a standard two camera stereo configuration is demonstrated.

4. Feature matching

An image contains a very large amount of data, from which only a relatively small amount of salient information must be extracted if real-time processing to be achieved. To perform this data compression, each image is analysed for features that are likely to correspond to 3-D scene elements. These features should recur through the sequence of images (while still in view) and be accurately locatable. Robustly matching these features is of the utmost importance in many vision based robot localisation and tracking schemes. Due to occlusions and missing parts some feature points in one image may not have a correspondence in another image. In addition, very often the scene under analysis reveals repetitive patterns and the matching algorithm may return multiple matches. Thus, a central problem in feature matching resides in selecting features that are robustly recovered in each image, immune to noise and invariant to geometric transformations. In general for this purpose features that have strong local support such as corners [14] and edges [16] are employed. In general edges generate large pixel structures and can result in computationally expensive matching algorithms. Corners are ideal features since they define areas in the image with strong textural characteristics, and are thus distinct with respect to their neighbourhood. For computational purposes, the SUSAN corner detector [17] was used.

4.1 Image normalisation

An important assumption in motion and structure analysis, optical flow estimation and stereo among others is that of constant image brightness (CIB). In spite of laborious camera calibration procedures, very often due to non-idealities in the optical and sensing equipment the image intensities of corresponding points may not be equal. One solution to this problem is to develop an expression that maps the intensity values between the images. It is common practice to apply a Laplacian of Gaussian (LoG) operator in order to eliminate the DC components of the images. Unfortunately, this image normalisation model compensates only for additive brightness variation, while the multiplicative brightness variation is not addressed. In order to compensate for this limitation, Cox et al [18] developed an image normalisation procedure called *dynamic histogram warping* (DHW) which consists of mapping the intensities in the left and right images using a grey-level global transformation derived from analysing the histograms of the images.

$$I_l(x,y) = \alpha I_r(x,y) + \beta \tag{1}$$

where α and β are the parameters of the transformation.

Our experiments indicate that this image normalisation technique outperforms the method based on the application of the LoG operator. It is important to note that the application of the DWH is restricted to cases where the number of occluded points is small compared to the overall number of pixels. This requirement is not fulfilled with images taken from widely varying viewpoints or affected by severe perspective distortions. These effects are minimal in our implementation since the cameras are arranged in a parallel configuration with a relatively small disparity.

4.2 Correspondence

Corner correspondence is estimated by taking a small region of pixels surrounding the corner to be matched in the reference image and comparing this with a similar window around each of the potential matching corners in subsequent images [19]. Each comparison yields a score, with the highest representing the best similarity. In our implementation the measure of similarity is evaluated using normalised SSD along the epipolar lines.

$$E = \frac{\sum_{x=0}^{M}\sum_{y=0}^{N}\left[I_r(x,y) - I_l(x+dx, y+dy)\right]^2}{\sqrt{\sum_{x=0}^{M}\sum_{y=0}^{N} I_r(x,y)^2} \times \sqrt{\sum_{x=0}^{M}\sum_{y=0}^{N} I_l(x+dx, y+dy)^2}} \tag{2}$$

where I_r is the right image, I_l is the left image, *dx* and *dy* are the disparities on the *x* and *y* axes.

We consider that the correlation window for each corner is affected only by translation. This assumption is valid as the trinocular system is arranged in a parallel manner. Matching algorithms in general select a corresponding point by imposing a number of constraints on the similarity measure. These are employed to avoid exhaustive searching and to increase robustness. In this respect, we apply the disparity constraint in order to select only the corners with *x* co-ordinates to the left of the corner in the reference image (rightmost image). To further reduce the number of investigations of potential corners we applied a maximum disparity constraint to eliminate corners situated too far away from the corner under investigation in the reference image. The maximum disparity constraint is set to 60 pixels in the short baseline pair and 120 pixels in the long baseline pair (image size: 768 x 576), with a vertical tolerance of 3 pixels to account for corner jitters. The corners left in the list of potential matches are ordered in agreement with the measure of similarity.

In order to select only robust matches the corner with the highest score is only considered a valid match if it respects two additional constraints. These are the confidence constraint and the uniqueness or distinctness constraint. The confidence constraint requires that the score of the best match has to be greater than a confidence threshold, set in agreement with the similarity measure. The aim of the uniqueness constraint is to select only distinct matches. This constraint is met if the difference between the scores of the best two matches is higher than a threshold *K*, E_n-$E_{n-1} \geq K$. Both threshold levels were determined experimentally [20]. Additionally, in order to establish matching consistency, double correlation (right, left and left, right correlation) was applied by searching for matches in the previous reference image using the corner list from the other images. To satisfy this constraint both sets of correlations must result in the same corner matches.

The corner matching is carried out individually on the short and long baseline pairs. The reference (rightmost) camera and the middle camera form the short baseline pair, while the reference camera and the leftmost camera form the long baseline pair. The trinocular constraint is applied to merge the resulting corners in both the long and short baselines. Therefore, reliable matches are retained if for each match over the short baseline, there is a corresponding long baseline match with double (±10%) the disparity value. Matches in the short baseline that do not respect this constraint are dismissed. Due to the inherent problems in matching over a long baseline [15], many corners that have been matched over the short baseline may not have been matched in the long baseline image pair. This considerably reduces the amount of final matched points after the trinocular constraint has been applied. However, the matched corners in the local neighbourhood surrounding the unmatched corner over the long baseline may still indicate that the trinocular constraint is upheld locally. This is verified using the following procedure:

- For each unmatched corner over the long baseline pair, it is investigated if the corner has a match over the short baseline.
- If so, over the short baseline, neighbouring matched corners are clustered, using an agglomerative clustering technique [21], with respect to disparity values. To ensure that the corner under analysis is not mismatched and the local coherence is preserved, its disparity should be similar to that of its local neighbourhood. In this respect the mean of the largest cluster must be similar with that of the corner's disparity.
- A similar procedure is applied with the long baseline, by clustering the local region around the unmatched corner.
- The largest local cluster over the long and short baselines should form a consensus with the trinocular constraint.

Fig 3 illustrates the principle, while Fig 7 shows the quantity of resulting corners matched in typically complex indoor scene.

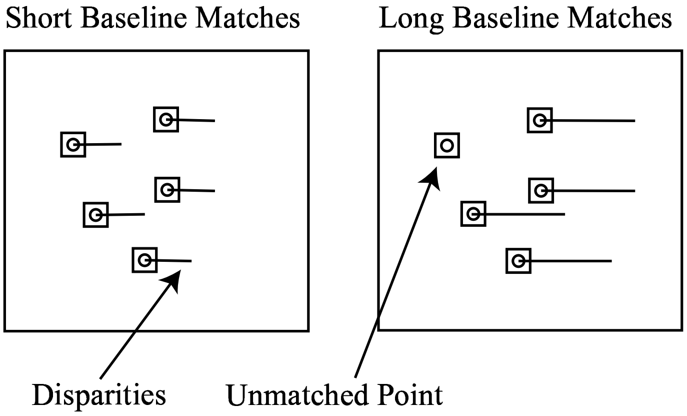


Fig. 3. Local regions around an unmatched corner. The unmatched point is retrieved by clustering local disparities in the short and long baseline matches. If the clusters respect the trinocular constraint locally, as shown, the corner is added to the list of robust landmarks.

5. Corner tracking and motion estimation

The landmark extraction method described above can now be used to determine the ego-motion of the cameras, and hence the motion of the robot. As mentioned all positional errors introduced to the mobile robot are in the form of rotations. If the landmarks in the robot's environment are tracked from frame to frame these rotations can be identified and compensated for. The corner matching method described operates in real-time, which offers much improved tracking capabilities in a dynamic environment. Tracking requires a quantity of the corner points from the previous frame to be present in the current frame. A search is then conducted to locate these same corners, which should form a consensus as to the motion of the cameras. Corner correspondence from frame to frame is performed as before, using the measure of similarity, with the search space constrained to ± 30 pixels around the corner under investigation and a 3 pixel vertical tolerance to allow for inaccuracies due to motion and corner jittering. The confidence and uniqueness constraints described above are also applied. In addition, the depth information of the corners must be respected. Depending on the velocity of the robot this information either remains unchanged, or can be confirmed from odometry data. Fig. 8 shows the resulting corners matched from a rotation. These rotations can be determined explicitly using the general cosine relationship. Fig. 4 illustrates the available reliable information derived from the algorithm.

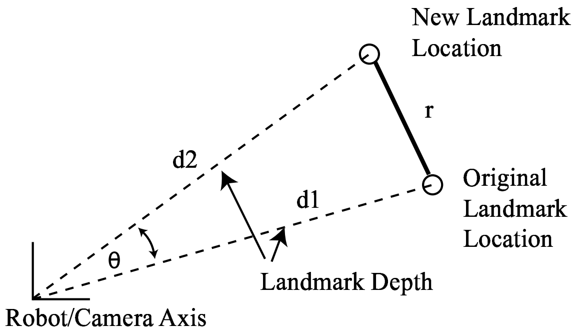


Fig. 4. Change in position of a landmark due to a rotation.

The angle in Fig. 4 is calculated for each match. The results are clustered as some matches may result from dynamics in the scene.

6. Experimental results

The trinocular head is designed of use in conjunction with a mobile robot, thus simultaneous image acquisition is essential in such dynamic scenes. In this regard an Ellipse RIO frame grabber was utilised, while three Hitachi KP-M1 cameras fitted with 6mm Computar lenses image the scene. Experiments were conducted on indoor laboratory scenes. Images are 768 x 576 pixels with 192 grey levels. Fig. 5 shows the three images captured from such a typical scene.

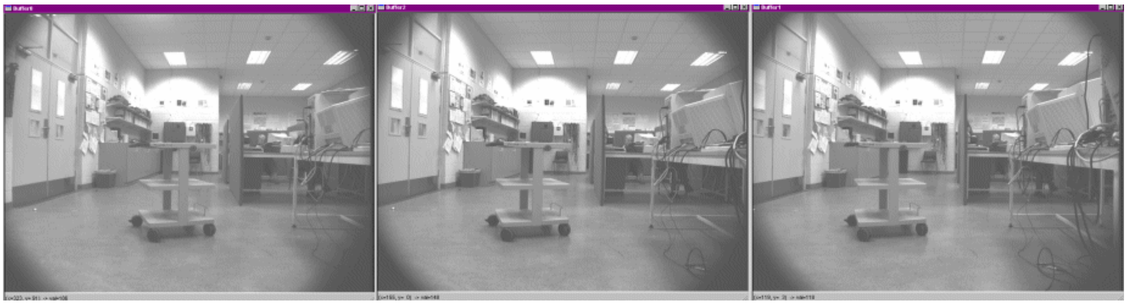


Fig. 5. Trinocular images of a typical indoor laboratory scene.

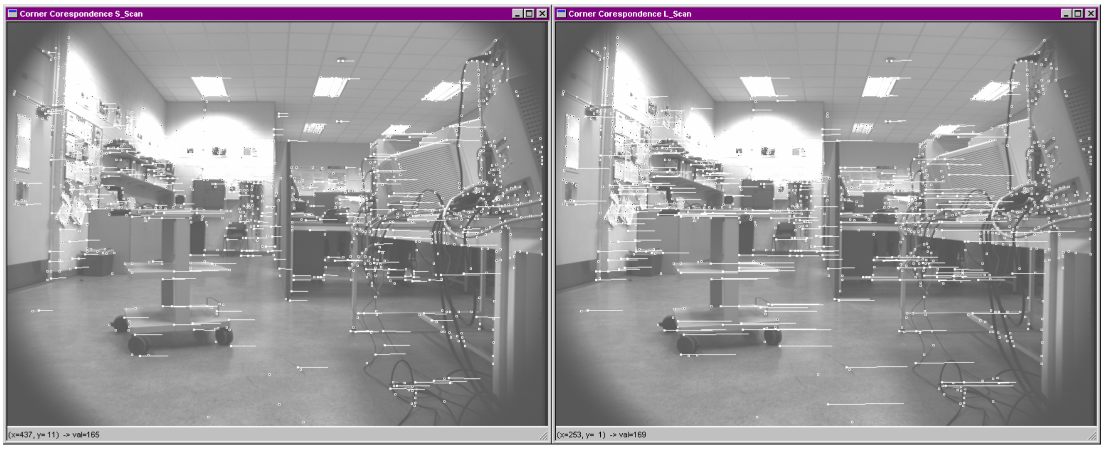

Corner Corespondence S_Scan
(x=437, y= 11) -> val=165
Corner Corespondence L_Scan
(x=253, y= 1) -> val=169

Fig. 6. Resulting corner matches over the short (left) and long (right) baselines with respect to the reference image (rightmost).

A major aim of this implementation was robust landmark detection. In this regard, corners were matched in the long and short baseline images using a 7 x 7 correlation window. A number of constraints including the confidence and uniqueness constraints were applied to the correlation scores. Fig. 6 shows the resulting matches in the short and long baseline images.

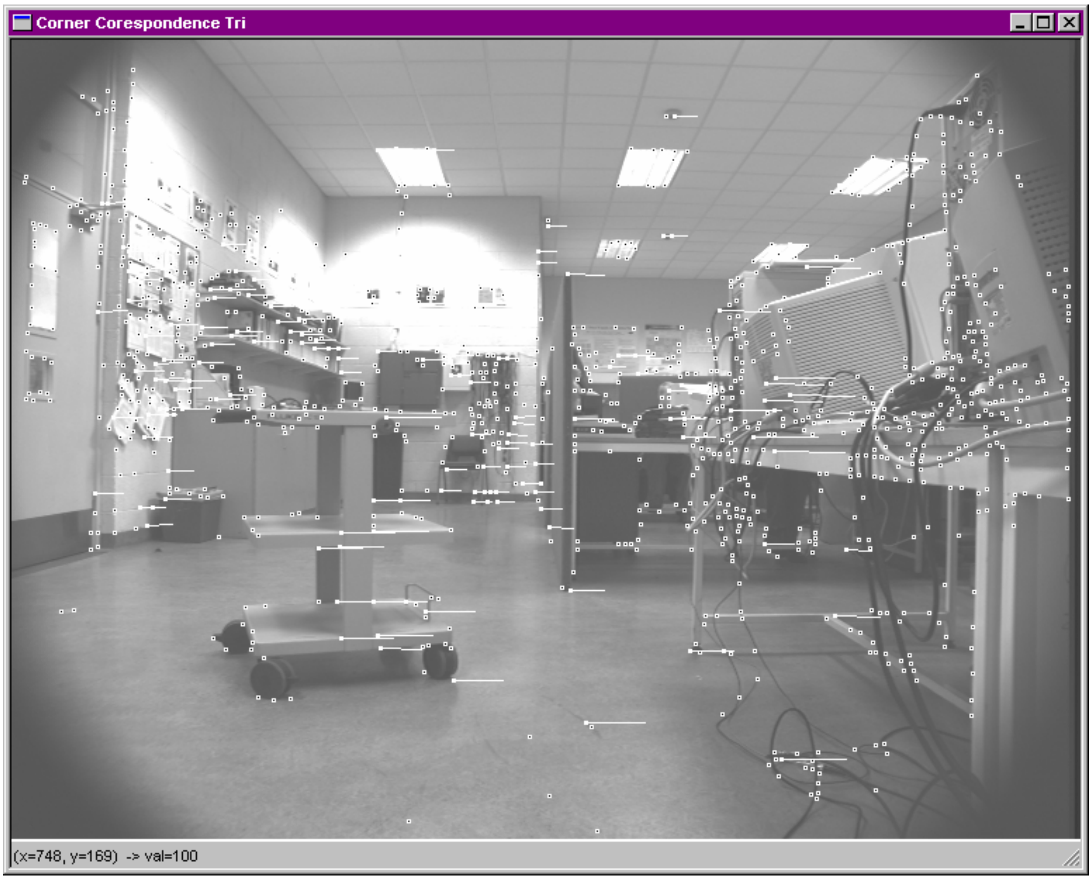
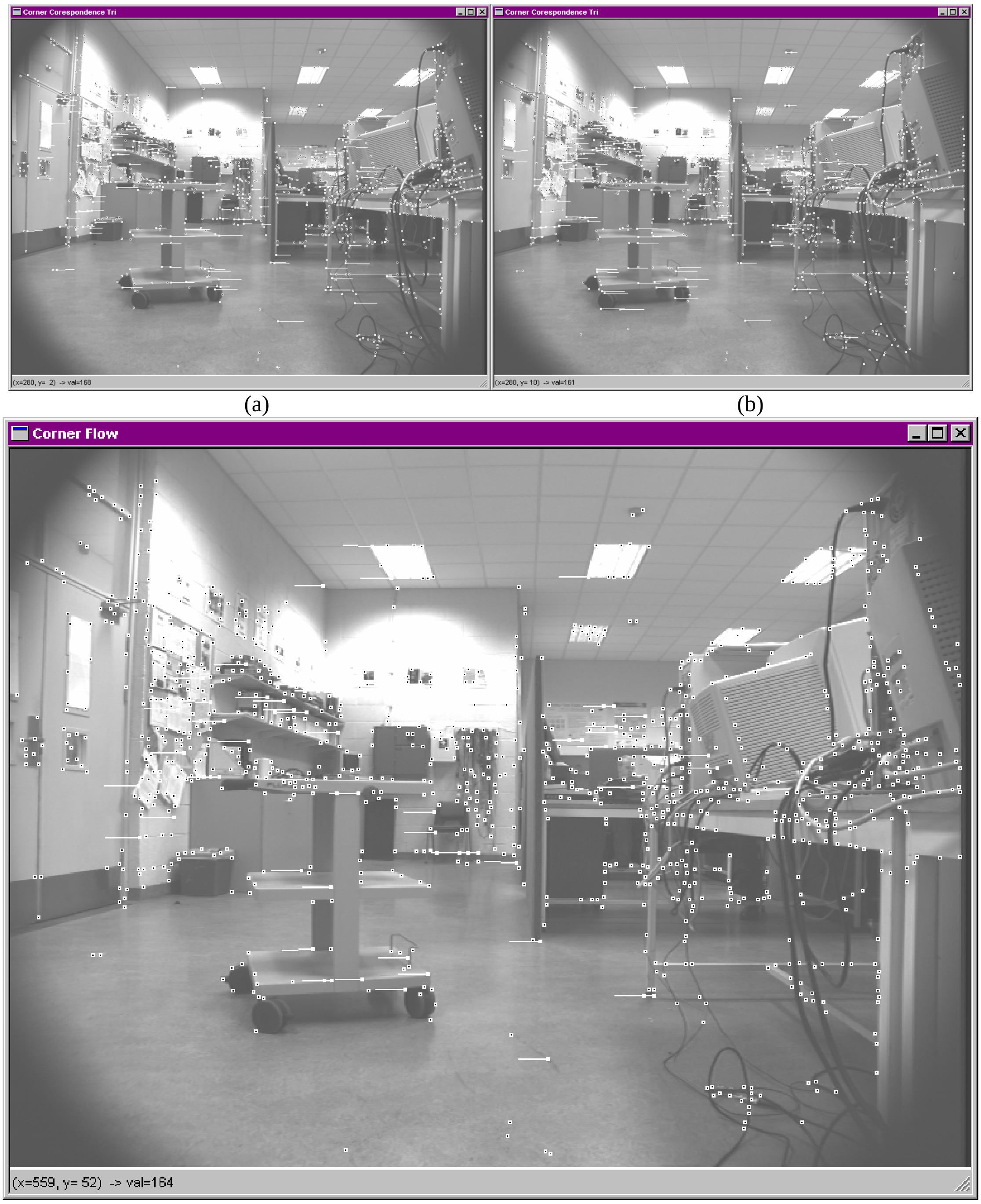


Fig. 7. The resulting corner correspondence that respect the trinocular constraint (note the figure shows the retention of short baseline matches).

The trinocular constraints are then applied to these results. Any matched pair of corners in Fig 6 that do not respect this constraint are dismissed. Also, a facility is provided to match any unmatched corner in the long baseline image (refer to Section 4.2). Fig. 7 shows the results of the application of this constraint.



(a) (b)

(c)

Fig. 8. Set of landmarks picked in two sequential frames (a and b). The resulting corner tracking results (c) are also shown, with approximately 46 corner points, all correctly tracked.

In order to detect the rotational errors of the robot, we apply frame to frame tracking using the corners above. Fig. 8 shows the quantity of corners matched between consecutive frames (shown) for an approximate clockwise rotation of 3 degrees.

As mentioned, the system gains real time performances. Time for processing the above images to the results shown in Fig. 7 is 970ms, running on a 500Mhz, 256MB RAM under Windows NT. It should be noted that code is not optimised for speed. Currently the extracted measures are in pixels. Specifically these are the disparity and corner tracking values. Further calibration work is planned to transform these measures into the metric domain.

Conclusion

This paper presents a robust real-time approach to the problem of using natural landmarks as visual feedback to the navigation system of a mobile robot. The main limitations of many such strategies are due to errors introduced from the landmark correspondence phase. In this regard, we have implemented a trinocular vision system, which increases the robustness of landmark correspondences very considerably compared to that of a stereo or monocular system. The algorithm first extracts corners from the images, which are then matched in the long and short baselines, subject to certain matching constraints. These results are combined in agreement with the constraints imposed by the trinocular system. The resulting set of corners are tracked from frame to frame in order to detect motion errors in the robot's trajectory. Experimental data indicates that this implementation offers a substantial increase in feature matching reliability for many applications.

References

[1] Alexander J. C. & Maddocks J. H. (1989), On the kinematics of wheeled mobile robots, *International Journal of Robotics Research,* 8(5) pp. 15-27

[2] Borenstein, J. & Feng, L. (1996), Measurement and correction of systematic odometry errors in mobile robots, *IEEE Transactions on Robotics and Automation* 12(5) pp. 869-880

[3] Se S., Lowe D. & Little J. (2001), Vision-based mobile robot localisation and mapping using scale-invariant features, *Proceedings of IEEE international conference on Robotics and Automation,* pp. 2051-2058

[4] Atiya S. & Hager G. D. (1993), Real-time vision-based robot localisation, *IEEE Transactions on Robotics and Automation* 9(6) pp. 785-800

[5] Cox I. J. (1989), Blanche: position estimation for an autonomous robot vehicle, *Proceedings of the IEEE/RSJ International workshop on Robots and Systems (IROS '98),* pp. 432-439

[6] Kriegman D. J., Triendl E. and Binford T. O. (1989), Stereo vision and navigation in buildings for mobile robots *IEEE Transactions on Robotics and Automation* 5(6) pp. 792-803

[7] Andersson R. L. (1989), Dynamic sensing in a ping-pong playing robot *IEEE Transactions on Robotics and Automation* 5(6) pp. 728-739

[8] Zhang Z. & Faugeras O. D. (1991), Tracking 3-D line segments: New developments, *Proceedings of International Conference on Advanced Robotics.* pp. 1365-1370

[9] Leonard J. L. & Durrant-Whyte H. F. (1991), Mobile robot localisation by tracking geometric beacons, *IEEE Transactions on Robotics and Automation* 7(3) pp. 376-382

[10] Lin C. & Tummala R. L. (1997), Mobile robot navigation using artificial landmarks *Journal of Robotic Systems* 14(2) pp. 93-106

[11] Espiau B., Chaumette F. & Rives P. (1992), A new approach to visual servoing in robotics, *IEEE Transactions on Robotics and Automation* 8(3) pp. 313-326

[12] Wang. C. M. (1988), Location estimation and uncertainty analysis for mobile robots *Proceedings of the International Conference on Robotics and Automation*, pp. 1230-1235

[13] Mallon J., Ghita O. & Whelan P. (2002), An integrated design towards the implementation of an autonomous mobile robot, *International Conference on Optimisation of Electrical & Electronic Equipment*, *OPTIM 2002,* Brasov, Romania.

[14] Harris C. (1992), Geometry from visual motion, In Blake A. & Yuille A. editors, *Active Vision*, MIT Press, pp. 264-284

[15] Okutomi M. & Kanade T. (1993), A multiple-baseline stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(4) pp. 353-363

[16] Lebegue X. & Aggarwal J. K. (1993), Significant line segments for an indoor mobile robot, *IEEE Transactions on Robotics and Automation* 9(6) pp. 801-815

[17] Smith S. M. (1992). *Feature based image sequence understanding.* PhD thesis, Robotics Research Group, Oxford University.

[18] Cox I. J., Roy S. & Hingorani S. L. (1995), Dynamic histogram warping of image pairs for constant image brightness, *IEEE International Conference on Image Processing,* Vol 2 pp. 366-369

[19] Smith P., Sinclair D., Cippola R. & Wood K. (1998). Effective corner matching. *Proceedings of the British Machine Vision Conference* Vol 2 pp. 545-556

[20] Ghita O., Mallon J. & Whelan P. F. (2001), Epipolar line extraction using feature matching, *Proceeding of the Irish Machine Vision & Image Processing (IMVIP)* pp. 87-97

[21] Anil K. J. and Dubes R. C. (1988) *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, New Jersey.