Embedded Stereo Vision Application for Automotives

[1]Rashmi.A, [2]Roopashree, [3]Bindu N.S
[1,2]BE, 6th SEM Student, Dept.of E&C
[3]Assistant Professor, Dept.of E&C
[1, 2,3]Vidyavardhaka college of Engineering, Mysore, India
[1]RASHMI94A@gmail.com
[2]roopashree9535@gmail.com
[3]bindu.ns5@gmail.com

***Abstract* — This paper deals with perception of 3D modules which are essential for safe operation of autonomous system. The challenge of fast stereo matching for embedded system is considered in this paper. Robotic and industrial applications do not permit the use of sophisticated stereo vision algorithm. The strength and weaknesses of different matching approaches have been analyzed and a well suited solution has been found in sparse census transform which halves the processing time with nearly unchanging matching quality. .system is robust easy to parameterize and offers high flexibility it also achieves high performance on several including resource limited system without losing the quality of stereo matching quality and processing time is compared to other algorithms on the middle bury stereo evaluation website reaching the middle quantity and top performance rank. Besides the detection of false positive and wrong matches are highly reduced due to the computation and analysis of dedicated confidence value.**

Keywords—stereo vision, stereo matching, epipolar geometry, hamming distance, intensity, pixel, disparity, rectified image

IJERT

I.INTRODUCTION

For modern mobile robot platforms, dependable and embedded perception modules are important for successful autonomous operations like navigation, visual surveying, or grasping. Especially 3D information about the area around the robot is crucial for reliable operations in human environments. State-of-the-art sensors such as laser scanners or time-of-flight methods deliver 3D information, which is either rough or has low resolution with respect to time and space. Stereo vision is a technology that is well suited for delivering a precise description within its field of view. Stereo is purely a passive technology that primarily uses only two cameras and a processing unit to do the matching and 3D reconstruction . Advanced Driver Assistance Systems (ADAS) for intelligent vehicles will mainly rely on dependable and reliable 3D be highly desirable.

Stereo Vision Algorithm mainly has three steps

- Rectification
- Stereo Matching Algorithm
- Post Processing

II.RECTIFICATION

Rectification ensures the compact implementation of stereo matching algorithm. It is very necessary to rectify captured images so that images are in epipolar geometry. Epipolar geometry of stereo vision is that when two cameras view a 3D scene from two distinct points, there are number of geometric relations between the 3D points and their projections on 2D images

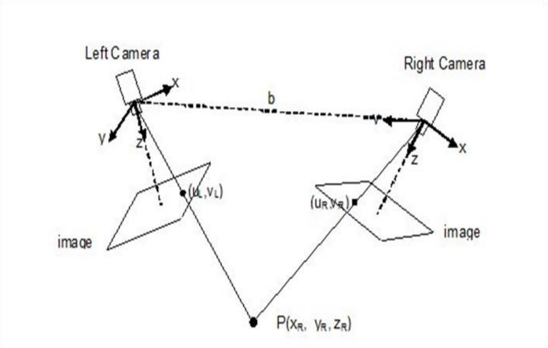


Fig 1.Epipolar geometry

This can be illustrated by Classical stereo vision which uses a stereo camera setup built up of two cameras, called a stereo camera head, mounted in parallel. It captures a synchronized stereo pair consisting of the left camera's and the right camera's image. A typical classical stereo process is shown in Fig.1 the distance between both cameras is called the baseline. Once the correct disparity for a pixel is found, it can be used to calculate the orthogonal distance between one camera's optical center and the projected scene point with

$$Z = b. f/d \qquad (1)$$

Where d is the disparity, b the baseline and f the camera's focal length. If 3D data should be given in camera coordinates, (2) can be used, where K is the camera calibration matrix, the pixel is given in homogeneous coordinates and is calculated with (1). K and f have to be determined by camera calibration which is essential for fast stereo matching. On the one hand, camera lens distortion can be removed, and on the other hand, the images can be rectified. Rectified images fulfill the epipolar constraint, which means that corresponding pixel rows share the same v-coordinate, so the search for corresponding pixels is reduced to a search along one pixel row instead of through the whole image. For this work it is always assumed that the cameras are calibrated and the stereo image pairs are rectified.

III.STEREO MATCHING ALGORITHM

Stereo Matching Algorithm is achieved by deploying census transform, which is highly robust and performs matching function by intensity variation within camera images independent of offset variation. Offset variation here refers to displacement i.e distance between two cameras. For optimal industrial application it is achieved by deploying Sum of Absolute Differences (SAD). Census transform enables high accuracy when compared to non real time algorithm.

Census transform consists of a comparison function ξ which is used to compare the center pixel intensity value i1 and center pixel intensity value i2 of the neighborhood region.

$$\xi(i_1, i_2) = \begin{cases} 1 & i_1 > i_2 \\ 0 & i_1 \leq i_2 \end{cases} \qquad (3)$$

In order make the cost of stereo vision algorithm to be less we calculate Hamming distance. For the cost function, the Hamming distance is calculated over the bit vectors. For a resource aware implementation it is necessary to reduce the computational complexity of the Census Transform. Here, we are using a so called sparse computation, where the hamming distance is not computed for all bits within the vector, but only for a dedicated amount of values. This way, the images are sub-sampled in a raster fashion.

IJERT

Requirements of embedded real time stereo matching A big advantage of stereo sensors is that they deliver a huge number of 3D points with a single measurement. This is what makes these so attractive for robotic applications. Of course, to ensure fast reactions of a robot to environmental changes, the sensor has to deliver data at high frame rates and low latencies. A minimum of 10 fps should be achieved in any case and the algorithm has to be suitable for real-time applications, which means the calculation has to be finished within that time frame and has to be independent from the actual scene. An area-based Census correlation algorithm fulfills all these requirements. And also use of certain neighborhoods allows the avoidance of double calculation and reduces the total number of comparisons. The mask configurations obtain a rather irregular structure which is very unfavorable for performance-optimized implementations on modern processors. The reliability of 3D data is also important. For instance, only 3D points with a high probability of correctness should be delivered and used for navigation. To fulfill this demand a confidence and a texture map are calculated which gives an opportunity to identify and filter uncertain matches and texture less areas Mobile robotic platforms have to often deal with different lighting conditions, so the matching algorithm has to be very robust in terms of different scene illumination of the stereo cameras.

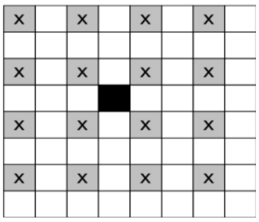
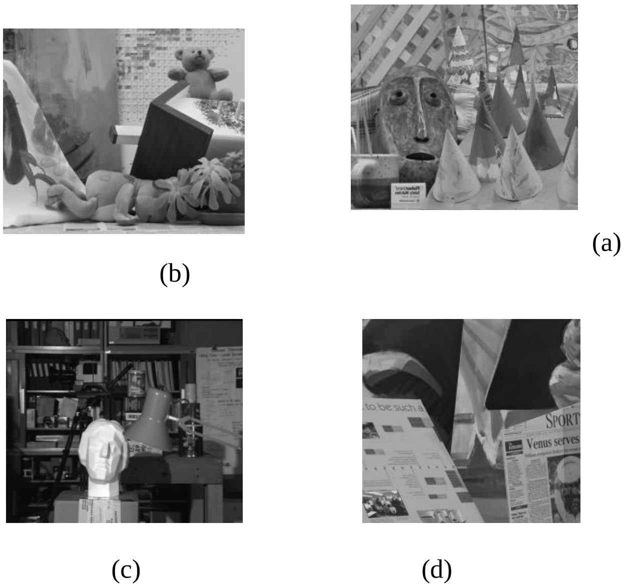


Fig.2 Computation of the Sparse Census Transform for sparse factor = 4.

The approach used in this work, keeps the mask size as large and symmetric as possible by using only every second pixel and every second row of the mask for the Census transform, as shown in Fig. 2 for an 8x8 mask. The filled squares are the pixels used for the Census and the sparse Census transform. Avoiding the double comparisons here is not the key to minimize the processing time, but it is assumed that large sparse Census masks perform better than small normal (dense) Census masks with the same weight of the resulting bit strings. Thus it is anticipated that sparse 16x16 Census performs better than 8x8 normal Census, where both have a bit string weight of 64 and thus need the same processing time.

To depict the difference in accuracy when reducing either the block size for the Census Transform and when increasing the sparse factor, Figure 3 presents the averaged accuracy for the Teddy, Cones, Tsukuba and Venus stereo images from the middlebury dataset [5], [6] (see Figure 3).



(a)

(b)

(c)

(d)

Fig. 3 Middlebury dataset: Teddy, Cones, Tsukuba and Venus images.

Here, the x-axis presents the number of hamming distance computations required for the cost function for the comparison of each block, while the y-axis presents the accuracy in pixel percentage where the resulting disparity value is within one pixel when compared to the ground truth.

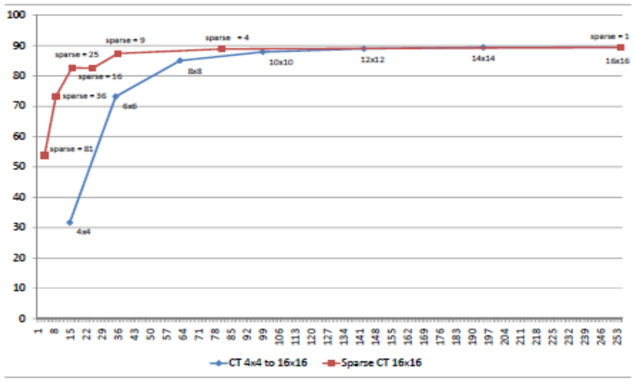


Fig. 4. Comparison of the accuracy achieved for different block sizes for the Census Transform and different sparse factors for block size 16 x16.

The results show, that until a sparse factor of 16, where only one pixel out of 16 is selected for the Census Transform, there is only a minor drop in accuracy. Reducing the block size leads to a far higher reduction in accuracy when compared to the total computational complexity. The detailed results for the accuracy emerging from the Sparse Census Transform are presented in Table I. Here, a sparse factor of 4 results only in a drop in accuracy of just 0.55%. Even if a sparse factor of 9, i.e. using every third pixelin x and y direction, would result in an even higher reduction of computational complexity, the reduction in accuracy is already 2.01% which is about 4 more loss than at factor 4. Thus, we are using sparse factor 4 in our work

.

Sparse	Teddy	Cones	Tsukuba	Venus	Age
81	55.62	40.85	55.66	62.94	53.77
36	75.51	68.18	70.30	79.16	73.29
25	83.43	82.38	78.58	86.34	82.68
16	83.59	82.36	78.63	85.71	82.57
9	86.00	89.88	83.71	89.18	87.43
4	87.19	91.55	85.94	90.89	88.89
1	86.74	91.70	87.07	92.24	89.44

TABLE I: ACCURACY OF THE SPARSE CENSUS TRANSFORM FOR DIFFERENT SPARSE FACTORS, USING THE TEDDY, CONES, TSUKUBA AND VENUS DATASETS FOR CENSUS BLOCK SIZE 16x16.

IV. POST PROCESSING

For the post-processing, we are using parabola-fitting for the sub-pixel refinement. Here, the cost values' absolute minimum matching costs and the neighboring costs are interpolated and a two bit sub-pixel refinement is implemented: Furthermore, occluded regions are detected using a Left/Right Consistency Check. Here, the disparity values centered on the left and right camera images are calculated and compared. Pixels showing a depth deviation of more than one pixel between both disparity maps are disregarded for the further computation. Industrial applications highly enforce the reduction of false positives within the depth map. Therefore, we implemented the computation of a confidence value. Here, the matching costs for the disparity range of each pixel are analyzed and both the absolute minimum, i.e., the global minimum, as well as the second lowest local minimum are being searched for. Here, the difference of these two matching costs determine the possibility that another disparity value also could have been a good match, or not, as depicted in Figure 4. If the difference is large, all other local minima can be considered to be rather bad matches. However, if the difference is low, there is a high possibility that other disparity levels would be good matches too and the difference is mainly caused by camera noise. A typical example for this situation is a chess board, where all fields look the same and the stereo matching algorithm cannot compute trustworthy results. Even if the chess board is a rather abstract example, this is quite often the case, due to the popularity of repeated textures in human design. Furthermore, un textured surfaces also show only slight differences in the disparity ranges' matching costs and also allow for a removal of uncertain values using our confidence value.

IJERT

Applications of stereo vision

- Stereo vision finds its main applications in people tracking , Surgeries , 3D underwater mosaicking
- Stereo vision system is mainly used in Advance Driver Automated System(ADAS)
- They find extensive applications in Robotics
- They play a major role in extraction of information in aerial surveys
- Stereo vision is used in target recognition of mobile robots.
- They are used in Forensics i.e crime scenes , Traffic accidents ,
- Mining and Mine face measurements
- Civil engineering and structure monitoring.
- Collision avoidance
- Manufacturing and process monitoring.

Advantages of stereo vision

- Robustness
- Gives a very dense depth
- Use to calculate shape of objects
- Human motion detection is possible instead of using sensors for it.

Disadvantages of stereo vision

- The system must be pre calibrated
- Has to be used in indoor environment
  Shadow and sunlight present in experimental area makes difficult in distance calculation.

Limitations of stereo vision

- Correspondence problem-
- Calibration problem
- Synchronization problem
- Shadow problem

- Sunlight problem

V. CONCLUSION

The design of a real-time stereo vision system suitable for automotive and industrial applications leads to very demanding requirements. The typical limitations of the computational resources in embedded systems require a stereo matching algorithm that offers a low complexity while the drop in accuracy has to be at a minimum level. Here, the implemented Sparse Census Transform reduces the computational complexity by a factor of 4, while the accuracy is still in the same levels as for the dense computation. The reduction of noise and false positives is another very challenging requirement. While most stereo vision algorithms depend on a left/right consistency check for the removal of occluded regions only, we implemented the computation of a confidence value that allows for a more dependable removal of mismatched areas. The concept can also be applied on upcoming multi-core DSP models with up to 6 DSP cores on-chip. This will enable future systems with higher frame rates at higher image resolution and still at reasonable energy requirements.

REFERENCES

[1] Kristian Ambrosch and Wilfried Kubinger. Accurate hardware-based stereo vision. Computer Vision and Image Understanding, In press, to appear in 2010, doi:10.1016/j.cviu.2010.07.008.

[2] Martin Humenberger, Christian Zinner, Michael Weber, Wilfried Kubinger, and Markus Vincze. A fast stereo matching algorithm suitable for embedded real-time systems. Computer Vision and Image Understanding, In press, to appear in 2010, doi:10.1016/j.cviu.2010.03.012.

[3] Bahador Khaleghi, Siddhant Ahuja, and Jonathan Wu. An Improved Real-Time Miniaturized Embedded Stereo Vision System (MESVS-II). In Proceedings of the 2008 Conference on Computer Vision and Pattern Recoginition Workshops, 2008.

[4] Li Mingxiang and Jia Yunde. Stereo Vision System on Programmable Chip (SVSoC) for Small Robot Navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,2006.

[5] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision, 47(1–3):7–42, 2002.

[6] Daniel Scharstein and Richard Szeliski. High-[3] Bahador Khaleghi, Siddhant Ahuja, and Jonathan Wu. An Improved Real-Time Miniaturized Embedded Stereo Vision System (MESVS-II). In Proceedings of the 2008 Conference on Computer Vision and Pattern Recoginition Workshops, 2008.

[4] Li Mingxiang and Jia Yunde. Stereo Vision System on Programmable Chip (SVSoC) for Small Robot Navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,2006.

[5] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision, 47(1–3):7–42, 2002.

[6] Daniel Scharstein and Richard Szeliski. High-Accuracy Stereo Depth Maps Using Structured Light. In Proceedings of the 2003 Conference on Computer Vision and Pattern Recoginition, 2003.

[7] Texas Instruments. TMS320C6472 Fixed-Point Digital Signal Processor, 2009. Lit. Number SPRS612D.

[8] Ramin Zabih and John Iseling Woodfill. Non-parametric Local Transforms for Computing Visual Correspondence. In Proceedings of the 3rd European Conference on Computer Vision, 1994.

[9] Christian Zinner, Wilfried Kubinger, and Richard Isaacs. Pfelib: A performance primitives library for embedded vision. EURASIP Journal on Embedded Systems, 2007(1):14, 1 2007.

IJERT