A Bottom-up and Top-down Based Disparity Computation

Jung-Gu Kim and Hong Jeong

Abstract

It is becoming apparent that stereo matching algorithms need much information from high level cognitive processes. Otherwise, conventional algorithms based on bottom-up control alone are susceptible to local minima. We introduce a system that consists of two levels. A lower level, using a usual matching method, is based upon the local neighborhood and a second level, that can integrate the partial information, is aimed at contextual matching. Conceptually, the introduction of bottom-up and top-down feedback loop to the usual matching algorithm improves the overall performance. For this purpose, we model the image attributes using a Markov random field (MRF) and thereupon derive a maximum a posteriori (MAP) estimate. The energy equation, corresponding to the estimate, efficiently represents the natural constraints such as occlusion and the partial informations from the other levels. In addition to recognition, we derive a training method that can determine the system parameters automatically. As an experiment, we test the algorithms using random dot stereograms (RDS) as well as natural scenes. It is proven that the overall recognition error is drastically reduced by the introduction of contextual matching.

I. Introduction

Stereo matching is the fundamental problem in the field of binocular stereo vision. The final goal of stereo matching is the reconstruction of the three dimensional world from left and right images which are taken from spatially separated cameras. To solve this problem, it is necessary to find corresponding points between two images. The matching result is represented as a disparity map and the three dimensional world is rebuilt from the reverse optical structure with the disparity map. From Julesz's experiment[9] using random dot stereo-grams it is well known that human can reconstruct the 3D world without higher recognition and recognize 3D objects even if one of the images has different properties such as color, noise, blurring and scale. This means that our visual system is invariant under of color, noise, blurring, and scale.

There are many different approaches to solving the matching problem, but they are basically based on point and pattern. Point based method use intensity[8,15], correlation[7], and intensity partial derivatives[10], whereas pattern based method use edges[2], line segments[1], special points[4], zero crossing[12] and peak values[13] in the image for matching primitives. Recently the phase[5, 14, 16] of the image has been used as the feature. However, good results using these schemes are possible only for specific images, which don't have the invariance of color, noise, blurring or scale.

In general, if only local information is used, there are inherent limitations in computing disparity. A typical problem is that within the occlusion regions, the corresponding points between two images do not exist and only by considering global information can the disparity map be computed. Furthermore, there is no way to accommodate high level partial information supplied externally. In this paper, we propose an additional level and introduce top down feedback to the conventional bottom up method.

Conceptually, in this tightly-coupled feedback loop, the results from the lower level, which will be called *matching*, are improved at the high level, which will be called *labeling* and also the results from the high level help the low level decisions. The net result is drastic reduction of recognition errors. Moreover, the system has the advantages of modular structure and capability of integrating partial information from other parts of the system. To construct such a system, we first hypothesize image attributes by an MRF and thereby define the stereo matching by MAP estimation. For the MAP estimate, we model the energy so that it accommodates natural constraints such as occlusion and linking capability

Manuscript received August 25, 1997; accepted January 12, 1998.

Jung-Gu Kim is with the Power Electronics Research Team, Research Institute of Industrial Science and Technology, Pohang 790-600, Korea.

Hong Jeong is with the Department of Electrical Engineering, Pohang University of Science and Technology, Pohang 790-600, Korea.

with lower and higher level modules. Then, this energy function undergoes great approximation using the mean field approach. The next step is to find minimizers of the energy equation and we derive a pair of relaxation equations. Arriving at the recognition method, we consider a training method that can determine the system parameters automatically and dynamically, adapting to the characteristics of incoming images. The organization of the rest of this paper is as follows. In § 2, we define stereo matching in terms of the MAP estimate. Then, we deal with a new matching algorithm, that can be linked with labeling, in § 3. For the high level stage, a labeling algorithm is introduced in § 4. Finally, § 6 addresses the issues of experimental results and performance analysis.

II. Defining Stereo Matching Problem

This section will define the overall problem in terms of Bayesian estimation and derive MAP estimates as suboptimal solutions.

1. Finding Optimal Disparity

Let's consider that (g^L, g^R) denote gray level images, of $N \times N$ pixels, taken from the left and right cameras. A *feature extraction* step F transforms the images into some *feature maps* (f^L, f^R), all $N \times N$ arrays. That is, $F : g \rightarrow f$ and in this research f is phase. In addition to the incoming images, assume that there are other information sources that will supply partial information **u** on stereo matching.

Given the information (f^L, f^R) and **u**, define the stereo matching $S : (f^L, f^R, \boldsymbol{u}) \rightarrow s$, where **s** is called *disparity volume*. In particular, $s = \{ s_{ijk} | i \in [1, N], j \in [1, N]$ and $k \in [1, M] \}$ for some nonnegative integers N and $M = 2d_{\max} + 1$, where $d_{\max}$ denotes the maximum disparity value that is allowed for a pixel. We can convert the *disparity map* d to **s** by T : d → **s** such that

$$T : s_{ijk} = \delta_{k-(\frac{M+1}{2} - d_{ij})}, \quad k \in [1, M] \tag{1}$$

where δ is Kronecker delta. Then, the stereo matching problem becomes

$$s = \arg\max P(s | f^L, f^R, \boldsymbol{u}). \tag{2}$$

This is a MAP estimate that determines the parameter **s** from the observation (f^L, f^R) and the partial information **u**. Instead of solving (2), we modify the problem into a simpler form. Consider a disparity d as an intermediate representation between f and s, as illustrated in Fig. 1. In this dependency diagram, **s** is used as an initial value of d. The transformation from **s** to **d** is given by s → d :

$$T^{-1} : d_{ij} = \frac{M+1}{2} - \sum_{k=1}^{M} k\delta_{1-s_{ijk}}. \tag{3}$$

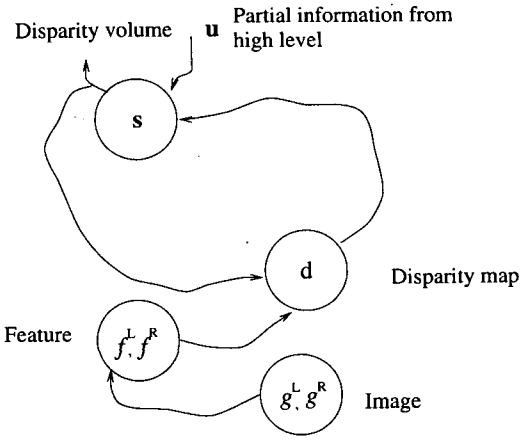


Fig. 1. The relationships between variables.

Note that (1) and (3) are inverses. Then, the problem becomes.

$$s = \arg\max \sum_d P(s | d, \boldsymbol{u}) P(d | f^L, f^R, d^0) \tag{4}$$

where d^0 is a initial disparity map and **u** is omitted in the second term assuming that only **s** is dependent on **u**. Instead of computing all the summations, we choose a suboptimal solution

$$(s, d) = \arg\max P(s | d, \boldsymbol{u}) P(d | f^L, f^R, d^0). \tag{5}$$

It is very difficult task to find **s** and d simultaneously. An alternative and practical approach is to first fix **s** and compute d by the first equation in (5), and *vice versa* :

$$\begin{cases} d = \arg\max_d P(d | f^L, f^R, s), \\ s = \arg\max_s P(s | d, \boldsymbol{u}). \end{cases} \tag{6}$$

The solution must satisfy the two equations in (6). These equations, denoted *matching* and *labeling* respectively, are dealt with in § 3 and § 4.

2. An Overall Block Diagram

The block diagram consists of three transformations: feature extraction, matching, and labeling and three representations: feature map, disparity map, and disparity volume. At the bottom of the figure, the input signals flow into the feature extraction stage are converted into a feature map. In the next step, the matching system receives the feature map and generates a disparity map. Located at the top, the labeling system uses some global context to find a better interpretation of disparity. It looks at a much wider scope of data than the matching stage. By the feedback link, the output of the labeling is delivered to the matching as an initial value. In this figure, the combination of feature

extraction and matching corresponds to the conventional bottom-up computation, whereas the combination of the labeling and matching corresponds to the top-down computation. Starting from some initial states, hopefully, the state will converge to an equilibrium point.

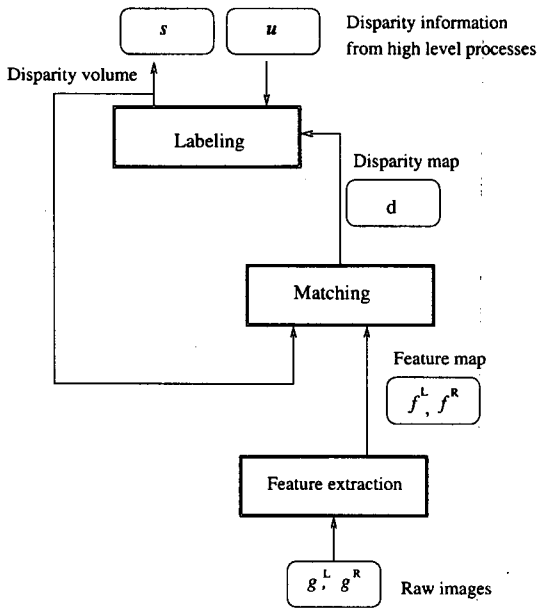


Fig. 2. The overall computation structure.

III. Local Stereo Matching

Solving (6) requires an iterative approach. This section deals with the first equation in (6). We follow the multilevel approach in Kim and Jeong[11] with slight modifications. One of the advantages of this algorithm is that we can easily integrate the information from the high level processes, *i. e.*, **s**.

1. Feature Extraction and Matching

Consider the solution of

$$\begin{cases} d = \arg\max_d P(d|f^L, f^R, d^0), \\ d = T^{-1}(s). \end{cases} \tag{7}$$

This is a typical stereo matching problem, except for the fact that it has an additional constraint d^0. This means that in addition to the observations, the solution must rely on the information that the high level process has provided in advance. Our scheme is to interpret this information as an initial value of the disparity. Consider a pyramid built from an $N \times N$ image plane W. The planes W_i, ($i \in [0, \log_2 N]$) are derived successively from the lower plane by reducing its size by half(however, we set $W_0 = W$.). Then W_i is an $(N/2^i)^2$ grid. For each layer i, the disparity vector $\mathbf{d}_i$ is defined by $\mathbf{d}_i = \{d_w, w, = (x,y) \in W_i\}$ and $\mathbf{d}_i \in R_i$, where R_i is the range space of the disparity vector. Our purpose is to find an optimal disparity d_1^* in R_i, for all i from top to bottom.

A set of pixels in W must be mapped into each site in W_i. We call the set of pixels *block* and with denote it by $B_i(i = 0,1,\ldots \log_2 N)$. In this manner, a site in W can be associated pixels in the original image plane W by the block B_i. Each cell in W_i has a block B_i of $(2^i)^2$ cells in W.

By to this method, one can determine $\mathbf{d}_i$ using all the phase values as a matching primitive in the original plane W. Let us consider a general scheme for finding an optimal solution d. Assuming that all the values inside a block B_i corresponds to the center frequency u_i of the associated bandpass filter, we find the solution in R_i. This value is now used as an initial condition in R_{i-1} and similar operation continues up until $i = 0$, with the final solution residing in R_0.

Consider the algorithm in the computational layers. An element (x, y) in W_i is uniquely identified by the subscript (x, y, i). Remember that for each (x, y, i), the block B_i associates a set of elements in W. Then the algorithm can be conveniently represented by

$$\begin{cases} d^{m+1}_{x,y,i} = d^n_{x,y,i} + \alpha\Big(\sum_{\omega\in} B_i(1-p^l_s \oplus p^r_s)(f^L_s - f^R_{\omega+d^*_{x,y,i}}) \dfrac{\partial f^R_{s+d^n x,y,i}}{\partial_s} \\ -\lambda \sum_{(k,l,i)\in N_{x,y,i}} (d^n_{x,y,i} - d^n_{k,l,i}) \Big(1 - \dfrac{1}{1+\exp[\gamma - \mu(d^n_{x,y,i} - d^n_{k,l,i})^2]}\Big)\Big) \end{cases} \tag{8}$$

where p^l_s, p^r_s and $\oplus$ represent respectively the left and right singular points, and the logical OR[11] operator. The initial conditions are $d^0_{x,y,i} = d_{x,y,i+1}$ and this computation must be performed $\forall(x,y)\in W_i$, ($i = \log_2 N, N\text{-}1,\cdots, 0$). Notice that this computation must proceed for each block in each layer from top to bottom.

2. Block Diagram for Local Matching

As Fig. 3 depicts, the local matching system consists of two parts: feature and disparity computation. The first block transforms the input image to obtain a phase-magnitude pair and thereby computes a singularity map. These intermediate

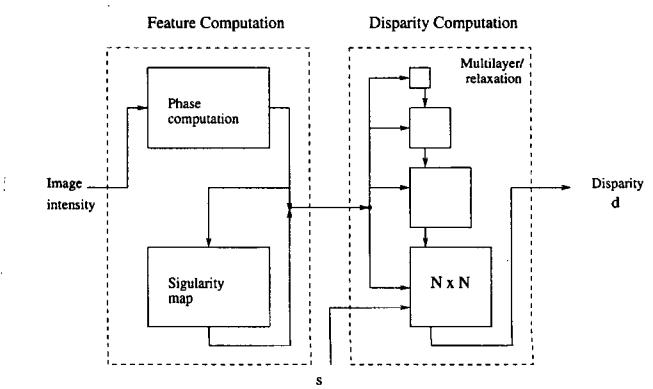


Fig. 3. Overall computation structure.

representations are used in the next block as inputs to the matching terms in (8). Inside this block the computation flows downwards and the conclusion the output exists as the final state in the bottom layer. The major difference with Kim and Jeong[11] is as follows. When this system is connected to the labeling system and is used second time, only the bottom part of this multilayer is used and it receives the disparity as an initial value from the labeling stage, that is, $d^0 = T^{-1}(s)$.

IV. Contextual Matching

The next step is to solve the second equation in (6). The basic idea is that the disparity in the signal level has some errors even when the selected features are good and matching algorithm is fine, since there will be many unexpected situations in a stereo image pair. Some recognition errors can not be corrected without using higher level knowledge. This and next section will deal with such topics.

1. Global Matching

The problem is to solve

$$\begin{cases} s = \arg\max_s P(s|s^0, u), \\ s^0 = T(d). \end{cases} \tag{9}$$

To begin with, Fig. 4 illustrates the mechanism of labeling. It contains two image strips aligned in the epipolar lines extracted from both images. The gray boxes indicate the feature strength of the image and the black boxes indicate the dominant feature within the neighborhood. Starting from the initial matching positions, which are indicated by arrows pointing upwards, one must find the correct matching positions in the search region.

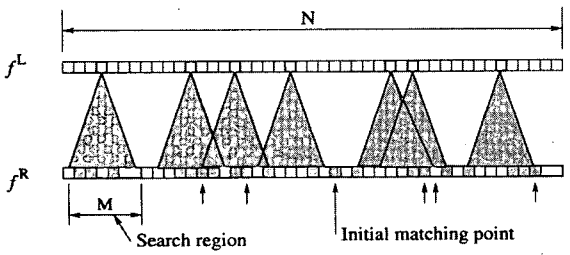


Fig. 4. Matching two image strips.

A result is shown in Fig. 5. The x and y axes represent each strip of the left and right images respectively. The $x = y$ diagonal indicates no shift between the two images and the black boxes the correct matching positions. Notice that d_{max} denotes the maximum disparity.

The relationship between d_{ij}, that is located at the (i, j)th position of the left signal, and s_{ij} is simple. For example, $d_{ij} = k$ means that s_{ij} = [0, ··· 1, 0, ···], where 1 is the $(\frac{M+1}{2} - k)$th element. The relationships are given by (1) and (3).

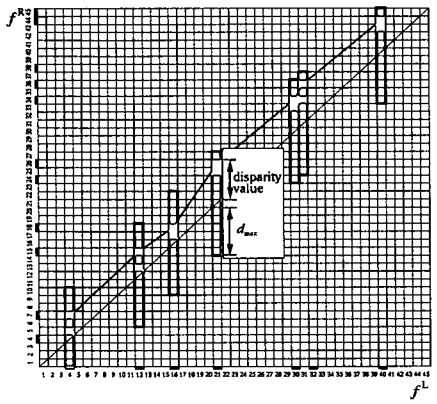


Fig. 5. Choosing the optimal matching points.

Now let us consider an important natural constraint. Matching of the two strips must obey the *precedence (or ordered constraint)* rule, that is

$$d_{ij} \geq d_{ij-x} - x, \tag{10}$$

where ij is the absolute position of the left image features and d is disparity value. In fact, (10) means that the slope in Fig. 5 is nonnegative.

The ultimate goal in labeling is to find the disparity volume **s** from the given disparity map d, that obeys the precedence rule as well as the partial knowledge **u**.

2. Energy Function

Applying Bayes rule to (9) yields

$$s = \arg \max_s P(s|s^0, u) \tag{11}$$

$$= \arg \max_s P(s^0|s)P(u|s)P(s) \tag{12}$$

where s^0 denotes the initial disparity volume computed from the disparity map d, *i.e.*, $s^0 = T(d)$.
We consider the first term in (12) as

$$P(s^0|s) \triangleq \frac{1}{Z_s^0} \exp\{-U(s^0|s)\}, \tag{13}$$

where Z_{s^0} is a partition function given by $Z_{s^0} = \Sigma_{s^0} \exp\{-U(s^0|s)\}$, and $U(\cdot)$ is an energy function

$$U(s^0|s) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha \|s_{ij} - s_{ij}^0\|^2, \tag{14}$$

where α is a constant. Analogous to **s**, for **u** we have

$$U(u|s) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{N} \beta \|s_{ji} - u_{ij}\|^2, \tag{15}$$

where β is a parameter.

As s_{ij} can be modeled as a first order Markov process, the *prior* term s in (12) can be represented by Gibbs distribution [3, 6], given by

$$P(s) \triangleq \frac{1}{Z_s}\exp\{-U(s)\}, \tag{16}$$

where $Z_s=\Sigma_s\exp\{-U(s)\}$, $U(s)=\Sigma_{c\in C}V_c(s)$. The potential function $V_c(s)$ is a real value defined on the clique c. Considering three types of cliques in the first order Markov process, we can model the energy function as

$$U(s) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{N}\{s_{ij}^T b + s_{ij-1}^T A s_{ij} + s_{ij}^T A s_{ij+1}\}, \tag{17}$$

where $b\in\mathbb{R}^M$, $A\in\mathbb{R}^{M\times M}$.

As a result, one arrives at the energy equation:

$$\begin{cases} U(s|s^0,u) = \sum_{i=1}^{N}\sum_{j=1}^{N}\{\alpha\|s_{ij}-s_{ij}^0\|^2+\beta\|s_{ij}-u_{ij}\|^2+s_{ij}^T b+s_{ij-1}^T A s_{ij}+s_{ij}^T A s_{ij+1}\}, \\ \sum_{k=1}^{M} s_{ijk} = 1, \quad \forall i \text{ and } j \\ s_{ijk} \geq 0, \quad \forall i,j \text{ and } k. \end{cases} \tag{18}$$

Note that the first term adjusts s to fit the data supplied by the matching stage, whereas the second term takes care of the partial information from the high level processes. Prior knowledge about s is represented by the third term. In this fashion, different types of information are merged naturally in this representation.

V. Obtaining Optimal Solutions

So far we have derived an energy function (18), that describes an optimal contextual matching. This section addresses how to solve the equation and also how to determine various parameters.

1. Minimizing the Energy Function

To solve the constrained optimization problem (18), we convert it into the Lagrangian $L(s,\lambda)$.

$$\begin{aligned} L(s,\lambda) &= U(s|s^0,u)+\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_{ij}\Big(\sum_{k=0}^{M}s_{ijk}-1\Big), \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\{\alpha\|s_{ij}-s_{ij}^0\|^2+\beta\|s_{ij}-u_{ij}\|^2+s_{ij}^T b+s_{ij-1}^T A s_{ij}+s_{ij}^T A s_{ij+1} \\ &\quad +\lambda_{ij}\Big(\sum_{k=1}^{M}s_{ijk}-1\Big)\}, \end{aligned} \tag{19}$$

where $\lambda=\{\lambda_{ij}|i\in[1,2,\cdots,N], j\in[1,2,\cdots,N]\}$ is the Lagrange multiplier. As a result, the problem is to find (s, λ) such that

$$(s,\lambda) = \arg \max_{\lambda} \min_{s\geq 0} L(s,\lambda). \tag{20}$$

Since (19) is a quadratic convex function, there exists a unique optimal solution. The necessary conditions are

$$\begin{cases} \nabla_{\lambda_{ij}} L(s,\lambda) = \sum_{k=1}^{M} s_{ijk} - 1 = 0, \\ \nabla_{s_{ij}} L(s,\lambda) = 2\alpha(s_{ij}-s_{ij}^0)+2\beta(s_{ij}-u_{ij})+b+A^T s_{ij+1}+\lambda_{ij}=0. \end{cases} \tag{21}$$

Note that the elements of s are non-negative. Rearranging the equations, one has

$$\begin{cases} s_{ijk} = \max\left(0, \dfrac{2\alpha s_{ijk}^0+2\beta u_{ijk}-b_k-(A^T s_{ij-1})_k-(A s_{ij+1})_k-\lambda_{ij}}{2(\alpha+\beta)}\right) \\ \lambda_{ij} = \dfrac{1}{M}\{\sum_{k=1}^{M}[2\alpha s_{ijk}^0+2\beta u_{ijk}-b_k-(A^T s_{ij-1})_k-(A s_{ij+1})_k]-2(\alpha+\beta)\}. \end{cases} \tag{22}$$

The equations can be easily converted to recursive forms:

$$\begin{cases} s_{ijk}^{\gamma+1} = \max\left(0, \dfrac{2\alpha s_{ijk}^0+2\beta u_{ijk}-b_k-(A^T s_{ij-1})_k-(A s_{ij+1})_k-\lambda_{ij}^T}{2(\alpha+\beta)}\right) \\ \lambda_{ij}^{\gamma+1} = \dfrac{1}{M}\{\sum_{k=1}^{M}[2\alpha s_{ijk}^0+2\beta u_{ijk}-b_k-(A^T s_{ij-1}^T)_k-(A s_{ij+1}^T)_k]-2(\alpha+\beta)\} \end{cases} \tag{23}$$

where γ is the iteration number and $(\cdot)_k$ denotes the kth element of the vector. To simplify the problem, we use s_{ijk}^0 as an initial value.

2. Determining Parameters

The equations in (23) contains four parameters α, β, **b**, and A. Recall that α and β are related to the matching terms in (14) and (15) and also that **b** and A denote the constants for the prior term in (17). The form of A can be derived from the precedence rule (10) :

$$\begin{cases} a_{i,j} = a_{i+j+1}, \quad \forall i,j\in[1,M], \\ a_{i,j+1} > a_{i,j}, \quad \forall j > i+1, \\ a_{i,j-1} > a_{i,j}, \quad \forall j \leq i+1. \end{cases} \tag{24}$$

Note that A is Toeplitz.

To obtain the parameters, we consider the maximum likelihood (ML) estimate:

$$\Theta = \arg \max_{\Theta} P(s^0,u,s|\Theta), \tag{25}$$

where $\Theta \triangleq [\alpha|\beta|b^T|A'^T]^T$. The procedure for building vector A' from matrix A is defined as follows. If $A=\{a_{ij}\ i,j\in[1,M]\}\in R^{M\times M}$, then $A'=\{a'_i|i,j\in[1,M]\}\in R^{M^2\times 1}$ is the vector with $a'_{k+M(l-1)} \triangleq a_{kl}$. Unfortunately, although the ML estimate is unique if it exists [17], it is computationally prohibitive due to the calculation of the partition function. Therefore, as an alternative to ML, we consider maximum pseudo likelihood (MPL) estimation where $P(s^0, u, s|\Theta)$ is represented as a product of local partition functions[17].

$$\begin{aligned} P(s^0,u,s|\Theta) &= \prod_{i=1}^{N}\prod_{j=1}^{N}\frac{1}{Z_{ij}}\exp\{-[\alpha\|s_{ij}-s_{ij}^0\|^2+\beta\|s_{ij}-u_{ij}\|^2+s_{ij}^T b \\ &\quad +s_{ij-1}^T A s_{ij}+S_{ij}^T A s_{ij+1}]\} = \prod_{i=1}^{N}\prod_{j=1}^{N}\frac{1}{Z_{ij}}\exp\{-\Theta^T\Phi(s_{ij})\}, \end{aligned} \tag{26}$$

where Z_{ij} is a local partition function: $Z_{ij}=\Sigma_{\forall s_{ij}}\exp\{-\Theta^T\Phi(s_{ij})\}$ and $\Phi(s_{ij})$ are the cliques in our system:

$$\Phi(s_{ij}) = \begin{bmatrix} \|s_{ij}-s_{ij}^0\|^2 \\ \|s_{ij}-u_{ij}\|^2 \\ s_{ij} \\ (s_{ij-1}s_{ij}^T+s_{ij}s_{ij+1}^T)' \end{bmatrix}. \tag{27}$$

This is the conditional probability like the Ising model that consists of potentials for each clique.

It can be shown that (26) is strictly concave with respect to Θ if and only if the parameters that comprise Θ are linearly independent[17]. Therefore, Θ can be found by the gradient search method:

$$\frac{\partial \Theta}{\partial t} = -\mu \nabla_{\Theta} \log P(s^0, u, s|\Theta). \tag{28}$$

Putting (26) into (28) gives

$$\begin{aligned} \Theta^{\tau+1} &= \Theta^{\tau} - \mu \sum_{i=1}^{N}\sum_{j=1}^{N} \{\Phi(s_{ij}) - \frac{1}{Z_{ij}} \sum_{\forall s_{ij}} [\Phi(s_{ij}) \exp(-\Theta^{\tau T} \Phi(s_{ij}))]\}, \\ &= \Theta^{\tau} - \mu \sum_{i=1}^{N}\sum_{j=1}^{N} \{\Phi(s_{ij}) - E[\Phi(s_{ij})]\}, \end{aligned} \tag{29}$$

where μ and τ are respectively an updating constant and an iteration index.

3. Computational Complexity

Here we summarize the overall scheme and discuss the computational complexities. From § 3 and § 4, we gather (8), (29) and (23) and put them together here:

$$\begin{cases} d_{x,y,i}^{n+1} = d_{x,y,i}^{n} + \alpha \Big\{ \sum_{w \in B_i} (1 - \mathbf{p}_s^l \oplus \mathbf{p}_s^r)(f_s^L - f_{w+d_{x,y,i}^n}^R) \frac{\partial f_{s+d_{x,y,i}^n}^R}{\partial s} \\ \qquad - \lambda \sum_{(k,l,i) \in N_{x,y,i}} (d_{x,y,i}^n - d_{k,l,i}^n)\Big(1 - \frac{1}{1+\exp[\gamma - \mu(d_{x,y,i}^n - d_{k,l,i}^n)^2]}\Big)\Big\}, \\ \Theta^{\tau+1} = \Theta^{\tau} - \mu \sum_{i=1}^{N}\sum_{j=1}^{N} \{\Phi(\mathbf{s}_{ij}) - E[\Phi(\mathbf{s}_{ij})]\}, \\ s_{ijk}^{\tau+1} = max\Big(0, \frac{2\alpha s_{ijk}^0 + 2\beta u_{ijk} - b_k - (A^T \mathbf{s}_{ij-1}^{\tau})_k - (A\mathbf{s}_{ij+1}^{\tau})_k - \lambda_{ij}^{\tau}}{2(\alpha+\beta)}\Big), \\ \lambda_{ij}^{\tau+1} = \frac{1}{M}\{\sum_{k=1}^{M} [2\alpha s_{ijk}^0 + 2\beta u_{ijk} - b_k - (A^T \mathbf{s}_{ij-1}^{\tau})_k - (A\mathbf{s}_{ij+1}^{\tau})_k] - 2(\alpha+\beta)\}. \end{cases} \tag{30}$$

The first equation denotes a stage of the disparity computation. The others are for computing disparity volume. In particular, the second equation computes the updating information on A and b. Using these parameters, the last two equations compute the disparity volume. Incidentally, the last equation supplies parameters for the second equation.

In more abstract form, Fig. 6. shows the overall system that computes (3). The first block denotes the transformation of the input image to the phase map. Standard stereo matching occurs in the second block. The last block denotes for the contextual matching. Also shown is the parameter updating part that compares the outputs of the two blocks and adjusts the parameters accordingly.

Assume consider that the image is an $N \times N$ array of pixels. As we can see in Fig. 2, the overall system consists of feature extraction, matching, and labeling. Among these, the phase computation in the feature extraction stage requires $\vartheta(N^2)$ multiplications. The necessary computation of the matching stage requires $\vartheta(N^2)$ multiplications[11], where K denotes maximum iteration number that the user must specify in advance.

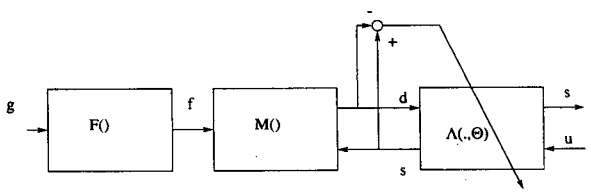


Fig. 6. Overall system.

In the labeling stage, two separate computations, *i.e.*, relaxation and parameter estimation, exist. Assume that the gate size is M. Both equations in (22) take $\vartheta(M^2N^2)$ time. Also the learning part in (29) takes $\vartheta(K \cdot M^2N^2)$time, where K is the maximum number of iterations. Note that the learning needs $\vartheta(KM)$ times the disparity computation time. In fact, the parameter need not be updated frequently. It must be updated only when the types of scenes, *i.e.*, indoors and outdoors, are changed.

For example, if N=512, M=16, and K=4, then the labeling stage takes 1 M multiplications and the learning part takes 1 G multiplications.

IV. Experimental Results

This section presents some experimental results of the new method for computing disparity. As typical examples, a 256 $\times$256 random dot stereogram and a 512$\times$512 pentagon image pair are chosen.

To see exactly what advantages and disadvantages the new scheme possesses, we planned three steps of experiments. First, a bottom-up process computing disparity volume from a given disparity map is examined. The purpose is to see how efficiently the additional stage improves the usual matching result. Conversely, the top-down process that computes disparity map from the given disparity volume, is investigated to show that information from the high level stage can improve the low level matching. Finally, an evaluation of the closed loop formed by the bottom-up and top-down controls, will display snap shot views that show how the solutions are improved dynamically as the computation flows along the loop. Of course, all parameters in the labeling system are automatically computed by (29).

1. Experiment Conditions

The scheme has been programmed in C following conventional procedural programming methods and run on a standard workstation. Two types of images are used throughout the experiments: synthetic and natural.

Fig. 7 shows a random dot stereogram and desired disparity map.

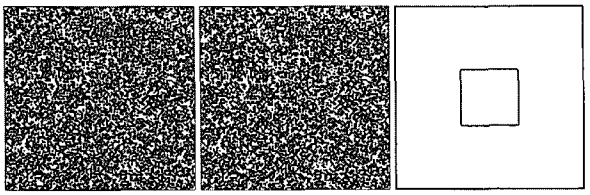


Fig. 7. A pair of random dot stereogram and its disparity map.

Only three different levels of disparity 0, 4, and 8 pixels are used.

Fig. 8 is a pair of 512×512 pentagon images and a manually generated reference disparity map.

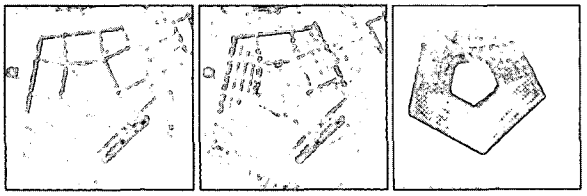


Fig. 8. A pair of 512×512 pentagon images and a manually generated disparity map.

With these images, we can start the following experiments.

2. Labeling after Matching

As a first experiment, a disparity volume has been computed from a given disparity map. The required equations are (29) for determining parameters and (23) for computing disparity volume. This experiment has been performed separately for the two types of images.

1) Disparity Volume of Random Dot Stereogram

Using the random dot stereogram, we first obtained the disparity map by (8) as explained in § 3. This input is used for computing disparity volume. Since the disparity volume can be easily converted to a representation in the form of a disparity map, we will show the result in this manner, though we still call it disparity volume.

At first, a learning stage described by (29) computes parameters A and **b**. The plot in Fig. 9 denotes the vector **b** which shows disparity distribution.

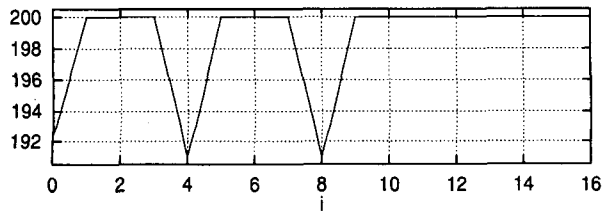


Fig. 9. The elements in **b**.

Notice that three positions of **b** have small values compared to the others. This means that the number of different disparities involved in the random dot stereogram are three. Since the center element **b** corresponds to zero disparity and the distance between the center element and another lower position of the vector indicates the disparity, the major disparities are 0, 4 and 8. Next, Fig. 10 shows a 3D representation of the parameter A.

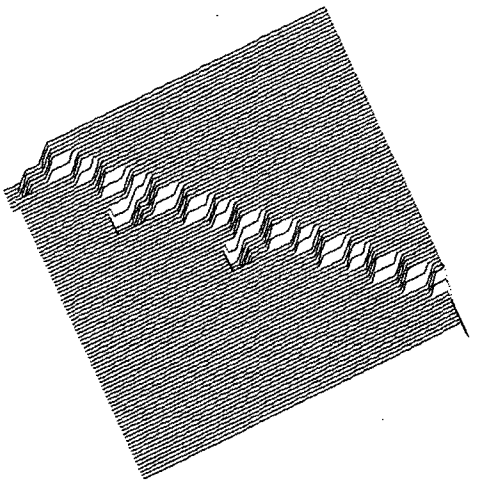


Fig. 10. A 3D plot of A.

In this figure, the columns and rows are represented by the x and y axes, respectively. The height, shown in log scale, of each element denotes the value at that position. The maximum and minimum numbers are 201.40 and -12.40 respectively.

Conceptually, the precedence rule has been transformed into the matrix A; the larger the number is, the harder the matching becomes. In fact the upper diagonal of this matrix are positions where the precedence rule is violated.

Upon seeing this figure, one can notice that the major disparities are 0, 4, and 8 again. Recall that the random dot stereogram has been generated with only three types of disparities. The diagonal elements of matrix A show the disparity distribution and the other elements show the disparity transitions. Since the matrix contains only a few types of elements, modeling this matrix by a few parameters will greatly simplify the computation.

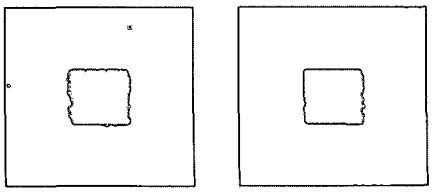


Fig. 11. Left: a given disparity map and right: disparity volume.

Fig. 11 shows a given disparity map and the disparity volume, respectively. The disparity volume is converted to a disparity map for convenience, as explained previously. With the left image as the input, the learning part updates (29) with μ = 0.01. Also, the estimated parameters α and β are 0.1 and 0.5, respectively.

It is obvious that the right image shows some improvements over the left image, especially near the left boundaries. Also, the right boundaries have been improved. This fact is apparent in Fig. 12 which shows only the occlusion areas.

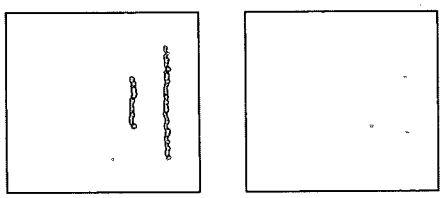


Fig. 12. Left: occlusion region of the disparity map and right: occlusion region of the disparity volume.

Almost all regions satisfy the precedence rule after the labeling computations. To estimate the algorithm quantitatively, let's compute the noise effect. This is possible since the exact solution, *i.e.*, the desired disparity map in Fig. 7 is available. The amount of noise is controlled by the signal-to-noise ratio (SNR):

$$SNR \triangleq \frac{\sigma_s^2}{\sigma_n^2}, \tag{31}$$

where σ_s^2 and σ_n^2 denote respectively the variance of the signal and *i.i.d.* Gaussian noise. Denoting $\mathbf{d}^*$ and $\mathbf{d}$ respectively the disparity for the original image and the estimated disparity from the noisy image, one can calculate the root-mean-square(RMS) error:

$$RMS\ error \triangleq [\frac{1}{N^2}\sum_{s\in S}(d_s^* - d_s)^2]^{\frac{1}{2}}. \tag{32}$$

Varying SNR for the RDS, we obtained the disparity map and the disparity volume analogously to the previous example. The corresponding RMS errors are summarized in Fig. 13.

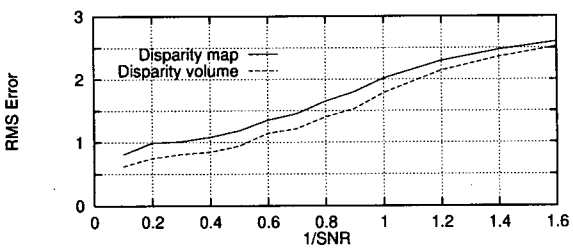


Fig. 13. The RMS errors vs 1/SNR for the disparity map and disparity volume.

Note that the disparity volume is better than the disparity map under any noise condition.

2) Disparity Volume of Pentagon Image

The parameters are computed first with the initial disparity map and reference disparity map shown in Fig. 8. The parameter **b** is shown in Fig. 14 and A is shown in Fig. 15. The magnitude of the elements of matrix A are in log scale. The maximum and minimum are, respectively, 203.0 and -8.0.

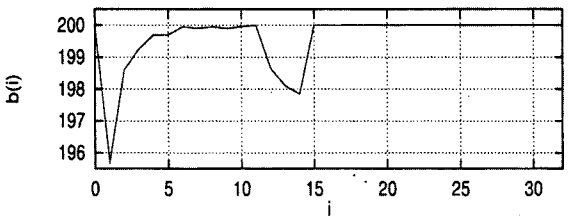


Fig. 14. A 2D plot of vector **b**.

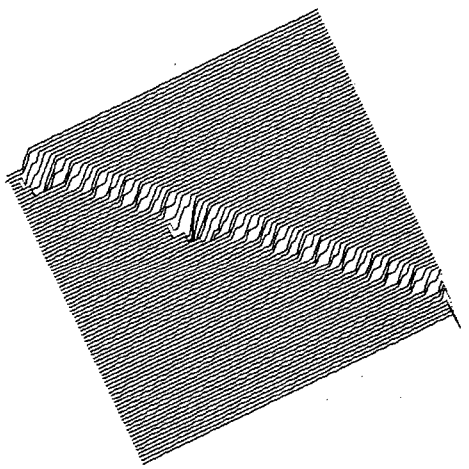


Fig. 15. The elements of A.

The learning part used α =12.0 and β =15.0. Also, μ in (29) is 0.01.

In Fig. 16, the left image shows the disparity map from the matching system and the right image shows the disparity map from the labeling system.

Fig. 16. Left: disparity map from the matching system and right: disparity map from the labeling system.

3. Matching Guided by Labeling

The second stage of experiment is to examine the mechanism of the downward control flow, computing the disparity map with disparity volume as an input. It is expected that the resulting disparity map must be better than if no hints from above are used.

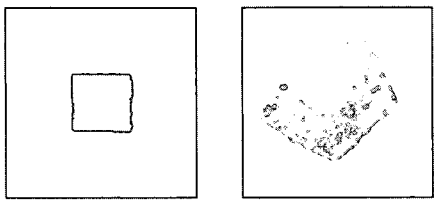


Fig. 17. Matching result.

The left image is the disparity map for the random dot stereogram without noise and the right for the pentagon image pair. We can see the improvement in the disparity maps from a comparison of the left image in Fig. 11 and in Fig. 17.

For the random dot stereogram, RMS errors vs 1/SNR is shown in Fig. 18 for the initial disparity map and improved one.

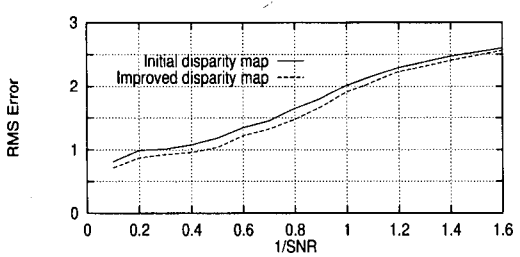


Fig. 18. The RMS errors vs 1/SNR for the initial disparity map and improved disparity map.

Notice that the disparity map after labeling computation is better than the disparity map without hints from the high level processes.

4. Experiment for Closed Loop

Fig. 19 shows a snap shot view for the random dot stereogram image pair with *SNR* = 10. Also a snap shot view for the disparity volume is shown in Fig. 20. For these conditions, RMS errors vs number of iterations is shown in Fig. 21.

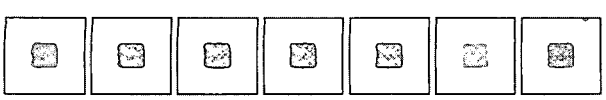


Fig. 19. A snap shot view for the disparity map.

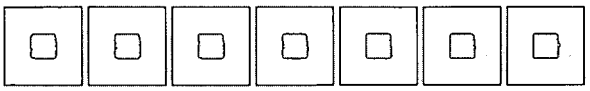


Fig. 20. A snap shot view for the disparity volume.

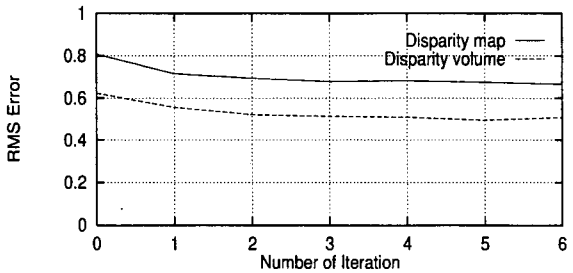


Fig. 21. The RMS errors of the disparity map and the disparity volume vs iteration number.

5. Discussion

From the experiment of parameter estimation, one can notice that the parameter **b** and A contain some information of the disparity map. The major disparity distributions can be found from the positions of the lower elements of **b** and lower diagonal elements of A. The precedence rule has been transformed into the matrix A. The upper diagonal elements of matrix A have large numbers to prevent violation of the rule. If we know the disparity distributions, we can make the parameters **b** and A manually and simplify the computation. In the total system, consisting of feature extraction, initial matching, parameter estimation, labeling and matching with disparity volume, the parameter estimation part takes the most time. Therefore we used only one row for estimating parameters in this experiment. The overall computation time for one loop is 8 minutes 6 seconds in ultra sparc workstation.

VII. Conclusion

From the assumption of that current binocular stereo image matching algorithms based on bottom-up control alone are susceptible to local minima, we introduced a bottom-up and top-down feedback computational algorithm to improve the overall performance.

Conceptually the result from the lower level is improved at the high level in the tightly coupled feedback loop and also the result from the high level helps the low level decisions. We first hypothesize the image attributes by an MRF and thereby define the stereo matching problem by MAP estimation for this bottom-up and top-down feedback loop system. For the conventional bottom-up computation, we

follow the multilevel approach in Kim and Jeong[11] with slight modifications.

In the top-down computation, we first derive energy equation for solving the MAP estimate and next derive iterative equations for obtaining the optimal solutions. For finding the parameters, we use the Lagrange multiplier and maximum likelihood estimation. The overall equations for finding the solution have iterative forms.

From experiments using RDS and pentagon image pairs, we first obtained the disparity volume at the labeling part based on the results of the matching system. Investigation the top-down process shows that the result of the labeling stage can improve the matching result. An evaluation of the performance of the closed loop was also performed. All the parameters in the labeling system are computed by the proposed method.

Acknowledgments

This research was conducted with the support from KRF, Korea in 1996.

References

[1] N. Ayache and B. Faverjon, "Efficient registration of stereo images by matching graph descriptions of edge segments," *International Journal of Computer Vision*, 1(2), pp. 107-131, 1987.

[2] H. H. Baker and T. O. Binford, "Depth from edges and intensity based stereo," *Proc. 7th Intern. Joint Conf. Artif. Intell.*, pp. 631-636, Aug., 1981.

[3] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Statis. Soc.*, Ser. B36, pp. 192-236, 1974.

[4] L. Dreshler and H. H. Nagel, "Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a stereo scene," *Comput. Graph. Image Process,* 20, pp. 199-228, 1982.

[5] D. J. Fleet, *Measurement of Image Velocity*, Kluwer Academic Publishers, 1992.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images", *IEEE Trans. on PAMI*, Vol. PAMI-6, pp. 721-741, 1984.

[7] F. Glazer, G. Reynolds and P. Anandan, "Scene matching by hierarchical correlation," *Proc. 1st IEEE Conf. Computer Vision Pattern Recognition*, pp. 432-441, June 1983.

[8] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, 17, pp. 185-203, 1981.

[9] B. Julesz, "Texture and visual perception," *Scientific America*, Feb., 1965.

[10] M. Kass, "A computational framework for the visual correspondence problem," *Proc. DARPA Image Understanding Workshop*, 1983.

[11] J. G. Kim and H. Jeong, "Multilayer stereo image matching based upon phase magnitude and mean field approximation," to *Journal of Electrical Engineering and Information Science,* Vol. 2, pp. 79-88, 1997.

[12] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Roy. Soc. London*, B. 204, pp. 301-328, 1979.

[13] J. E. W. Mayhew and J. P. frisby, "Psychophysical and computational studies towards a theory of human stereopsis," *Artificial Intelligence*, Vol. 17, pp. 349-385, 1981.

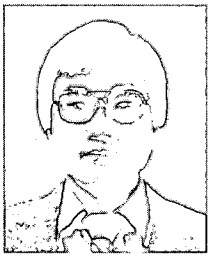
[14] T. D. Sanger, "Stereo disparity computation using Gabor filters," *Biological Cybernetics*, 59, pp. 405-418, 1988.

[15] A. M. Waxman, "An image flow paradigm," *Proc. Workshop on Computer Vision: Representation and Control*, pp. 49-57, 1984.

[16] J. Weng, "Image matching using the windowed Fourier phase," *International Journal of computer Vision*, 11:3, pp. 211-236, 1993.

[17] C. S. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields," *CVGIP: Graphical Models and Image Processing*, Vol. 54, No. 4, pp. 308-328, July, 1992.

Jung-Gu Kim received the B.S. degree in the Department of Electrical Engineering from Kyoung-Buk National University, Korea, in 1991, and the M.S. and the Ph.D. degrees in the Department of Electronic and Electrical Engineering from Pohang University of Science and Technology(POSTECH) in 1993 and 1998, respectively. He is a senior researcher in Power Electronics Research Team of Research Institute of Industrial Science and Technology(RIST), Pohang, Korea. His current research interests include image processing, stereo vision, computer and machine vision and it's application.



Hong Jeong received B.S. degree in the Department of Electrical Engineering from Seoul National University, Korea, in 1977. In 1979, he received the M.S. degree in the Department of Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST). In 1984, 1986, and 1988, he received the S.M., E.E., and Ph.D. degrees, respectively, all in the Department of Electrical Engineering and Computer Science at M.I.T., Cambridge, Massachusetts, U.S.A. During the period of 1979-1982, he was a faculty staff at the Department of Electrical Engineering at the Kyoung-Buk National University, Daegu, Korea. He is a Sigma Xi member. During 1994-1995, he worked as a vicechairman in the Special Interest Group on Neurocomputing in the Korea Information Science Society. Since 1988, he has worked in the Department of Electrical Engineering at the Pohang University of Science and Technology, where he now works as a Processor. His research interests include digital signal processing, computer vision, speech recognition, and radar signal processing.