Algorithms for Stereo Matching

Jin Linhao

October 2015

1 Introduction

The recent development of Stereo Imaging focus on Stereo Matching as it is the most important part in finding disparity map of image pairs. A good number of algorithms have been proposed so far and they can be categorized into two groups, sparse output and dense output. Sparse output uses feature based methods stem from human vision studies and are based on image segments or edges between two images. It is mostly used in ancient times when computational powers were very weak.However, with the recent development of CPU and GPU, the demand for dense output is increasing dramatically.

Dense matching algorithms are classified into local and global algorithms. **Local methods** (area-based) trade accuracy for speed. They are window-based methods because disparity computation at a given point depends only on intensity values within a finite support window. **Global methods** (energy-based) on the other hand are time-consuming but very accurate.Their goal is to minimize a global cost function, which combines data and smoothness terms, taking into account the whole image. Of course there are many other methods such as **Dynamic Programming** which is a fair trade-off between the complexity of computation need and the quality of the results obtained. In every aspect, DP stands between local algorithms and global optimization ones.

This report will first discuss the basic algorithms of dense matching before 2008, followed by modern approaches including convolutional neural network, scene flow and semi-global methods.

2 Local Methods

2.1 Sum of absolute differences (SAD)

Nearly every algorithm uses *matching cost function* to find the correspondence of two pixels of an image pair. The matching cost are commonly aggregated in SAD over a rectangular support region with fixed or adapted size. SAD function is mathematically defined as

$$SAD(x, y, z) = \sum_{x,y \in W} |I_l(x, y) - I_r(x, y - d)| \tag{1}$$

where I_l, I_r are the intensity values in left and right image, (x,y) are pixel coordinates, d is the disparity value under consideration and W is the aggregated support region. This function takes the absolute difference between each pixel in the original block and the corresponding pixel in the block being used for comparison and takes the summation. It is the simplest and most widely used algorithm for in local methods of Stereo Matching. Most development of methods are based on this.

2.2 Sum of Square Differences (SSD)

The method of SSD is similar to SAD with the only difference in mathematical representation

$$SSD = \sum_{x,y \in W} (I_l(x,y) - I_r(x,y-d))^2 \tag{2}$$

2.3 Normalized Cross Correlation (NCC)

cross-correlation is a measure of similarity of two series as a function of the lag of one relative to the other. For image-processing applications in which the brightness of the image and template can vary due to lighting and exposure conditions, the images can be first normalized. This is typically done at every step by subtracting the mean and dividing by the standard deviation.

$$\widehat{f} = \frac{f - \bar{f}}{\sqrt{\sum (f - \bar{f})^2}} \tag{3}$$

$$\widehat{g} = \frac{g - \bar{g}}{\sqrt{\sum (g - \bar{g})^2}} \tag{4}$$

so

$$NCC(f.g) = C_f g(\widehat{f}, \widehat{g}) = \sum_{i.j \in R} \widehat{f}(i,j)\widehat{g}(i,j) \tag{5}$$

For the case of stereo imaging

$$NCC(x,y,z) = \frac{\sum_{x,y \in W} I_l(x,y) \cdot I_r(x,y-d)}{\sqrt{\sum_{x,y \in W} I_l^2(x,y) \cdot I_r^2(x,y-d)}} \tag{6}$$

The score of NCC varies from -1 to 1 where 1 is the best match, -1 is the worst match.

3 Global Methods

The goal of global methods is to minimize a global cost function E, which combines data and smoothness terms.

$$E(d) = E_{data}(d) + \lambda \cdot E_{smooth}(d) \tag{7}$$

where E_{data} takes consideration the (x,y) pixel's value through out the image, E_{smooth} provides the algorithm's smoothing assumptions and λ is a weight factor.

3.1 Color Segmentation

Color image segmentation simplifies the vision problem by assuming that objects are colored distinctively, and that only gross color differences matter. It therefore discards information about color and brightness variations that provides many valuable cues about the shapes and textures of 3D surfaces. But the resulting simplified (and impoverished) image can be processed very rapidly, which can be important in mobile robot applications.

Several techniques involve the use of thresholds in a three dimensional color space. Several color spaces are in wide use, including Hue Saturation Intensity (HSI), YUV and Red Green Blue (RGB). The choice of color space for classification depends on several factors including specific situation and digital hardware.

3.2 Graph cut and graph theory

A Cut of a graph is a partition of the vertices in the graph into two disjoint subsets based on graph theory. Many image segmentation problems uses techniques for graph cut in graph theory. In graph theory, a graph G is a pair of sets (V, E) where V is a nonempty set of items called vertice or nodes, E is a set of two item subject of V called edges. An undirected graph is defined as a set of nodes (vertices V) and a set of undirected edges E that connect the nodes. A directed graph is defined as a set of nodes (vertices V) and a set of ordered set of vertices or directed edges E that connect the nodes. For an edge e=(u,v), u is called the tail of e, v is called the head of e. A cut is a set of edges $\mathbf{C} \subset \mathbf{E}$ such that the two terminals become separated on the induced graph.

$$G' = (V, E \setminus \mathbf{C}) \tag{8}$$

A flow network is defined as a directed graph where an edge has a nonnegative capacity and A flow in G is a real-valued (often integer) function that satisfies the following three properties:

- Capacity Constraint: for all u,v $\in V$, $f(u,v) \leq$ c(u,v)
- Skew Symmetry: for all u,v $\in V$, $f(u,v) = -f(u,v)$
- Flow Conservation: for u $\in (V \setminus \{s,t\})$, $\sum f(u,v) = 0$

Where s denotes a source terminal and t denotes a sink terminal. See Figure 1.To find the minimum cut in graph G, the maximum source-to-sink flow

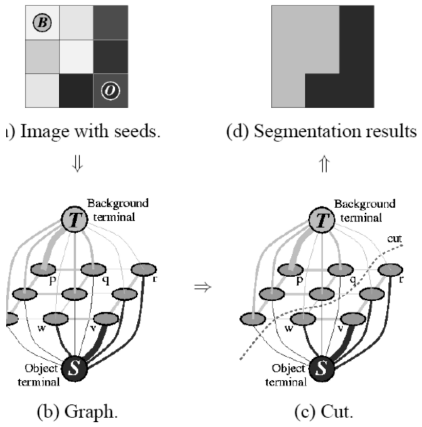

Figure 1: Source terminal and edges

possible is equal to the capacity of the minimum cut in G. There are several algorithms like

- Ford-Fulkerson Algorithm
- Push-Relabel Algorithm
- New Algorithm by Boykov, etc.

For example, the main operation in Ford-Fulkerson Algorithm is Starting from zero flow, increase the flow gradually by finding a path from s to t along which more flow can be sent, until a max flow is achieved. The Ford-Fulkerson algorithm uses a residual network, which defined as the network of edges containing flow that has already been sent of flow, to find the solution.

Solving image segmentation problem, data element set P represents the image pixels, where Neighborhood system as a set N representing all pairs $\{p, q\}$ of neighboring elements in P (ordered or unordered). A be a vector specifying the assignment of pixel p in P, each A_p can be either in the background or the object.

$$A = (A_1, A_2, ..., A_p, ..., A_{|p|}) \tag{9}$$

Implementing the global cost function Function. 7 yields

$$E(A) = \lambda \cdot R(A) + B(A) \tag{10}$$

Where

$$R(A) = \sum_{p \subset \mathbf{P}} R_p(A_p) \tag{11}$$

$$B(A) = \sum_{p \subset \mathbf{P}} \sum_{\{p,q\} \subset N} B_{\{p,q\}} \cdot \delta(A_p, A_q) \tag{12}$$

$$\delta(A_p, A_q) = \begin{cases} 1, & A_p \neq A_q \\ 0, & A_p = A_q \end{cases} \tag{13}$$

The global minimum of result E(A) gives the best segmentation hence the best match.

3.3 Markov Random Field

A Markov Random Field (MRF) is a graph $G = (V, E)$ which satisfies its random variable (RV) u_j to satisfy

$$p(u_i \mid \{u_j\}_{j \in V \setminus i}) = p(u_i \mid \{u_j\}_{j \in N_i}) \tag{14}$$

N_i is called the Markov blanket of node i. The key to MRFs is that, through local connections, information can propagate a long way through the graph. This communication is important if we want to express models in which knowing the value of one node tells us something important about the values of other, possibly distant, nodes in the graph.

The main idea is the distribution over an MRF satisfies Equation 14 can an be expressed as the product of (positive) potential functions defined on maximal cliques of G Such distributions are often expressed in terms of an energy function E, and clique potentials Ψ:

$$p(u) = \frac{1}{Z}(-E(u, \theta)) \tag{15}$$

Where

$$-E(u, \theta) = \sum_{c \in C} \Psi_c(\bar{u}_c, \theta_c) \tag{16}$$

Here,

- C is the set of maximal cliques of the graph (i.e., maximal sub-graphs of G that are fully connected)
- The clique potential Ψ_c, c$\in C$, is a non-negative function defined on the RVs in clique $\bar{u}_c$, parameterized by θ_c
- Z, the partition function, ensures the distribution sums to 1:

$$Z = \sum_{u_1 \ldots u_N} \prod \exp(-\Psi_c(\bar{u}_c, \theta_c)) \tag{17}$$

The partition function is important for learning as it's a function of the parameters $\theta = \{\theta_c\}_{c \in C}$. But often it's not critical for inference.

MRF is often used along with graph theory model which regards each pixel as a node connected with or without noise. Apply global cost function Equation

7 into the model will generate a penalize for smoothness for the disperancy of data v and solution u. For Stereo Imaging, a typical graph cut method, α - expansion move is used with a iterative circle through all labels. α expansion is defined as: For a given label $\alpha \in$ L, let any RV whose current label is in L$\backslash$ α either switch to α, or remain the same. Given a labeling u, and the label α, construct a new graph such that the min-cut labeling $\widehat{u}$ minimizes E($\widehat{u}$). So the move start with an arbitary label, find the lowest E within a single α-expansion, go there if it has a lower E than current labeling, if E does not decrease, then stop.

4 Dynamic Programming

Dynamic Programming (DP) is a method for solving complex problems by breaking them down into simpler sub-problems which contains the steps of defining sub-problems, write down the recurrence that relates the problems and recognize and solve the basic case. DP can be classified into 1-dimensional DP, 2-dimensional DP, interval DP, tree DP and subset DP. In stereo imaging, DP is the method stands between local and global algorithms. DP approaches the matching problem of finding the correct disparities on a scan line is regarded as a search problem. The matching costs of all points on a scan-line describe the disparity search space. Finding the correct disparities is akin to finding the path in this space which takes the shortest route through the cost values. Special rules for how to transverse the search space can be added in order to handle occlusions.