PIXEL BASED CORRELATION OF SAD AND SSD ALGORITHMS FOR DISPARITY MAP USING MULTIVIEW IMAGES

[1]CHIRAG D. PATEL [2] KAPIL S. RAVIYA

[1]M.E. (E.C.) Student, Dept. of Electronics & Communication Engineering/C.U.Shah College of Engineering and Technology/Wadhwan City/Gujarat/India..

[2] Lecturer /C.U.Shah College of Engineering and Technology/Wadhwan City/Gujarat/India.

Chirag.beec@yahoo.com,raviyakapil@yahoo.co.in

ABSTRACT: **Stereo vision systems aim at reconstructing 3D scenes by matching two or more images taken from slightly different viewpoints. The main problem that has to be solved is the identification of corresponding pixels, i.e. pixels that represent the same point in the scene and reconstruction of 3-D coordinates form the 2-D coordinate. In the area of computer vision, this correspondence problem is the hardest and has been studied extensively. Stereo Matching is not difficult to understand in theory but it is not easy to solve in practice. Broad categories of stereo algorithms are use for finding the disparity. Algorithm identified based upon differences in image geometry, matching primitives, and the computational structure used. The matching approach to solve the supposed correspondence problem in static, binocular stereo vision has intrinsic limitations. Specifically, matching is of no use in occluded areas because there is nothing to match in those regions. Other kinds of problems, like large regions of the image with a very uniform surface result in erroneous matching in almost every case. Disparity in such regions can be determined with a different approach, based on well known details and principles of stereo vision We describe the performance of stereo correspondence algorithms and the techniques using the various performance parameter like PSNR,MSE,RMS error etc used for acquiring our image data sets and ground truth estimates.**

Key Words: STEREO MATCHING, DISPARITY, PSNR, MSE RMS Error

1. INTRODUCTION:

Computer vision tries to copy the way how human beings perceive visual information by means of using cameras acting as eyeballs and computers to process the information in an intelligent way as does the human brain. The cameras are designed based on the knowledge of the eyeball operation so that some comparisons can be established: the shutter corresponds to the human iris, the camera lenses and detector corresponds to the human lens and retina.

The ability of humans to perceive the three-dimensional world from the two-dimensional projections on the retina is both important and fascinating. The ability is partly dependent on perspective effects, that is, the fact that three-dimensional objects look different from different viewpoints. In human vision our two eyes have slightly different perspectives, but perhaps even more important are perspective changes due to movement.

It should also be noted that occlusion, size and other effects are very important for our own three-dimensional perception. How depth information can be inferred from a two or more 2 D images taken from different viewpoints. Stereo vision systems aim at reconstructing 3D scenes by matching two or more images taken from slightly different viewpoints. The main problem that has to be solved is the identification of corresponding pixels, i.e. pixels that represent the same point in the scene. In the area of computer vision, this correspondence problem has been studied extensively. Stereo matching is basic to obtaining depth information from a pair of images and analyzing the 3-dimensional (3D) structure of objects. Because the process is easily affected by local environments, stereo matching is seemed as an ill-posed problem. Many sophisticated studies have explored and solved such problems and decreased

The problem arises due to mismatching between pairs of stereo images caused by noise from texture less regions, depth discontinuities, occlusion, etc. In this paper, we compute the SAD(sum of difference),SSD(sum of squared difference)algorithm use and compare all results with Ground truth and finding PSNR,MSE RMS Error.

2. STEREO MATCHING

The problem of reconstructing a three dimensional scene from several viewpoints was first investigated in the fields of aerial photography and human stereopsis. Until relatively recently, the scene reconstruction problem was typically treated as a matching problem where the objective was to match points or features between two or more images. Having obtained a match, the three dimensional position of a point could be determined by triangulation assuming the camera positions were known.

The matching of image points is performed by comparing a region in one image, referred to as the reference image, with potential matching regions in the other image and selecting the most likely match based on some similarity measure. The resulting scene estimate is then invariably represented using a depth-map relative to the reference camera.

As an example of the stereo matching process, consider estimating the three dimensional position of a point P shown in Fig. 1. By correctly matching this point between the two images, the relative shift or displacement of the point can be used to calculate the depth of the point.

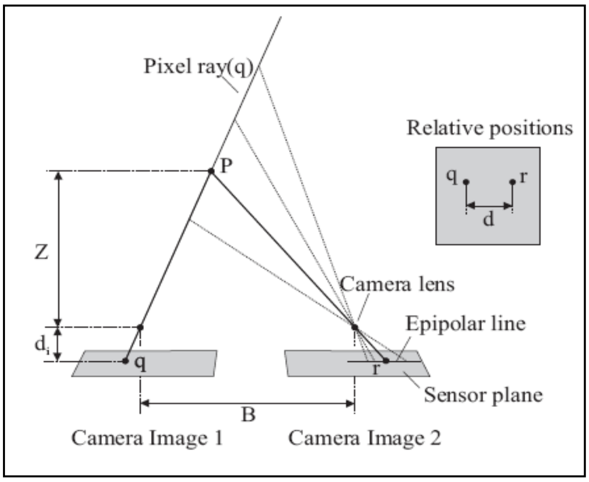


Figure 1 Demonstration of disparity

As an example consider the point P in Fig. 1. This has image coordinates (x, y) as viewed from camera 1 and image coordinates (x+ d, y) when viewed from camera 2.

By correctly matching this point between the two images the relative shift, or disparity, d, of the point can found. This can then be used to calculate the depth of the point. If all cameras have the same focal length, are parallel to each other, and located on the same plane, the magnitude of this disparity is related to the depth, Z,

$$Z = \frac{B \cdot d_i}{d} \quad \text{-----} \quad (1)$$

Where B is the baseline distance between two cameras and di is the distance of the image plane behind the principal point. One problem with this approach is that it is difficult to determine matches reliably because of ambiguities and occlusions. To reduce the number of ambiguities, regions in the image are matched in order to improve the reliability of matching, instead of individual pixels. This is based on the assumption that nearby pixels are likely to have originated from a similar depth.

However, difficulties arise in regions which do contain several depths, because the observed region will appear different between the various cameras. The spatial resolution of the reconstructed scene will also be reduced in proportion to the size of the matching region used.

Another difficulty with traditional stereo matching is which surfaces that are visible within the reference image may be occluded or hidden from view in one or more of the other images. In this situation false matches will occur as a true match does not exist. To avoid these problems occluded regions must be identified. Matches must then only be formed with images where the corresponding surfaces are visible. Identifying these surfaces is difficult with traditional stereo matching, since the matching is performed directly in 2D image space where occlusions cannot be properly modeled.

2.1 SOLVING THE CORRESPONDENCE PROBLEM

The correspondence problem consists in finding correct point-to point correspond-dance between images or models. If we can identify the same 3D point in both views we can estimate its 3D coordinates. Accurately solving the correspondence problem is the key to accurately solving the Stereo Vision problem.

.

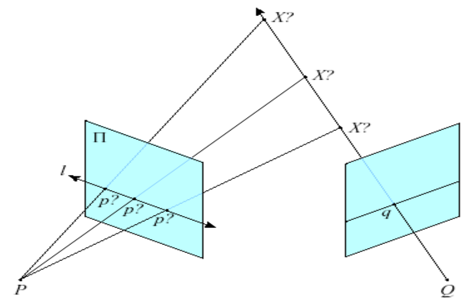


Figure 2: The geometry of the epipolar constraint:

In image point p corresponding to image point q must lie on epipolar line l, which is intersection of image plane with plane spanned by q and centers of projection P, Q

The fundamental hypothesis behind multi-image correspondence is that the appearance of any sufficiently small region in the world changes little from image to image. In general, appearance might emphasize higher-level descriptors over raw intensity values, but in its strongest sense, this hypothesis would mean that the color of any world point remains constant from image to image. In other words, if image point p and q are both images of some world point X, then the color values at p and q are equal.

This color constancy (or brightness constancy in the case of grayscale images) hypothesis is in fact true with ideal cameras if all visible surfaces in the world are perfectly diffuse (i.e., Lambertian). In practice, given photometric camera calibration and typical scenes, color constancy holds well enough to justify its use by most algorithms for correspondence

The geometry of the binocular imaging process also significantly prunes the set of Possible correspondences, from lying potentially anywhere within the 2D image, to lying necessarily somewhere along a 1D line embedded in that image[2][3]. Suppose that We are looking for all corresponding image point pairs (p, q) involving a given point q (Figure 2). Then we know that the corresponding world point X, of which q is an image, must lie somewhere along the ray through q from the center of projection Q. The image of this ray Qq in the other camera's image plane ∏ lies on a line l that is the intersection of ∏ with the plane spanned by the points P, Q and q. Because X lies on ray Qq, its projection p on ∏ must lie on the corresponding epipolar line l.(When corresponding epipolar lines lie on corresponding scan lines, the images are said to be rectified; the difference in coordinates of corresponding image points is called the disparity at those points.) This observation, that given one image point, a matching point in the other image must lie on the corresponding epipolar line, is called the epipolar constraint. Use of the epipolar constraint requires geometric camera calibration, and is what typically distinguishes stereo correspondence algorithms from other, more general correspondence algorithms.

Based on color constancy and the epipolar constraint, correspondence might proceed by matching every point in one image to every point with exactly the same Color in its corresponding epipolar line. However, this is obviously awed: there would be not only missed matches at the slightest deviation from color constancy, but also potentially many spurious matches from anything else that happens to be the same color. Moreover, with real cameras, sensor noise and finite pixel sizes lead to additional imprecision in solving the correspondence problem. It is apparent that color constancy and the epipolar constraint are not enough to determine correspondence with sufficient accuracy for reliable triangulation. Thus, some additional constraint is needed in order to reconstruct a meaningful three-dimensional model.
Marr and Poggio proposed two such additional rules to guide binocular correspondence[5][6]:

Uniqueness, which states that “each item from each image may be assigned at most one disparity value,”

Continuity, which states that “disparity varies smoothly almost everywhere."

In explaining the uniqueness rule, Marr and Poggio specified that each item corresponds to something that has a unique physical position, and suggested that detected features such as edges or corners could be used. They clearly cautioned against equating an item with a gray-level point, describing a scene with transparency as a contraindicating example. However, this latter interpretation, that each image location be assigned at most one disparity value, is however very prevalent in practice; only a small number of stereo algorithms attempt to find more than one disparity value per pixel. This common simplification is in fact justifiable, if pixels are regarded as point samples rather than area samples, under the assumption that the scene consists of opaque objects: in that case, each image point receives light from, and is the projection of, only the one closest world point along its optical ray.
In explaining the continuity rule, Marr and Poggio observed that matter is cohesive, it is separated into objects, and the surfaces of objects are generally smooth compared with their distance from the viewer. These smooth surfaces, whose normals vary slowly, generally meet or intersect in smooth edges, whose tangents vary slowly. When projected onto a two-dimensional image plane, these three dimensional features result in smoothly varying disparity values almost everywhere in the image, with only a small fraction of the area of an image composed of boundaries that are discontinuous in depth. In other words, a reconstructed disparity map can be expected to be piecewise smooth, consisting of smooth surface patches separated by cleanly defined, smooth boundaries.
These two rules further disambiguate the correspondence problem. Together with color constancy and the epipolar constraint, uniqueness and continuity typically provide sufficient constraints to yield a reasonable solution to the stereo correspondence problem.

3. STEREO MATCHING ALGORITHMS

Stereo Matching Algorithms like SAD(sum of absolute difference), SSD (sum of squared difference) use for finding the disparity use the stereo images .

3.1.SAD(SUM OF ABSOLUTE DIFFERENCE)

This algorithm is based on pixel based approach is use for finding the disparity. Sum of absolute differences (SAD) is an algorithm for measuring the similarity between image blocks. It works by taking the absolute difference between each pixel in the original block and the corresponding pixel in the block being used for comparison. These differences are summed to create a simple metric of block If the left and right images exactly match, the resultant will be zero

Disparity measurement

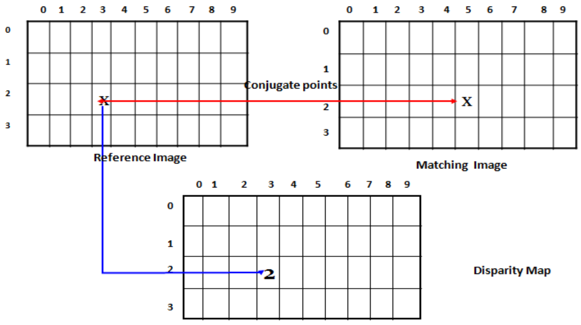


Figure 3: Disparity measurement :

The sum of absolute differences may be used for a variety of purposes, such as object recognition, the generation of disparity maps for stereo images, and motion estimation for video compression

Equation

$$\sum_{(i,j)\in W} |I_1(i,j) - I_2(x+i, y+j)|$$

Steps for finding the disparity

1) Take stereo image (Take two image for the same object from the different view)

2) Select the image pixel I1(i, j) in the first image

3) Finding matching pixel or correspondence pixel in second image

4) find the absolute difference between them (its called disparity)

3.2.SSD **(SUM OF SQUARED DIFFERENCE)**

SSD (sum of squared difference) is an algorithm for measuring the similarity between image blocks. It works by taking sum of squared difference between each pixel in the original block and the corresponding pixel in the block being used for comparison. These differences are summed to create a simple metric of block If the left and right images exactly match, the resultant will be zero

EQUATION

$$\sum_{(i,j)\in W} (I_1(i,j) - I_2(x+i, y+j))^2$$

Steps for finding the disparity

1) Take stereo image (Take two image for the same object from the different view)

2) Select the image pixel I1(i, j) in the first image

3) Finding matching pixel or correspondence pixel in second image

4) Find the sum of squared difference between them (its called disparity)

4. Evaluation Methodology

Quality measures are computed with known ground truth data:

MSE

The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR)are the two error metrics used to compare image compression quality.

The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error.

The lower the value of MSE, the lower the error.

$$MSE = \frac{\sum_{M,N}[I_1(M.N) - I_2(M.N)]^2}{M * N}$$

PSNR

To compute the PSNR, the block first calculates the mean-squared error using the following equation:

In the previous equation *M* and *N* are the number of rows and columns in the input images, respectively. Then the block computes the PSNR using the following equation:

$$PSNR = 10\ \log_{10}\frac{R^2}{MSE}$$

R= size of the image.

RMS (root-mean-squared) error (measured in disparity units) between the computed disparity map dC(x, y) and the ground truth map in fig (4) dT (x, y), i.e.

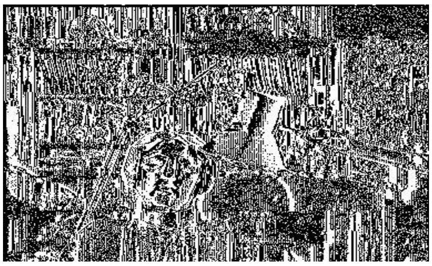
$$RMS\ Error = \left(\frac{1}{N}\sum_{(x,y)} |d_c(x,y) - d_t(x,y)|^2\right)^{\frac{1}{2}}$$

5. Observation & Results

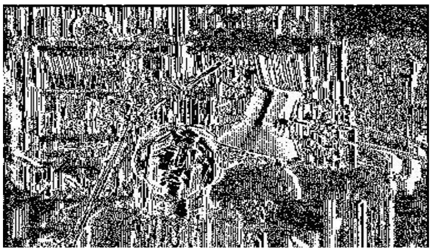
A) LEFT IMAGE

B)Right image

Result of SAD



Result of SSD



Ground truth image

OSERVATION TABLE

SRNO	ALGORITHM	RMS ERROR	MSE	PSNR
1	SAD	128.5731	82	24.04
2	SSD	128.0096	76	29.32

TABLE1.COMPARISION OF RMS.MSE AND PSNR

6.CONCLUSION

Matlab R2007b has been chosen for implementing different Stereo Matching Algorithms. Stereo Matching Algorithms like SAD(sum of absolute difference), SSD(sum of squared difference) have been implemented to generate disparity Map. SSD gives good disparity Map. MSE error is small compare to the SAD and give more accurate PSNR

7. FUTURE SCOPE

Disparity Maps are successfully generated by implementing Stereo Matching Algorithms, but still there is a scope for improvement. Performance of the Stereo Matching Algorithms is affected by the illumination conditions, shape and the camera characteristics. Effects of these three on Disparity Map. Depth Map can be Generate. Height and Width of an object can also be tried to be calculated. 3D view can also be generated by using Disparity Map and Depth Map.

8. REFERENCES

[1] R. Gonzales, P.Wintz, woods, Digital Image Processing, Addition-Wesley, 1987

[2]E.Trucco,V. Robeiot,S.Tinonin ,M.Corbatto SSD disparity estimation for dyanamic stereo

[3]Nadir Nourain Dawoud,Brahim Belhaouri samir,josefina janier fast template matching method based optimizied sum of absolte difference algorithm forface localiztion internanal journal of computer application vol.18 no8 ,march2011

[4]A. Ian H. Witten, Eibe Frank, The Geometry from Multiple Images, MIT Press, 2000.

[5] A Computational Theory of Human Stereo Vision by D. Marr and T.Poggio

[6] Cooperative Computation of Stereo Disparity by D. Marr and T. Poggio

[7] A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms by Daniel Scharstein and Richard Szeliski

[8] A Cooperative Algorithm for Stereo Matching and Occlusion Detection by C. Lawrence and Kanade

[9] www.middlebury.edu

