ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

Analysis of local and global techniques for disparity map generation in stereo vision

1. Vrushali Baxi, 2.Ms. Priyanka Tirkha 3.Bilav I. Prajapati
1, 2. Department of Computer Engineering SBCET, Jaipur, Rajasthan.
3. Department of Computer Engineering A D Patel institute of Technology, Gujarat.

Abstract

The field of computer vision is still open to generate 3D images (depth) using stereo image pairs. Obtaining reliable depth estimation has importance in robotic applications and autonomous systems. Stereo vision is the one of the methods that can yield depth information of the scene. The depth generation of images used for applications in multi view facial analysis, mobile multi-view coding, video surveillance etc, Disparity obtained from stereo images used to create realistic 3D scenes and environments for virtual reality and virtual studio applications. In this paper, methods for comparison of two images are given and to find disparity map from stereo image pairs are described. Different techniques on stereo for comparing stereo images are studied in this work. The disparity estimation method for local and global matching are briefly covered.

Keywords: stereo, multiview, disparity, depth.

.

1. Introduction

The goal of multi-view stereo is to reconstruct a complete 3D object model from a collection of images taken from known camera viewpoints. The situation in binocular stereo, the goal is to produce a dense depth map from a pair of images.

The method can be described as follows. Initially, for each point of the left image find the best match into the right image. Then, reverse the role of the two images and for each point of the right image find the best match into the left image. Finally, keep only those matches that turn out to be coherent when matching left-to-right (direct matching phase) and right-to-left (reverse matching phase).

It is worth observing that in both phases the match associated with each pixel is established independently of those found at neighboring pixels, since the other matching phase will highlight ambiguous matches.

Stereo Image: If we capture two images from a scene at appropriate angles, we can then view the scene later in three dimensions, this denotes a stereo pair.

The following figure shows how to view images,

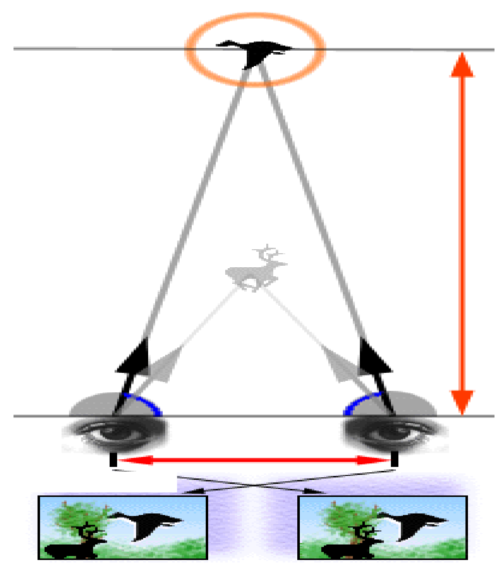

Fig.1: Figure showing two different images taken by eye

In this figure, by knowing the distance between reference points (red), and the angle of each eye (blue) when pointed at the distant subject (orange circle), we can determine the distance (orange) to the subject.

2. METHODS USE FOR MATCHING STEREO IMAGES

2.1 General Purpose methods for stereo matching:

The stereo matching is categorized in to two broad categories:

1. Intensity based: compare the images based on intensity pixels.

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

2. Feature based: compare the images, based on features of image like edges, corner, curves etc. The comparison in both the methods can be done in terms of *similarity*, *unique*, ordering. Our aim is to find Disparity Estimation so, let's take overview of some Disparity Estimation methods available.

2.2 Disparity Estimation methods:

According to recent stereo algorithms [1] that generate dense depth measurements can be roughly divided into two classes, namely *global* and *local* algorithms. Global algorithms rely on iterative schemes that carry out disparity assignments on the basis of the minimization of a global cost function. These algorithms yield accurate and dense disparity measurements but exhibit very high computational cost that renders them unsuited to real-time applications. *Local* algorithm also referred to as area-based algorithms; calculate the disparity at each pixel on the basis of the photometric properties of the neighboring pixels. There are different techniques available for local matching and Global matching algorithm. Some of the frequently used Local matching techniques are Squared intensity Differences (SD), Sum of Absolute intensity Differences (SAD) and Global matching techniques are Dynamic programming, Graph cuts and Belief propagation. In this paper the study of all these techniques is given.

2.2.1. Squared Intensity Differences:

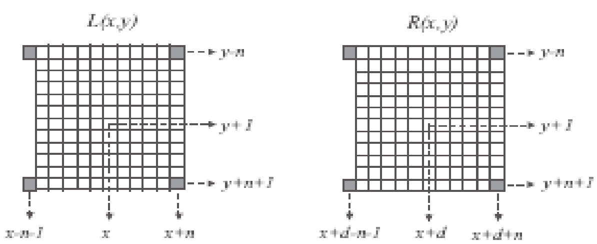

Fig. 2: Figure shows SD on a group of left and right image pixels. [5]

One of the commonly used method for local matching is Squared intensity Differences. As shown in figure we can compare intensity of each image by calculating difference of related pixel of right image to the left image and making sum of their squares we can map disparity *d*.

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 . \quad (1)$$

In the formula *yi* is left image pixel and *y* is right image pixel and n is number of pixels. This method is not preferable.

2.2.2. SAD [5] (Sum of Absolute intensity Differences):

This is an alternate method used widely for finding local matching. It is almost similar as squared intensity difference. But instead of calculating square differences it calculates absolute differences only so it becomes very simple. This is the reason why it used widely. It works by taking the absolute value of the difference between each pixel in the original (left image) block and the corresponding (right image) pixel in the block being used for comparison. These differences are summed to create a simple metric of block similarity, the of the difference image.
SAD is an extremely fast metric due to its simplicity; it is effectively the simplest possible metric that takes into account every pixel in a block. Therefore it is very effective for a wide motion search of many different blocks.

SAD(x, y, d) =SUM | L(x, y)-R(x,y)| (2)

In formula L(x, y) is left image pixel and R(x, y) is Right image pixel, d shows the difference.

2.2.3. Dynamic Programming:

This is the well known technique used for global matching. As shown fig 3 the left image pixel P and relative pixel of right image pixel Q with disparity d is shown.

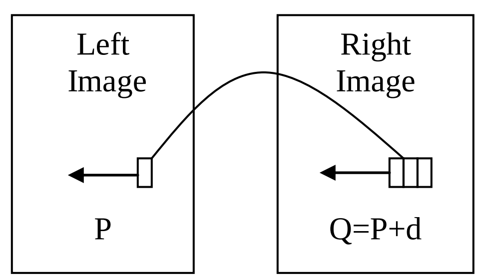

Fig 3: Figure shows disparity difference on left and right pixels

The Left scan line and Right scan line shows the Occluded Pixels by Left image and Right image, respectively. Find the minimum path with given penalty as shown as following steps.

1. Horizontal direction shows Left occluded pixels

2. Vertical direction shows Right occluded pixels

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

3. Diagonal direction shows visible pixels

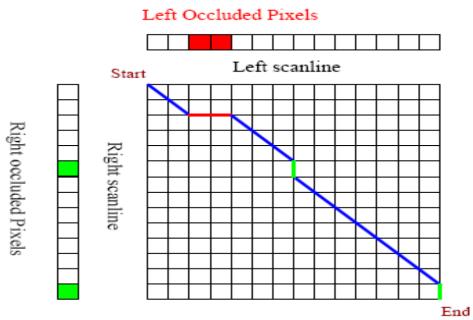

Fig 4: Minimum path using dynamic programming

The fig 4 shows left occluded pixels in red and right occluded pixels in green color and the minimum path.

2.2.4. Graph cut method:

All the problems that have been solved using graph-cuts have one property in common, they are all optimizing labeling means Trying to compute the most likely **labeling** from a very large solution space to explain some measured data. These labels are intensity values, disparity values in images, objects in images, or distinct classes in large training datasets. The individual entities which are labeled are represented as sites in the graphs and the contextual information which constrains the labeling are used to setup edges in the graph. In other words the probability of a particular site getting a specific label depends on the labeling of its neighbors. This contextual information is rich in images, video, textures, large training sets and is exploited in the graph cuts framework. The notion of a cut in such a graph is analogous to the idea of partitioning the sites using labels. The graph construction needs to be done in a way such that the minimum cut will yield the most likely labeling or partitioning of the measured data.

Greig *et al.* [4] were first to discover that powerful min-cut/max-flow algorithms from combinatorial optimization can be used to minimize certain important energy functions in vision. In this section we will review some basic information about graphs and flow networks in the context of energy minimization. A directed weighted graph *G= (E, V)*, consists of a set of nodes *V* and a set of directed edges *E* that connect them. Usually the nodes correspond to pixels, voxels or other features. A graph normally contains some additional special nodes that are called terminals, usually called the source *s* ∈V and the sink *t* ∈V. In the context of vision, terminals correspond to the set of labels that can be assigned to pixels. In Figure 5 we show a simple example of a two terminal graph (due to Greig et al. [4]).

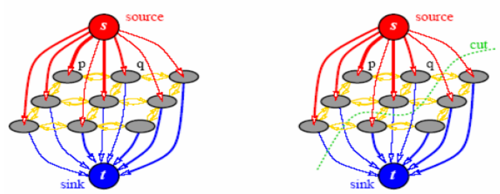

Fig 5: Graph cut method when applied on different pixels [1]

Normally, there are two types of edges in the graph*: n-links* and *t-links*. n-links connect pairs of neighboring pixels or voxels. Thus, they represent a neighborhood system in the image. The cost of n-links corresponds to a penalty for discontinuity between the pixels. T-links connect pixels with terminals (labels), and the cost of a t-link connecting a pixel and a terminal corresponds to a penalty for assigning the corresponding label to the pixel.

2.2.5. Belief Propagation

BP solves the stereo matching problem by pixel labeling[6]. The set S of sites is the set, or a subset of all M X N pixels (say, of the left image), and the set L of labels is defined by possible disparities, say L = {0, 1… dmax}

In general it is assumed that labels should only vary smoothly within an image, except at some edges (i.e., discontinuities) of regions. A standard form of energy functional, used for Characterizing a labeling function f, is as follows:

$$E(f) = \sum_{p\in P}\left(D_p(f_p) + \sum_{(p,q)\in A} V(f_p - f_q)\right)$$

We aim at minimizing this energy functional, which combines a data term *Dp(fp)* is the cost of assigning a label *fp* to pixel *p* with discontinuity term (in the inner sum):*V (fp - fq)* is the cost of assigning labels *fp* and *fq* to two adjacent pixels *p* and *q*, respectively; it represents the discontinuity cost.

Label *fp* is a disparity, and *Dp(fp)* may be defined by the difference in intensities between pixels in left and right image defined to be corresponding' when applying disparity fp. The task is a minimization of *E (f)* within the class of all possible labeling *f*.

3. Conclusion:

The study of different local and global matching algorithm concludes that SD and SAD are generally used for local matching techniques for stereo image depth map generation. In Squared intensity differences the calculation takes more time as well complex calculations. On the other hand the SAD provides very simple calculations, which makes it popular. That's the reason SAD used widely. The global matching algorithms like Dynamic programming, Graph cuts and Belief propagation are used frequently. Dynamic programming gets less computation time and obtains a robust dense disparity map. Belief propagation and Graph cut methods takes graph calculation and long procedures and comparisons of image pixels. Dynamic programming method used widely, because of less computation time. The observation shows only local algorithms gets confused near discontinuities of image. Belief propagation work better for 2d images. The combination of local and global can work better and produce more depth in stereo images. Generally the SAD is used for local matching and Dynamic programming method is used for Global matching.

4. References

[1] Sudipta N. Sinha ,"Graph Cut Algorithms in Vision, Graphics and Machine Learning" , *IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI), vol 26, no.9, pp 1124-1137, Sept 2004*.

[2] Y. Boykov and V. Kolmogorov, ".An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision.", *IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI), vol 26, no.9, pp 1124-1137, Sept 2004*.

[3] Barbara Zitova, Jan Flusser "Image registration methods: a survey", *ELSVIER November 2001; received in revised form 20 June 2003; accepted 26 June 2003*

[4] V. Kwatra, A. Schodl, I. Essa, G. Turk and A. Bobick,."Graphcut Textures: Image and Video Synthesis Using Graph Cuts." *SIGGRAPH 2003, pp. 277-286.*

[5] L. Di Stefano et al.,"A Fast Area-Based Stereo Matching Algorithm", *IEEE transactions on Image recognition*

[6] Andreas Klaus, Mario Sormann and Konrad Karner,"Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure", *VRVis Research Center 8010 Graz, Austria.*

[7] Jyothi Digge and Yashraj Digge, Stereo vision for Robotics. IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET 2012), pp. 33-39, 2012.