2008

Stereo correspondence using moment invariants

Prashan Premaratne
University of Wollongong, prashan@uow.edu.au

Farzad Safaei
University of Wollongong, farzad@uow.edu.au

Stereo Correspondence Using Moment Invariants

Prashan Premaratne and Farzad Safaei

School of Electrical, Computer & Telecommunications Engineering,
The University of Wollongong, North Wollongong 2522, NSW, Australia
prashan@uow.edu.au

Abstract. Autonomous navigation is seen as a vital tool in harnessing the enormous potential of Unmanned Aerial Vehicles (UAV) and small robotic vehicles for both military and civilian use. Even though, laser based scanning solutions for Simultaneous Location And Mapping (SLAM) is considered as the most reliable for depth estimation, they are not feasible for use in UAV and land-based small vehicles due to their physical size and weight. Stereovision is considered as the best approach for any autonomous navigation solution as stereo rigs are considered to be lightweight and inexpensive. However, stereoscopy which estimates the depth information through pairs of stereo images can still be computationally expensive and unreliable. This is mainly due to some of the algorithms used in successful stereovision solutions require high computational requirements that cannot be met by small robotic vehicles. In our research, we implement a feature-based stereovision solution using moment invariants as a metric to find corresponding regions in image pairs that will reduce the computational complexity and improve the accuracy of the disparity measures that will be significant for the use in UAVs and in small robotic vehicles.

1 Introduction

Stereo vision is a mechanism for obtaining depth information from digital images. The challenge in stereovision is how to find corresponding points in the left image and the right image, known as the correspondence problem. Once a pair of corresponding points is found the depth can be computed using triangulation. There are two prominent approaches to finding such corresponding pairs namely, area-based and feature-based techniques. In the area based techniques, every pixel in a designated area of one image is compared with the pixels in the same row of the other image. This is done with few constraints such as maximum disparity to avert any false matches. Some of the well-known techniques in this approach are Hierarchical Block Matching [1], Census [2], Correlation Matching [3-4] and Zitnick-Kanade (Cooperative Algorithm for Stereo Matching and Occlusion Detection) [5-6] algorithms. The feature-based methods rely on finding special features in corresponding pairs and may result in fewer depth values lowering the computational complexity.

Our approach is very much aimed at controlling small robotic vehicles using stereovision for depth calculation. This depth information will be used in control algorithms to detect and avoid obstacles. If this depth information is to be useful, they need to be estimated in realtime which requires any stereovision algorithm to be less computationally expensive.

With our great success in using moment invariants for recognizing hand gestures, moment invariants can be of great use in finding corresponding matching regions in stereo pairs [7-8]. Moment invariants are invariant to rotation, scale and shift and the rotation invariant property is especially beneficial to the stereo correspondence problem as any misalignment or non-flat ground conditions can create slightly rotated versions of any scene in any one of the cameras. In our approach, we rely on edge-corner detection algorithms such as Harris corner detection [8] to produce reliable feature points. This will result in fewer points of interest compared to area-based techniques [9-18]. An image can be separated to a collection of blocks and they can be marked as candidates or not depending on whether they occupy corners. These blocks can be matched with the help of moment invariants and disparity of the identified features can be simply calculated.

In this paper, section 2 details the general stereo matching approaches and the moment invariant based technique is presented in detail in section 3. This is followed by our experimental results and the conclusion.

2 Stereo Matching Approaches

In area based stereo matching, for a given pair of stereo images, the corresponding points are supposed to lie on the epipolar lines [19]. Area based techniques rely on the assumption of surface continuity, and often involve some correlation-measure to construct a disparity map with an estimate of disparity for each point visible in the stereo pair. Area based techniques produce much denser disparity maps, which is critical in obstacle detection and avoidance.

Since corresponding points are the images of the same real point in the taken scene projected into both pictures, we can assume that their surroundings in both pictures will be quite similar. Area-based methods use this similarity for corresponding points detection [9-13]. It is computed from the difference in local neighborhoods (usually a constant size square) of the points. Computing the similarity of two points is the elementary step in the method and cannot be accelerated. The main problem is to look for the corresponding point in the picture. The naive area-based algorithm chooses a point from the first image, and run through all the points in the second image to find its corresponding point. This inefficient process can be accelerated by restraining the search area to a specific region around the corresponding pixel by specifying a maximum disparity. The most efficient method is adapted from Epipolar Geometry known as epipolar constraint.

Area based methods are considered to be computationally expensive due to the exhaustive nature of the metrics being used. Sum of Squared Difference (SSD), Sum of Absolute Difference, Sum of Sum of Squared Difference (SSSD) and Cross Correlation based metrics use every pixel to calculate these metrics making them exhaustive.

Feature-based stereo matching techniques focus on local intensity variations and generate depth information only at points where features are detected. In general, feature-based techniques provide more accurate information in terms of locating

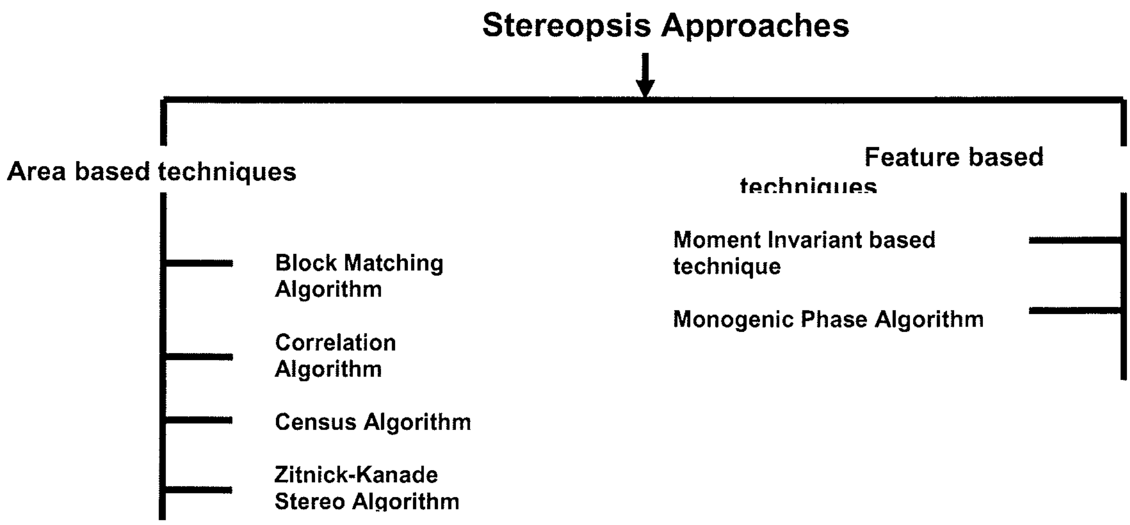

Fig. 1. Stereopsis Approaches

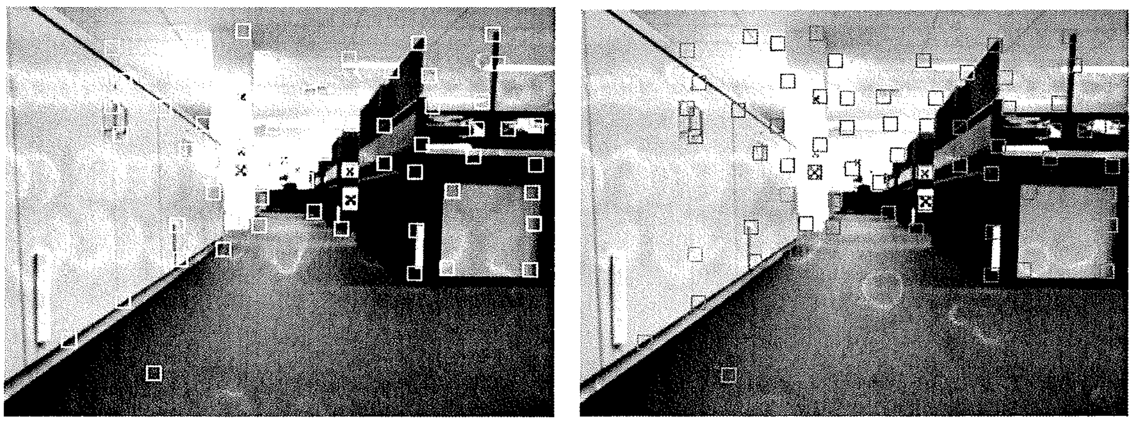

Fig. 2. Left and Right stereo images with 'corners' marked using Harris corner detection

depth discontinuities and thus achieve fast and robust matching. However they yield very sparse range maps and may have to undergo expensive feature extraction process. Edge elements, corners, line segments, and curve segments are features that are robust against the change of perspective, and they have been widely used in many stereovision work. Features such as edge elements and corners are easy to detect however, they may suffer from occlusion whereas line and curve segments require extra computation time, but are more robust against occlusion. Higher level image features such as circles, ellipses, and polygonal regions have also been used as features for stereo matching, these features are, however, restricted to images of indoor scenes. Nevertheless, feature based techniques are associated with computation of conjugate pairs with subpixel accuracy. They will also include object-dependent constraints in the solution of the correspondence problem such as 'corners' when using Harris corner detection algorithm.

3 Moment Invariant Based Stereo Matching

Using invariant moments to locate corresponding features in stereo pairs will be less computationally intensive as the number of block comparisons will depend on the disparity constraint as well as number of features. In our approach, 'corners' will be used as features as shown in Fig. 2. Then the left image is divided into 20x20 pixel blocks and will be marked with occupying features. Then the blocks containing features will be used to calculate the first 4 moments using equations 3 to 6. Even though the moment invariants can calculate upto 7 such moments, 4 moments will be adequate to uniquely represent a square. Using epipolar and disparity constraints, we can now evaluate the adjoining 20x20 pixel blocks in the Right image for matches for the blocks containing the features. The 'closeness' of these moments will be decided using a threshold that is dependent on the image scenery. When such blocks are identified, the simple depth calculation formula can be used to calculate the depth to the identified feature as follows:

$$d = b\frac{f}{D}$$

Where d is the depth the object from the camera plane, f being the focus of the camera, D the disparity a b is the baseline distance. Here the underlying assumption is that the epipolar lines run parallel to the image lines, so that corresponding points lie on the same image lines.

3.1 Moment Invariants

Moment invariants algorithm has been known as one of the most effective methods to extract descriptive feature for object recognition applications. The algorithm has been widely applied in classification of aircrafts, ships, ground targets, etc [20, 21]. Essentially, the algorithm derives a number of self-characteristic properties from a binary image of an object. These properties are invariant to rotation, scale and translation. Let $f(i,j)$ be a point of a digital image of size $M \times N$ ($i = 1,2, \ldots, M$ and $j = 1,2, \ldots, N$). The two dimensional moments and central moments of order $(p + q)$ of $f(i,j)$, are defined as:

$$m_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} i^p j^q f(i,j) \tag{1}$$

$$U_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} (i-\bar{i})^p (j-\bar{j})^q f(i,j) \tag{2}$$

$$\text{Where } \bar{i} = \frac{m_{10}}{m_{00}} \qquad \bar{j} = \frac{m_{01}}{m_{00}} \tag{3}$$

$$\phi_1 = \eta_{20} + \eta_{02} \tag{4}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2 \tag{5}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{6}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{7}$$

Where η_{pq} is the normalized central moments defined by:

$$\eta_{pq} = {U_{pq}} \Big/ {U_{00}^{r}} \; .$$

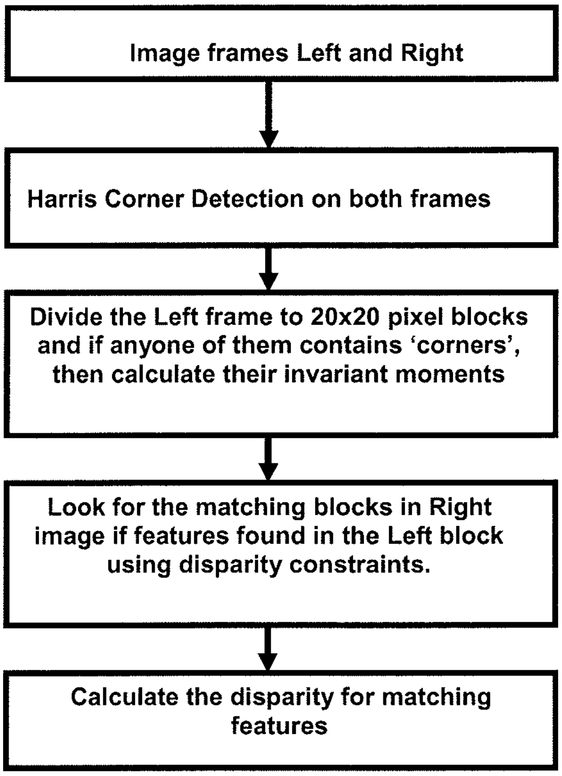

Fig. 3. Summary of the proposed technique

If certain blocks do not contain any features, the search area in the other image (Right) can be restricted using maximum disparity constraint. This will further cut down the computational requirements as opposed to area-based correlation techniques.

4 Experimental Results

We ran the proposed algorithm on 100 stereo pairs generated using a BumbleBee™ stereo rig mounted on a mobile robot producing 320x240 pixel frames. The path the robot covered had special markers to provide feature points as most of the cubicles of the indoor setup had monotonous flat featureless walls and partitions. This was followed by running correlation based stereo matching algorithm using SSD and SAD metrics. The results are shown in Fig. 4. The processing system comprised of a Pentium 4 system running at 2GHz with 2GB of RAM. The system was capable of processing 10 frames per second for both moment invariant based approach and correlation based technique using SAD metric however, it only managed 5 frames per second using SSD. This was expected as the SSD involved more computational complexity compared to SAD. However, it should be pointed out that the proposed approach computed handful of disparity values similar to Fig. 2. We also used the Zitnick-Kanade algorithm to estimate the disparity even though the results are not presented due to the inability of the algorithm to run anywhere near realtime with our modest processing power. Since this algorithm expects multiple iterations to refine its estimates, it is simply not useful for realtime applications that we are interested in. Fig. 3 summarizes the major steps in the proposed algorithm.

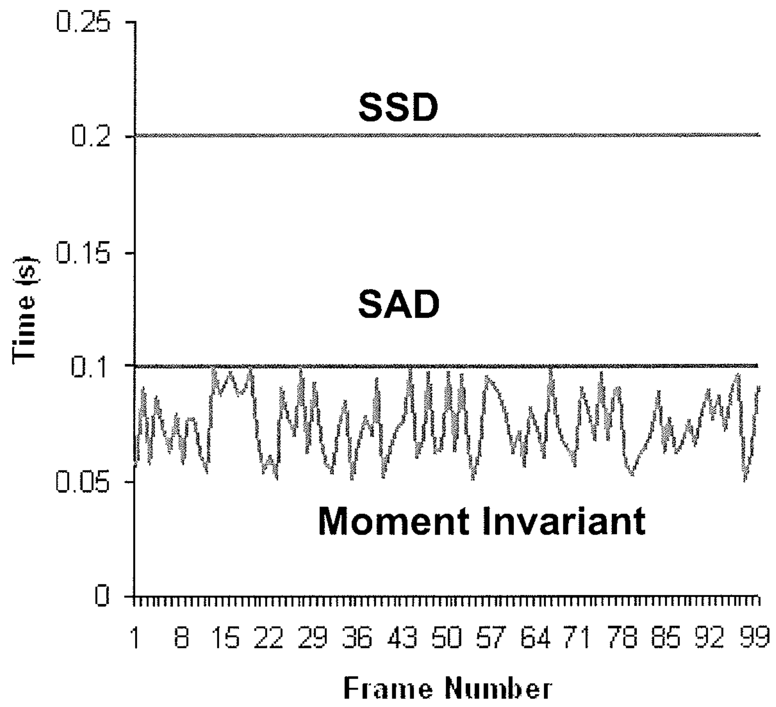

Fig. 4. Comparison of Moment Invariant technique with correlation method using SSD and SAD metrics

The major advantage of local approaches presented here is speed and suitability for hardware implementation. Global optimization algorithms commonly require 2 to 3 orders of magnitude more time than even the software implementations of local methods.

5 Summary

In many respects feature based algorithms are established as the most robust way to implement stereo vision algorithms for the industrial-type stereo problems. The advantages offered by using features are that feature-based representations contain desirable statistical properties and provide algorithmic flexibility to the programmer. The flexibility is that algorithmic constraints can be applied explicitly to the data structures rather than implicitly as with area based correlation techniques. In particular, the use of 'corners' leads to algorithms which are as locally accurate as the precision to which the edges can be extracted.

Even though, feature-based techniques do not produce denser disparity map, their values are more accurate than the area-based techniques. Some of the reasons for this is that presence of shadows produce erroneous results; some surfaces were non-uniformly reflecting light from; backgrounds are usually flat single-colored surfaces and some parts of the first image were occluded in one of the images.

We managed to demonstrate that feature-based techniques relying on moment invariants for matching can process a frame in the order of tenths of seconds in software implementations. This implies that the algorithm can comfortably reach higher video rates using DSP and FPGA implementations [22]. At the moment, there is no technique for achieving simultaneously the high quality range obtained from global optimization with the fast run-times of local schemes.

References

1. Koschan, Q., Rodehorst, V., Spiller, K.: Color Stereo Vision Using Uierarchical Block Matching and Active Color Illumination. In: Proc. 13th Int. Conf. Pattern Recog., vol. 1, pp. 835–839 (1996)
2. Zabih, R., Woodfill, J.: Non-parametric Local Transformers for Computing Visual Correspondence. In: Third Eurpean Conf. Computer Vision (1994)
3. van Beek, J.C.M., Lukkien, J.J.: A Parallel Algorithm for Stereo Vision Based on Correlation. In: Proc. 3rd Int. conf. High Performance Computing (1996)
4. Hirschm utller, H., Innocent, P.R., Garibaldi, J.M.: Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. Int. Journal of Computer Vision 47(1/2/3), 229–246 (2002)
5. Zitnick, C., Kanade, T.: A Cooperative Algorithm for Stereo Matching and Occlusion Detection. Tech. Report CMU-RI-TR-99-35, Robotics Institute, Carnegie Mellon University (October 1999)
6. Zitnick, C., Kanade, T.: A Cooperative Algorithm for Stereo Matching and Occlusion Detection. IEEE Trans. Pattern Analysis and Machine Intellig. 22(7), 675–684 (2000)
7. Premaratne, P., Nguyen, Q.: Consumer electronics control system based on hand gesture moment invariants. IET Computer Vision 1(1), 35–41 (2007)
8. Premaratne, P., Safaei, F., Nguyen, Q.: Moment Invariant Based Control System Using Hand Gestures: Book Intelligent Computing in Signal Processing and Pattern recognition. In: Book Series Lecture Notes in Control and Information Sciences, vol. 345, pp. 322–333. Springer, Heidelberg (2006)
9. Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In: Proc. 4th Alvey Vision Conference, pp. 147–151 (1988)
10. Di Stefano, L., Marchionni, M., Mattoccia, S., Neri, G.: A Fast Area-Based Stereo Matching Algorithm. In: 15th IAPR/CIPRS International Conference on Vision Interface, Calgary, Canada, May 27-29 (2002)

11. Fusiello, A., Roberto, V., Trucco, E.: Experiments with a New Area-Based Stereo Algorithm. In: International Conference on Image Analysis and Proceedings, Florence (1997)
12. Barnard, S.T., Thompson, W.B.: Disparity Analysis of Images. IEEE Trans. of PAMI PAMI-2, 4 (1980)
13. Hannah, M.J.: Bootstrap Stereo. In: Proc. Image Understanding Workshop (1980)
14. Hannah, M.J.: SRI's Baseline Stereo System. In: Proc. of DARPA Image Understanding Workshop, pp. 149–155 (1985)
15. Hannah M.J.: A System for Digital Stereo Image Matching. In: Photogrammatic Engineering and RemoteSensing, pp. 1765–1770 (1989)
16. Burt, P., Julesz, B.: Modifications of the Classical notion of Panum's Fusional Area. Perception 9, 671–682 (1980)
17. Lane, R.A., Thacker, N.A., Seed, N.L.: Stretch-Correlation as a Real-Time Alternative to Feature Based Stereo Matching Algorithms. Image and Vision Comp. Journal 12(4) (May 1994)
18. Lane, R.A., Thacker, N.A., Seed, N.L., Ivey, P.A.: A Stereo Vision Processor. In: Proc. of IEEE Custom IntegratedCircuits Conference (1995)
19. Weng, J.: Camera calibration with distortion models and accuracy evaluation. IEEE Trans. Patt. Anal. Machine Intel. 14, 965–980 (1992)
20. Premaratne, P.: ISAR ship classification, An alternative approach. CSSIP-DSTO Internal Publication, Australia (March 2003)
21. Zhongliang, Q., Wenjun, W.: Automatic ship classification by superstructure moment invariants and two-stage classifier. In: ICCS/ISITA 1992 Comm. on the Move, pp. 544–547 (1992)
22. van der Horst, J., van Leeuwen, R., Broers, H., Kleihorst, R., Jonker, P.: A Real-Time Stereo SmartCam, Using FPGA, SIMD and VLIW. In: Proc. of the 2nd Workshop on Applications of Computer Vision (May 12, 2006)