Multilayer Stereo Image Matching Based upon Phase-Magnitude and Mean Field Approximation

Hong Jeong, Jung-Gu Kim, and Myoung-Sik Chae

Abstract

This paper introduces a new energy function, as a maximum a posteriori(MAP) estimate of binocular disparity, that can deal with both random dot stereo-gram(RDS) and natural scenes. The energy function uses phase-magnitude as features to detect only the shift for a pair of corrupted conjugate images. Also we adopted Fleet singularity that effectively detects unstable areas of image plane and thus eliminates in advance error-prone stereo matching. The multi-scale concept is applied to the multi layer architecture that can search the solutions systematically from coarse to fine details and thereby avoids drastically the local minima. Using mean field approximation, we obtained a compact representation that is suitable for fast computation. In this manner, the energy function satisfies major natural constraints and requirements for implementing parallel relaxation. As an experiment, the proposed algorithm is applied to RDS and natural stereo images. As a result we will see that it reveals good performance in terms of recognition errors, parallel implementation, and noise characteristics.

I. Introduction

It is well known that we human can recognize 3D world using only a slightly different image pair taken from spatially separated position. This stereo vision is the fundamental and basic topic from the beginning of research in the field of computer vision. The final goal of stereo vision is a reconstruction of 3D world using two or several images taken from single or several cameras. However, it is yet far from completion since there are several unexpected conditions such as noise, lighting difference between two images, quantization error, acquisition errors, computational demand, and so forth. This section first introduces the stereo matching and then discusses typical stereo algorithms.

1. Stereo Matching

Stereo image matching from the assumption of epipolar line and parallel image plane is focused on finding conjugate points corresponding to two images, and this matching result is used for the 3D world reconstruction. As can be seen from Fig. 1, available data in binocular stereo vision system are two different intensity valued images and some mechanical parameters.

In this figure, two cameras with *focal length F* are separated with the so called *base line B*. When we let the position of

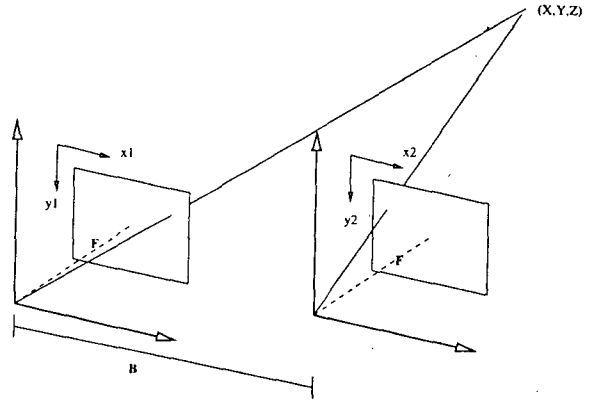


Fig. 1. Binocular stereo vision system.

arbitrary 3D object point as (X, Y, Z) and two reference axes for two images as (x_1, y_1) and (x_2, y_2), the distance between (X, Y, Z) and the center of lens is easily obtained from trigonometric relation:

$$Z=\frac{B\cdot F}{d}, \quad d=x_1-x_2, \tag{1}$$

where d called *disparity* is the distance between two corresponding points of the left and right images. The goal of matching in stereo vision is to find this disparity value for all points in image.

Manuscript received March 26, 1997; accepted September 24, 1997.

H. Jeong and J. G. Kim are with Department of Electronics Engineering, Pohang University of Science and Technology, Pohang 790-784, Korea.

M. S. Chae is with SK Telecom, Hwa-Am dong 58-4, Daejeon 305-348, Korea.

2. Algorithms for Stereo Matching

Our approach to the problem is based upon the MRF(Markov random field) hypothesis of the image attributes such as image intensity, phase, disparity, and discontinuities. By this hypothesis, we can mathematically define an optimal matching as a MAP estimation, though other estimation measures are possible. This line of research ultimately leads to the most crucial step that must define an energy function in an equilibrium state that describes the optimal solution satisfying faithfully the natural constraints. After then we can suggest a computation architecture that minimizes the energy function.

The energy function must make the following selections: image intensity vs. phase-magnitude as features describing the original images, single-layer vs. multi-layer as computational structures for effectively searching the optimal solutions, and selecting candidate matching sites at a given scale. In conclusion, our concern is how to represent the phase-magnitude, multi-scale, and *singularity map* compactly in a single energy function. The point is that the phase-magnitude has an advantage of detecting shift only from noisy signals. The multi-scale can guide the algorithm to escape from the local attractors and instead to converge to the global minima. The other help is the singularity map that eliminates uncertain sites for matching. Combined together, all these properties are aimed at finding optimal matches avoiding attracting local minima.

In many ways, the feature-based method contrasts with the area-based method. Conceptually, the area-based method mainly uses intensity[20, 35], intensity correlation[16], intensity partial derivative[22] and intensity directional variance[29], and the feature-based method uses several possible features such as line segments [1, 2, 8], edges[3, 17, 18, 19, 23, 27, 32], peak points[3, 27] and zero crossing points[26] for matching primitives. However, intensity-based method has some problems in real situation that two matching points in the image usually have different intensity values[30]. On the other hand, the feature-based method needs one more step for interpolation after matching. As a new approach, some researchers proposed a phase difference method[9, 10, 11, 21, 25, 33, 36, 37]. However, these methods have difficulties in constructing unique surface and ignore the prior property of disparity. To cope with this problem, we heavily use prior knowledge by modeling disparity field, along with other attributes, as Markov random field(MRF).

The concept of multi-scale is best described by the multi-layer architecture. At the top of the multi-layer, a global view of interpretation is first obtained and naturally limits the search space in the next lower level. At the lower level, the child processes start from the initial values inherited from the parent and find optimal solutions for the given conditions. As a result, the state in the finest level contains the solution. This downward control flow can drastically narrow down the search spaces pruning unnecessary branches. Note however, as we will see, that this algorithm is different from the pyramid algorithm in that each layer uses the phase values of the sites in the finest level. Finally, the singularity map is an effective way to remove uncertain areas from matching and can be naturally integrated into the problem formulation.

The organization of the rest of this paper is as follows. In § 2, we define the stereo matching as a MAP estimate. This estimate is further reduced to the energy function in § 3. After then, we will show how to implement our algorithm in a parallel relaxation scheme in § 4. § 5 contains experimental results analyzing and comparing our algorithm with others.

II. Problem Formulation

To begin with, this section addresses defining the stereo matching in terms of Bayes estimation. In order to do this, we must first transform the signal into phase domain and then consider the problem in relation with phase. Naturally, this section addresses Garbor transformation, phase singularities, multi-layer algorithm, and MAP estimation.

1. Phase Representation

A signal can be expressed in terms of many forms such as amplitude, frequency, and phase. Among these, the phase has very important informations of the original signal as can be seen from the experiment of Oppenheim and Lim[31]. One of the most popular method of local phase extraction from image is Gabor function[9, 10, 11, 33]. Many different functions such as modulated window of square and sinusoidal waves[36, 37], Hilbert transform pair[13, 34], log-normal filter[8], and the first and second derivatives of Gaussian function[38] also have been proposed for the phase extraction. Among these functions, we will use Gabor function for phase computation.

Let's denote the intensity functions of the image at x as $F(x)$. Also we consider $G(x, u_0)$ as the result of signal $F(x)$ filtered with the Garbor filter $h(x, u_0)$ In other words, we have

$$G(x, u_0) = F(x) * h(x, u_0). \tag{2}$$

The Garbor function $h(x, u)$ has the form of

$$h(x, u) = e^{-x^2/2\sigma^2} e^{j2\pi ux}, \tag{3}$$

where σ represents the parameter determines the window size. For the purpose of improvement of frequency resolution, we used variable size of windows. Small size of window was used for high frequency signal region and large size of window was used for low frequency region. And we set

$$\sigma = \frac{K}{u}, \tag{4}$$

where u is center frequency of filter and K is a constant. In this

paper, we set $K = 0.5$. Finally, the phase is defined as

$$\phi \triangleq \arg[G(x, u_0)]. \tag{5}$$

2. Singularity in Phase Space

When we calculate phase, there are some parts of signals, called *singular points*, where the value of phase is very sensitive to the filter size. Since these portions correspond to the distortion of input signal, careful consideration has to be done before matching.

Fleet[10] proposed the following conditions for detection of these singular points.

$$\begin{cases} \sigma(u) \left| \frac{\partial}{\partial x} \phi(x, u) - 2\pi u \right| & \geq \tau_1, \\ \sigma(u) \frac{\left| \frac{\partial}{\partial x} A(x, u) \right|}{A(x, u)} & \geq \tau_2, \end{cases} \tag{6}$$

where $\sigma(u)$ denotes that σ is a function of u. τ_1 and τ_2 are parameters for the arbitrary threshold. The first in (6) indicates the region of signal that the difference between frequency of filtered signal and center frequency is greater than some threshold; the second denotes the region where amplitude decreases too much. Fig. 2 shows an example for the singular points.

Fleet removed the singularity points before disparity computation and reduced disparity errors. After matching, disparity values at the singular points must be interpolated with the value of neighbors. Analogously, we will derive an energy function that includes this singularity in an efficient manner.

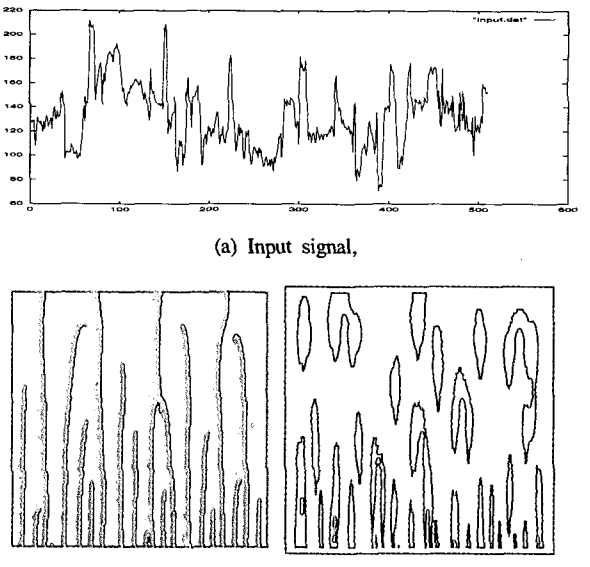
3. Multilayer Structure

Let's consider a pyramid built from an $N \times N$ image plane S. The plane S_i, ($i \in [0, \ldots, \log_2 N]$) are derived successively from the lower plane by reducing its size by half. (However, we set $S_0 = S$.) Then S_i is denoted by the $(N/2^i)^2$ grid. For each layer i, the disparity vector d_i is defined by $d_i = \{d_s, s = (x, y) \in S_i\}$ and $d_i \in D_i$, where D_i is the range space of disparity vector. Our purpose is to find an optimal disparity d^* in $D_i, \forall i$, from top to bottom.

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

It is easy that $\sum_{s \in B_i} 1 = 2^{2i}$, $\sum_{(x,y) \in S_i} 1 = (N/2^i)^2$, and therefore, $\sum_{s \in B_i} \sum_{(x,y) \in S_i} 1 = N^2$. This relationships will be used in estimating computational complexity. For example, for a 16×16 S, Fig. 3 shows S_2, S_3 and S_4 that are respectively 4×4, 2×2, and 1×1. Also shown are the blocks B_2, B_3 and B_4 that are respectively 4×4, 8×8, and 16×16.

According to this method, one can determine d_i using all the phase values in the original plane S. Let's consider a general scheme for finding an optimal solution a. 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 D_i. This value is now used as initial conditions in D_{i-1} and similar operations continues up until $i = 0$; the final solution resides in



(a) Input signal,

(b) Phase information with increasing scale, (c) Singular points.

Fig. 2. Detection of singular points.

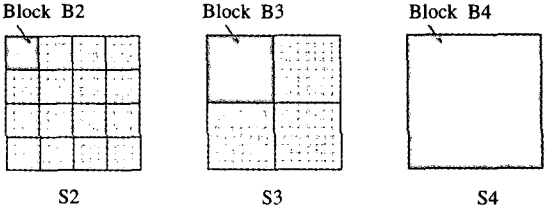


Fig. 3. Blocks B_2, B_3 and B_4 when S is 16×16.

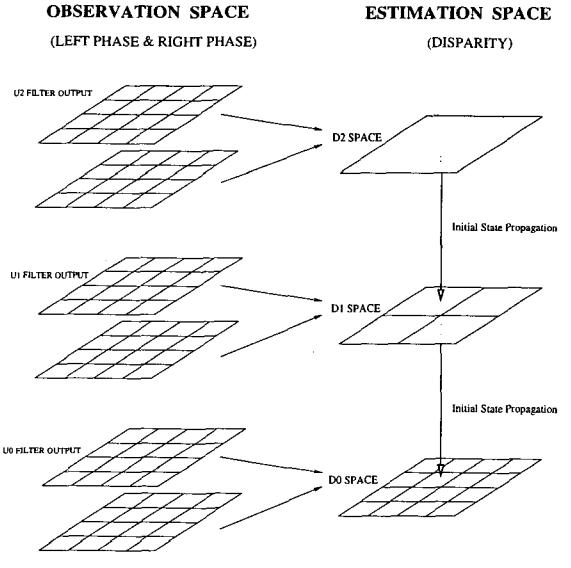


Fig. 4. Finding solutions successively in different filter scale.

D_0. This method is described in Fig. 4.

This computation is very similar to conventional pyramid structure[5, 6] but basically different in that the phase for all the N×N cells are used for each layer.

4. MAP Estimate

Let S be the set of N^2 lattice and F a random field on this lattice: $S=\{s_1, s_2, \cdots, s_{N2}\}$ and $F=\{F_s, s\in S\}$. Our intention is to consider S as an image plane and F as the attributes defined on it. When the field F satisfies the following condition of local dependency, it is MRF.

$$\begin{cases} P(f)>0, \quad \forall f \\ P(F_s=f_s|F_r=f_r, r\neq s)=P(F_s=f_s|F_r=f_r, r\in N_s). \end{cases} \tag{7}$$

where N_s denotes the neighborhood of lattice s. However, this MRF condition cannot be implemented to the real computation of conditional probability. It is well known that, according to the Hammersley-Clifford theorem[4] on the equivalence of MRF and Gibbs, the distribution of total field has the form[15]:

$$P(F=f)=\frac{1}{Z}\exp[-U(f)], \tag{8}$$

where Z is a partition function and $U(f)$ is an *energy function* which can be represented as a sum of click potential V_c,

$$\begin{cases} U(f) = \sum_{c\in C} V_c(f), \\ Z = \sum_f \exp[-U(f)], \end{cases} \tag{9}$$

where the click is defined as a point or subset of S in which all points are neighbors with each other. Fig. 5 shows clicks in the second order neighborhood system.

We call a *click system* C as a set of clicks. The image attributes we are assuming MRF here are image intensity, disparity, and disparity discontinuity.

Let's assume that the disparity field is real MRF within the space S_l, which consists of lattice B_is. We further introduce *line processes* $L=(L^h, L^v)$ to denote the discontinuities in the horizontal and the vertical directions of the disparity map. In particular, our concern is the MAP estimate of the disparity map:

$$P(d,h,v|\phi^L,\phi^R)=\frac{P(\phi^L|\phi^R,d,h,v)P(d,h,v|\phi^R)}{P(\phi^L|\phi^R)}, \tag{10}$$

where ϕ^L and ϕ^R stand for phase values of the left and right images obtained on the lattice S using the filter of center frequency u_l. Keep in mind that this is for a layer S_l. As we will see, the values inherited from the parent is used as an initial value.

The phase is limited in $[-\pi, \pi]$ with distance $1/u_0$ since the filtered signal with the low pass filter of center frequency u_0 consists of sinusoidal functions of the frequency u_0. Therefore, corresponding points can be found when the disparity is smaller than $1/2u_0$. For this region, hierarchical pyramids of input images

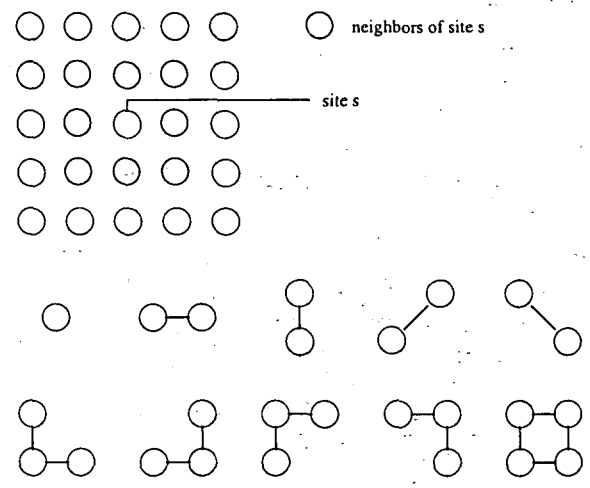


Fig. 5. Overall computation structure.

or filtered images were used in [10, 37]. In this paper, we used the method of reducing of state space of disparity vector corresponds to the filter center frequencies.

III. Energy Function for Disparity Computation

Our goal is to obtain a disparity map using the phase information and singularity map. As an assumption, the attributes, such as phase, disparity, and discontinuities, assigned to each pixel can be modeled by MRF. Then, this model makes it possible to derive a MAP estimate of the disparity. Finally, as a likelihood of this estimate, an energy function is defined.

1. Phase Matching

$P(\phi^L \mid \phi^R, d, h, v)$ in (10) denotes the amount of matching error, with the given disparity and line processes, for the two phases at a given site of the left image. It is natural to assume that this relationship is independent from the discontinuities and remove the line processes from the expression. If we assume this matching error as Gaussian with variance σ^2, then $P(\phi^L \mid \phi^R, d, h, v)$ becomes

$$\begin{cases} P(\phi^L|\phi^R,d,h,v) \triangleq \frac{1}{Z_2}\exp[-U(\phi^L|\phi^R,d)], \\ U(\phi^L|\phi^R,d)=\sum_{(x,y)\in S_l}\sum_{s\in B_i}\frac{1}{2\sigma^2}(1-q_s^l\oplus q_s^r)(\phi^L_{s+d(B_i)}-\phi^R_s)^2. \end{cases} \tag{11}$$

where the subscript s denotes a lattice point of S, and q_s^l and q_s^r represent indicators of singular points for each left and right image. This indicator is set to *1* at the singular point and *0* elsewhere. The operator $\oplus$ stands for logical OR. B_l is a block in S corresponding to $(x, y)\in S_l$; $d(B_i)$ is disparity value in B_l. The term $(s+d(B_i))$ represents the shift (or the amount of disparity) along the epipolar line.

2. Prior Knowledge

In (10), $P(\phi^L \mid \phi^R)$ is a constant and can be ignored for the further discussion. The term $P(d, h, v \mid \phi^R)$ represents the prior for disparity and line processes. Assuming that the disparity and line processes are MRF and considering that ϕ^R and (d, h, v) are mutually independent, one obtains the Gibbs distribution:

$$\begin{cases} P(d,h,v|\phi^R) \triangleq \frac{1}{Z_1}\exp[-U(d,h,v)], \\ U(d,h,v) = \sum_c V_c(d,h,v), \end{cases} \tag{12}$$

where Z_1 is a normalization constant and $V_c(d, h, v)$ is the click potential for a click \$c\$. We model the potential function as

$$\begin{aligned} V_c(d,h,v) \triangleq\ & \mu[(d_{x,y}-d_{x,y-1})^2(1-l^h_{x,y})+(d_{x,y}-d_{x-1,y})^2 \\ & (1-l^v_{x,y})]+\gamma(l^h_{x,y}+l^v_{x,y}), \end{aligned} \tag{13}$$

where $d_{x,y}$ is the same meaning with $d(x,y)$ for $(x,y)\in S_i$. Notice that the line process $(l^h_{x,y}, l^v_{x,y})$, representing discontinuities in the 3D surface, turns on and off the smoothness terms for the disparity. The second term in (13) is introduced to inhibit too much line processes. As a result, the energy will be minimal when the line processes are turned on at the actual discontinuous sites and when the disparity functional is smooth on the object surface.

Substituting (13) and (11), (1) becomes

$$\begin{cases} P(d,h,v|\phi^L,\phi^R)=\frac{1}{Z}\exp[-U(\phi^L|\phi^R,d)-U(d,h,v)], \\ Z=\sum_{d,h,v}\exp[-U(\phi^L|\phi^R,d)-U(d,h,v)]. \end{cases} \tag{14}$$

where

$$\begin{aligned} U(d,h,v|\phi^L,\phi^R) =\ & \sum_{(x,y)\in S_i}\sum_{s\in B_i}\frac{1}{2\sigma^2}(1-q^l_s\oplus q^r_s)(\phi^L_{s+d(B_i)}-\phi^R_s)^2 \\ & +\mu\sum_{(x,y)\in S}[(d_{x,y}-d_{x,y-1})^2(1-l^h_{x,y}) \\ & +(d_{x,y}-d_{x-1,y})^2(1-l^v_{x,y})] \\ & +\gamma\sum_{(x,y)\in S}(l^h_{x,y}+l^v_{x,y}). \end{aligned} \tag{15}$$

3. Approximating Energy Function

So far, we have derived an energy function (15) for finding disparity and line processes. The next step will be to find the optimal solution that minimizes the energy function. There are several methods for this purpose such as stochastic relaxation[24] and deterministic relaxation method. Since the stochastic relaxation method takes too much time for computation, we will use deterministic method. In order to achieve this goal, we must simplify further the energy equation by the mean field approximation[7].

Note that in (14) the partition function Z is the sums of all (d, h, v). Since the line process can be replaced by the disparity difference, the energy function can be represented solely by a functional of a:

$$\begin{aligned} Z \triangleq\ & \sum_d \exp[-U_e(d)], \\ U_e(d) =\ & \frac{1}{2\sigma^2}\sum_{x,y}\{\sum_{s\in B_i}(1-q^l_s\oplus q^r_s)(\phi^L_s-\phi^R_{s+d(B_i)})^2\}+\sum_{x,y}2\gamma \\ & -\sum_{x,y}\log[1+\exp(\gamma-\mu(d_{x,y}-d_{x-1,y})^2)] \\ & -\sum_{x,y}\log[1+\exp(\gamma-\mu(d_{x,y}-d_{x,y-1})^2)]. \end{aligned} \tag{16}$$

We will call $U_e(d)$ an *effective energy function*. From this effective energy, we can define a new pdf $P^m(d_{x,y})$ which consists of $d_{x,y}$ at (x,y) and mean values at otherpositions. This means that one can neglect $d_{R,m} \mid_{R\neq x, m\neq y}$ when we compute mean $\langle d_{x,y}\rangle$ of $d_{x,y}$.

$$\begin{cases} P^m(d_{x,y}) \triangleq \frac{1}{Z^m}\exp[-U^m(d_{x,y})], \\ U^m(d_{x,y}) = \frac{1}{2\sigma^2}\{\sum_{s\in B_i}(1-q^l_s\oplus q^r_s)(\phi^L_s-\phi^R_{s+d(B_i)})^2\} \\ \qquad -\sum_{(k,m)\in N_{x,y}}\log[1+\exp(\gamma-\mu(d_{x,y}-\langle d_{k,m}\rangle)^2)], \end{cases} \tag{17}$$

where $N_{x,y}$ is the first order neighborhood at (x,y). $U^m(d_{x,y})$ is the energy we have sought for so far.

Ⅳ. Realization

Computing disparity is reduced to a problem for finding minima of the energy function $U^m(d_{x,y})$. In this section, we suggest a relaxation method that successively find the solutions in parallel and iterative manner.

1. Suboptimal Condition

The suboptimal value of $d_{x,y}$ in (17) can be obtained by

$$\left.\frac{\partial U^m(d_{x,y})}{\partial d_{x,y}}\right|_{d_{x,y}=\langle d_{x,y}\rangle}=0. \tag{18}$$

An optimal solution $\langle d_{x,y}\rangle$ must satisfy (18) but a closed form is impossible. Instead, we are looking for a solution with the gradient of energy function. To proceed further, we define the auxiliary function

$$l_{x,y}(x-1,y) \triangleq \frac{1}{1+\exp(\gamma-\mu(d_{x,y}-\langle d_{x-1,y}\rangle)^2)}. \tag{19}$$

Putting (19) into the left term of (18), we have

$$\begin{aligned} \left.\frac{\partial U^m(d_{x,y})}{\partial d_{x,y}}\right|_{d_{x,y}=\langle d_{x,y}\rangle} =\ & -\frac{1}{\sigma^2}\sum_{s\in B_i}(1-q^l_s\oplus q^r_s)(\phi^L_s-\phi^R_{s+d_{x,y}})\frac{\partial\phi^{R,+d}}{\partial d_s} \\ & +\sum_{(k,m)\in N_{x,y}}2\mu(d_{x,y}-\langle d_{k,m}\rangle)(1-l_{x,y}(k,m)). \end{aligned} \tag{20}$$

Since we have to know the mean values before calculation, an iterative calculation will be suitable in real implementation.

2. Parallel Relaxation

For finding $\langle d_{x,y}\rangle$ which satisfies (18), a gradient descent method was adopted:

$$\langle d_{x,y}\rangle^{n+1}=\langle d_{x,y}\rangle^n-\tau\frac{\partial U^m(d_{x,y})}{\partial d_{x,y}}, \tag{21}$$

where τ is a parameter governing the convergence speed. Substituting (20) into (21) yields

$$\begin{cases} \langle d_{x,y}\rangle^{n+1} = \langle d_{x,y}\rangle^{n} + \alpha\Big[\sum_{s\in B_i}(1-q_s^l \oplus q_s^r)(\phi_s^L - \phi^R_{s+\langle d_{x,y}\rangle^n})\dfrac{\partial \phi^R_{s+\langle d_{x,y}\rangle^n}}{\partial d_s} \\ \qquad -\lambda \sum_{(k,m)\in N_{x,y}}(\langle d_{x,y}\rangle^n - \langle d_{k,m}\rangle^n)(1-l^n_{x,y}(k,m))\Big], \\ l^n_{x,y}(k,m) = \dfrac{1}{1+\exp(\gamma-\mu(\langle d_{x,y}\rangle^n - \langle d_{k,m}\rangle^n)^2)}, \end{cases} \quad (22)$$

where $\alpha=\frac{\tau}{\sigma^2}$ and $\lambda=2\sigma^2\mu$. For a given level, this computation can be achieved in parallel.

Let's denote the algorithm considering the computational layers. In 3, we defined *block* $B_i(i=0, 1,\ldots, \log_2 N)$. An element defined on (x,y) site in S_l 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 S. Then the algorithm can be conveniently represented by

$$\begin{cases} \langle d_{x,y}\rangle^{n+1} = \langle d_{x,y}\rangle^{n} + \alpha\Big[\sum_{s\in B_i}(1-q_s^l \oplus q_s^r)(\phi_s^L - \phi^R_{s+\langle d_{x,y}\rangle^n})\dfrac{\partial \phi^R_{s+\langle d_{x,y}\rangle^n}}{\partial d_s} \\ \qquad -\lambda \sum_{(k,m)\in N_{x,y}}(\langle d_{x,y}\rangle^n - \langle d_{k,m}\rangle^n)(1-l^n_{x,y}(k,m))\Big], \\ l^n_{x,y}(k,m) = \dfrac{1}{1+\exp(\gamma-\mu(\langle d_{x,y}\rangle^n - \langle d_{k,m}\rangle^n)^2)}, \end{cases} \quad (23)$$

where the initial conditions are $\langle d_{x,y,i}\rangle^0 = \langle d_{x,y,i+1}\rangle$ and this computation must be done $\forall(x,y)\in S_i$, $(i=\log N, N-1, \ldots 1)$. Notice that this computation must proceed for each block in each layer from top to bottom.

As Fig. 6 depicts the overall scheme, the system consists of two parts: feature and disparity computation.

The first block transforms the input image to obtain phase-magnitude pair and thereby computes a singularity map. These intermediate representations are used in the next block as inputs to the matching terms in (23). Inside this block the computation flows downwards from top and after all the output is contained as a state of the bottom layer.

In more detail, a cell computing (23) can be realized by Fig. 7.

It describes the activities inside a cell (or processor) and the connections between cells and inputs. The disparity downloaded from the parent is used as an initial value of the children cells. The processor is connected with its neighbor processors and receives their disparity values. Also external inputs are the $N\times N$ phase values, processed with scaled filters, and the values of their siguralities.

3. Computational Complexity

Let's consider the computational loads of this algorithm. At first, we can observe the first block in Fig. 6, that takes $O(N^2)$ multiplications. As we will see shortly, however, the second block takes much more time. To see this, let's consider (23) for a layer i and that the iteration is done only once for each pixel. The second term must be calculated for N^2 cells. The third term must be calculated for $(N/2^i)^2$ times. Therefore, $O(\alpha N^2+\beta(N/2^i)^2)$

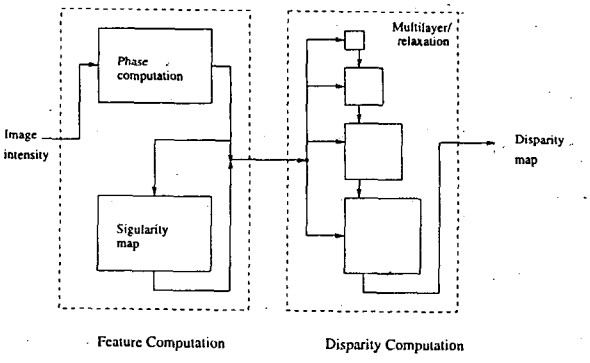


Fig. 6. Overall computation structure.

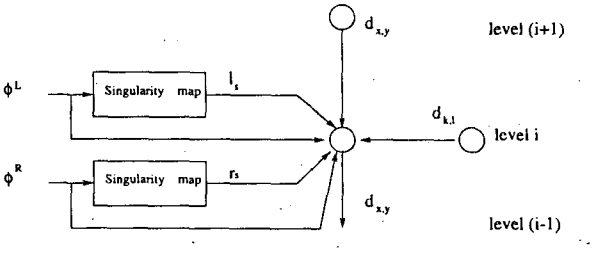


Fig. 7. Computation structure for a processor.

computations are need for a layer i. Here, typically α=3 and β =10 are constants. Therefore, we can expect that the computation load will increase as the computation proceeds from top to bottom.

Remember that we must compute the layers for i=0,1,..., $\log_2 N$. Summing for all the layers, we will reach $\alpha O(N^2\log_2 N)$. If the iteration is done K times for each pixel. Then we have the overall computation $K\cdot\alpha O(N^2\log_2 N)$. For example, if K=10 and N=512, then the overall computation will be around 128M multiplications. The recent computers with 64MFLOPS will execute this computation in a couple of seconds. However, actual computation is much slower, contrary to the expectation, due to communication bottleneck between CPU and I/O.

V. Experimental Results

The algorithm has been programmed and tested by a SUN ultra Sparc workstation. As typical examples, we chose a random dot stereo-gram and a pentagon image. All the size of the test images is 512×512 and the filter scales are σ=16, 8 and 4. The final disparity maps were obtained using (23) for each image pair. For each test, we used the parameters λ=2.0 μ=1.0 and γ=7.0 that were heuristically chosen.

1. Experiment with Random Dot Stereogram

The random dot stereo-gram image pair in Fig. 8 has three layers of different disparities: 0, 8 and 15 pixels.

The left and right image pair are respectively shown in top. Also the desired disparity map and its 3D perspective are shown

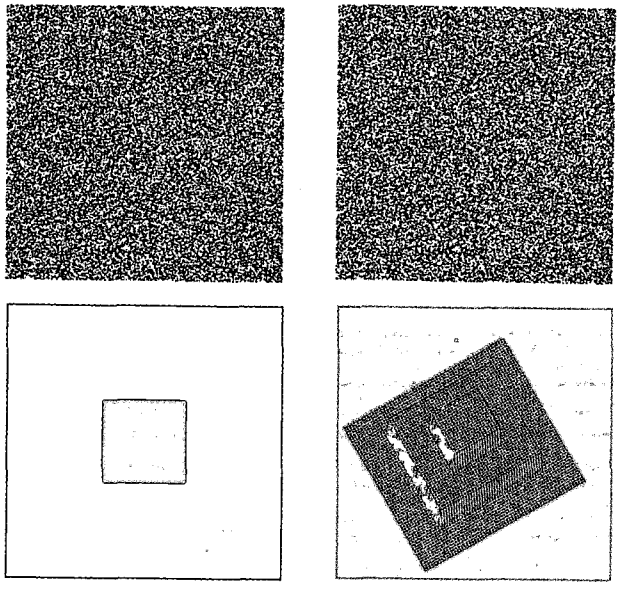


Fig. 8. Random dot stereo-gram; at top an RDS pair and below a disparity map and its 3D view are shown.

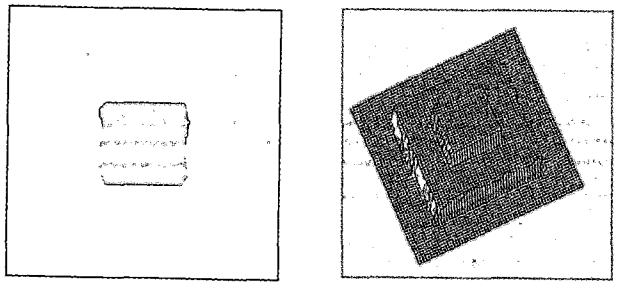


Fig. 9. Our result: a disparity map and its 3D plot.

in below. Our result is shown in Fig. 9.

Although the result contains errors, the boundaries are relatively sharp and occlusion area is relatively clean. To estimate the algorithm quantitatively, we computed the noise effect of our algorithm. This is possible since we know the exact solution, *i.e.*, the disparity map in Fig. 8. The amount of noise is controlled by the signal-to-noise ratio (SNR):

$$SNR \triangleq \frac{\sigma_s^2}{\sigma_n^2}, \tag{24}$$

where σ_s^2 and σ_n^2 denote respectively the signal variance and the amount of noise, in term of *i.i.d.* Gaussian. Denoting d^* and α respectively the disparity for the original image and the estimated disparity from the noisy image, we can calculate the root-mean-square-error (RMS):

$$\text{RMS} \triangleq \left[\frac{1}{N^2}\sum_{s\in S}(d_s^* - d_s)^2\right]^{\frac{1}{2}}. \tag{25}$$

Varying SNR for the RDS, we ran the program to obtain the disparity maps and calculated corresponding RMS. The graph in

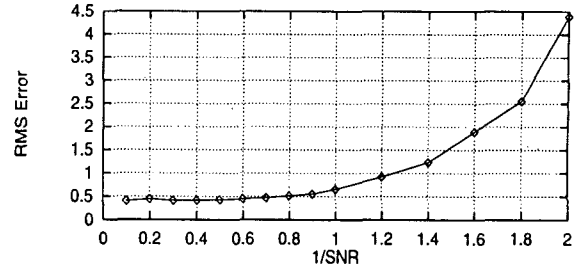


Fig. 10. RMS Error vs. 1/SNR for disparity computation.

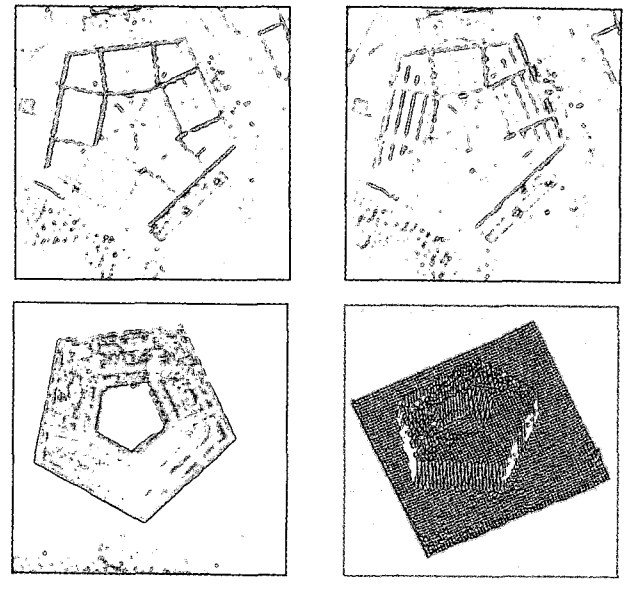


Fig. 11. Top: a 512×512 pentagon image pair and below: disparity map and 3D view.

Fig. 10 is the result.

As we can observe in this graph, the performance of this algorithm is gradually degraded, that is one of the desired characteristics for this type of algorithms. Further notice will reveal that the error is exponentially proportional to 1/SNR. However, the graph has some bias indicating that there exists an inherent errors that does not diminish as 1/SNR goes to zero. In fact, this errors are due to occluding regions between two images, that exist in the left sides of the rectangles and cannot be removed. Actually, these regions are filled with random noise in RDS so that the viewers cannot perceive the gaps.

2. *Experiment with Pentagon Image*

As a natural scene, the pentagon image pair in Fig. 11 is tested. The top images is a left and right image pair. Below them are shown the reference disparity map and its 3D rendering. As usual, the reference map have been obtained manually. For each pentagon image, Fig. 12 shows the phase images and singular maps at various scales (σ =16, 8, 4).

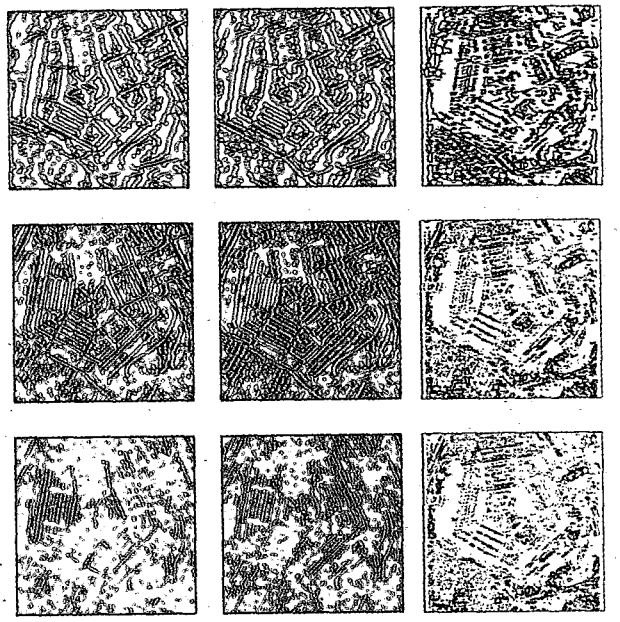


Fig. 12. The phase images and singular maps for the pentagon (σ=16, 8, and 4).

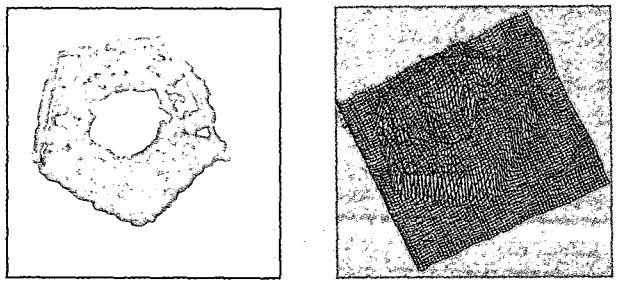


Fig. 13. Disparity map and its 3D view from our proposed algorithm.

Obtained disparity map with our algorithm is shown in Fig. 13. Compare to the reference disparity map, the boundaries and inside parts of the main building are not clear. However, the obtained disparity map shows better result outside of the main building.

3. Discussion

We have observed so far the results both for RDS and pentagon image. One of the result is that the new algorithm has the property of gradual degradation. There are four parameters in (23): α, λ, μ and γ. In this experiment, we determined all the values heuristically. The parameters might be derived from (17) by some Lagrangian approach. In this fashion, all the parameters are also determined automatically depending upon the image quality and types. However, this might ask too much computation.

VI. Conclusion

In this paper, we introduced a new multi-layer architecture for matching stereo images of RDS and general scene. As an optimal criteria for stereo matching, we derived a new energy function, based upon the MAP estimate, that contains natural constraints of disparity computation.

The result is a multi-layer relaxation algorithm with $O(N^2 \log N)$ multiplications. This algorithm is relatively fast, has low recognition errors, is robust to noise, and in particular can be used in general purpose stereo matching. One of the open problems will be the enhancement of computational complexity, the reduction of recognition errors, and the automatic updation of the internal parameters.

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*, Vol. 1, No. 2, pp. 107-131, 1987.

[2] N. Ayache and F. Lustman, "Fast and reliable trinocular stereovision," *Proc. Ist Conf. Computer Vision*, pp. 422-427, June, 1987.

[3] 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.

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

[5] P. J. Burt, *The Pyramid as a Structure for Efficient Computation*, Springer Verlag, 1984.

[6] V. Cantoni and S. Levialdi(editors), *Pyramidal systems for computer vision*, Springer Verlag, 1986.

[7] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.

[8] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am.*, A 4, pp. 2379-2393, 1987.

[9] D. J. Fleet, and A. D. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, Vol. 5, pp. 77-104, 1990.

[10] D. J. Fleet, A. D. Jepson, and M. Jenkin, "Phase-based disparity measurement," Comput. Vision, Graphics, Image Processing, Vol. 53, pp. 198-210, 1991.

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

[12] D. J. Fleet, "Stability of phase information," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-15, No. 12, pp. 1253-1268, 1993.

[13] W. T. Freeman and E. H. Adelson, "The degign and use of steer-able filters," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-13, pp. 891-906, 1991.

[14] D. Gabor, "Theory of communication," *J. IEE*, Vol. 93, pp. 429-457, 1946.

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

[16] 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.

[17] W. E. L. Grimson, "A computer implementation of a theory of human stereo vision," *Phil. Trans. Royal Soc. London*, Vol. B292, pp. 217-253, 1981.

[18] W. E. L. Grimson, *From Image to Surfaces: A Computational Study of the Human Early Visual System*. Cambridge, MA: M.I.T. Press, 1981.

[19] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-7, No. 1, pp. 17-34, Jan. 1985.

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

[21] A. Jepson and M. Jenkin, "The Fast Computation of Disparity from Phase Difference," *Proc. IEEE CVPR*, pp. 398-403, 1989.

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

[23] Y. C. Kim and J. K. Aggarwal, "Positioning 3D objects using stereo images," *IEEE J. Robotics and Automation*, Vol. RA-3, No. 4, pp. 361-373, Aug. 1987.

[24] S. Kirkpatrick C. Gelatt and M. Vecchi, "Optimization by Simulated Annealing," *Science*, Vol. 220, pp. 671-680, 1983.

[25] L. Langley, T. J. Atherton, R. G. Wilson and M. H. E. Larcombe, "Vertical and horizontal disparities from phase," *First European Conference on Computer Vision*, pp. 315-325. April 1990.

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

[27] 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.

[28] G. Medioni and r. Nevatia, "Segment based stereo matching," *Comput. Vision, Graphics, Image Processing*, Vol. 31. pp. 2-18, 1985.

[29] H. P. Moravec, "Towards automatic visual obstacle avoidance," *Proc. 5th Int. Joint Conf. Artificial Intelligence*, pp. 584, 1977.

[30] O. Nakayama and Y. Shirai, "Stereo matching for occlusion boundaries using normalization light intensity," *ACCV'95 Second Asian Conference on Computer Vision*, pp. II-321-325, 1995.

[31] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE*, Vol. 69, No. 5, pp. 529-541, May, 1981.

[32] S. B. Pollard, J. E. Mayhew, and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, Vol. 14. pp. 449-470, 1981.

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

[34] E. P. Simoncelli, W. T. Freeman, E. H. Adelson and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Info. Theory*, Vol. 38, pp. 587-607, 1992.

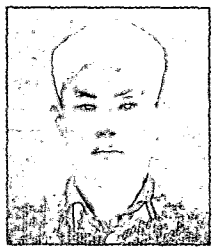
[35] A. M. Waxman, "An image flow paradigm," *Proc. Workshop on Computer Vision: Representation and Control*, pp. 49-57, 1984.

[36] J. Weng, "A theory of image matching," *Proc. 3rd ICCV*, pp. 200-209, 1990.

[37] J. Weng, "Image matching using the windowed Fourier phase," *International Journal of computer Vision*, Vol. 11, No. 3, pp. 211-236, 1993.

[38] C. J. Westelius, *Preattentive Gaze Control for Robot Vision, Thesis:* LIU-TEK-LIC-1992:14, Dept. of E. E. Linkoping Univ, Sweden.

Hong Jeong received the B.S. degree from the Department of Electrical Engineering at the Seoul National University, Korea, in 1977. In 1979, he received the M.S. degree from the Department of Electrical Engineering at the 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 from 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 members in the Department of Electrical Engineering at the Kyoung-Book National University, Daegu, Korea. He is a member of Sigma Xi. During 1994-1995, he worked as a vice-chairman in the Special Interest Group on Neurocomputing at 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 is now a Professor. His research interest include digital signal processing, computer vision, speech recognition, and radar signal processing.



Myoung-Sik Chae received the B.S. degree from the Department of Electrical Engineering at the Han-Yang University, in 1995, and the M.S. degree from the Department of Electrical Engineering at the Pohang University of Science and Technology (POSTECH), in 1997. He is a researcher in the SK Telecom. His current research interests include image processing, and image communication.

Jung-Gu Kim received the B.S. degree from the Department of Electrical Engineering at the Kyoung-Book National University, in 1991, and the M.S. degree from the Department of Electrical Engineering at the Pohang University of Science and Technology(POSTECH), in 1993. Since 1993, he has been working towards the Ph.D. degree in Electrical Engineering at POSTECH. He is a researcher in the Power Electronics Research Team of the Research Institute of Industrial Science and Technology(RIST). His current research interests include image processing, stereo vision, computer and machine vision and their application.