STEREO MATCHING ALGORITHM BASED ON MODIFIED WAVELET DECOMPOSITION PROCESS

YONG-SUK KIM, JUN-JAE LEE and YEONG-HO HA

1Department of Electronic Engineering, Kyungpook National University, Taegu 702-701, South Korea
2Department of Computer Engineering, Dongseo University, Pusan 671-716, South Korea

(Received 1 June 1995; in revised form 29 July 1995; received for publication 9 August 1996)

Abstract—Multiresolutional representation such as pyramidal structures is useful for stereo matching as coarse-to-fine strategy. However, conventional pyramidal structures using Gaussian or Laplacian filters lose much information due to their low-pass filtering characteristics and also cannot obtain any spatial orientation selectivity. The adoption of wavelet transform can remedy these problems, but at the time of image translation, it changes wavelet coefficients. In this paper, a pyramid using modified wavelet decomposition process is proposed to have translation invariance. The image transformed by the proposed method is converted into appropriate multiple features without loss of information. Since the importance of each feature is determined heuristically in the multiple feature-based stereo matching method, it is very difficult to fuse them adequately. In the proposed algorithm, the weight of each feature, that is, the relative importance of each feature, is decided from the similarity between the intensities in the local region of each left and right wavelet channels. Since the window size used for the decision of weight and disparity values greatly influences the processed result, the window is adaptively determined from the disparities estimated in the coarse resolution and low-varying channel of fine resolution. The window size must be large enough to obtain signal-to-noise ratio, but not too large as to induce the effects of projective distortion. Also, a new relaxation algorithm which can reduce false matches without blurring the disparity edge is proposed. By integrating adaptive weight, variable window selection method, and relaxation process, an accurate and stable disparity map is obtained. Experimental results for various images show that the proposed algorithm has good performance even if the image has the unfavorable conditions. © 1997 Pattern Recognition Society. Published by Elsevier Science Ltd.

Stereo matching Wavelet transform Adaptive weights Relaxation algorithm
Variable window selection method

1. INTRODUCTION

Depth perception in human is known to be derived from the 12 sources of information about distance.1 In computer vision system, binocular parallax information is considered effective to obtain depth by stereo matching, and various methods have been applied to solve the correspondence problem.2-6 Of the steps adopted for depth analysis, image matching is one of the most difficult problems in computer vision and depends on the choice of feature primitives. In stereo matching, the difficult problems that must be solved are: (1) difficulty in feature extraction; (2) fusion of extracted features; (3) determination of window size and shape; (4) consideration of large disparity; and (5) dealing with disparity edges and occluded regions in which some constraints cannot be satisfied. A new algorithm is presented in this paper to solve these problems.

The existing techniques for stereo matching are grouped into two categories according to the matching primitives. The feature-based methods which use sparse high-level features such as zero-crossing points of the filtered image,7-9 connected edges,10 segmented edges,11 and corner points12 have accurate disparity values at the feature points. However, these methods need complicated processes such as edge thinning and linking to avoid false matching and post-processing such as interpolation to obtain full resolution disparity map. Especially, they are inadequate when the texture of scenes is too dense or too sparse. The intensity-based methods which use dense low-level feature and intensity value itself have disparities in all pixels without feature extraction or interpolation,13-15 but require pre-processing for noise reduction because they are very sensitive to noise. When stereo images have large disparities, it is difficult to treat a false match. Therefore, multiple scenes,16 hierarchical structure,17 or neural network with various constraints18 are used to solve these problems. Other algorithms for stereo matching include segmented region-based method,19 phase-based method,20,21 topological method,22 tree matching,23 stereo matching using Gabor filters,24,25 and neural networks learned by constraints.26

Recently, some algorithms combining the intensity-based and the feature-based techniques have taken advantage of the reliable primitives of each technique. Weng et al.27 used some primitives simultaneously, i.e. intensity value, edgeness, and corneriness, to determine the correct disparity taking into account possible structural discontinuities and occlusions. Cochran and

...
Medioni\(^{(28)}\) first performed the intensity-based techniques which use the local variance of intensity pattern, then obtained accurate disparities using edge information as a feature-based primitive from the blurred disparity map. This method applies a set of constraints to identify and remove the low-confidence matches, then performs surface interpolation to obtain the full resolution disparity map. The performance of stereo vision system based on the above methods depends mainly on the extraction of the optimal features, high-level or low-level primitives, which is insensitive to image translation and noise, and optimal fusion of these features. However, it is a complicated and difficult process. On extracting features, the features can induce duplication or loss of information due to the extraction of each feature being processed separately. On fusing, relative importance of each feature is determined heuristically. This paper presents an approach to estimating the disparity of stereo image pairs in terms of features with low-varying and detail components at each channel which is obtained by wavelet transform.

The wavelet transform has been applied to many tasks such as multiresolution analysis,\(^{(29-32)}\) image compression,\(^{(33,34)}\) and singularity detection.\(^{(35,36)}\) The wavelet-transformed images are decomposed into blurred image (LL channel), horizontal edges (LH channel), vertical edges (HL channel), and diagonal edges (HH channel) without loss of information. These characteristics of the wavelet transform make the feature extraction process very simple and the pyramidal structure obtained by wavelet transform can be used for stereo matching as a coarse-to-fine strategy. However, the pyramidal structures using Gaussian or Laplacian filters lose much information and suffer from the difficulty that data at separate levels are correlated. Furthermore, Laplacian pyramid does not introduce any spatial orientation selectivity into the decomposition process. These characteristics result in inaccurate matching at each level of image pyramid, whereas the wavelet pyramid provides all features without duplication or loss of information. Nevertheless, wavelet transform is unstable with respect to the translation of the input signal because wavelet filters do not have ideal filter characteristics.

In this paper, a coarse-to-fine stereo matching algorithm based on the wavelet representation in terms of detail and low-varying components of images is proposed. The process of wavelet decomposition for preserving the characteristic of translation invariance, that is the simple translation of the transformed coefficients by the translation of input signal, is introduced. The matching in each level is achieved by minimizing weighted sum of mean absolute difference (WMSAD) considering the relative importance of each wavelet channel in the local window whose shape and size are decided adaptively. The relative importance of each wavelet channel is determined by normalized cross-correlation (NCC) which is increased in the channel of larger signal-to-noise ratio (SNR) value. This method easily obtains the fusion of multiple features consisting of low-varying and detail components used for stereo matching, while the fusion of each feature is complicated and difficult in the conventional methods. The disparities in each channel are initially estimated independently at the position obtained by minimizing MAD (mean absolute difference) between left and right channels. The weights of each channel are then obtained at each finally matched position. The initial weights are updated iteratively to determine the disparity per pixel by maximizing the matching possibility. The matching process is performed by the proposed relaxation algorithm. The new relaxation algorithm to improve the matched result obtains the disparity by using the matching possibility that is calculated from the locally adaptive weights and variable window in wavelet-transformed stereo images. The initial possibility values computed from the similarity in each wavelet channel are updated iteratively to impose global consistency. In each iteration, the possibility value is adjusted according to the difference in current possibility and neighboring values. The algorithm performs well without the blurring of disparity edge. In addition, to obtain precise and stable estimation of correspondence, it is important to select an appropriate window size, which depends on the signal and disparity variation within the window. However, while the signal variation is measurable, the disparity variation is not. In the proposed algorithm, it is possible to determine the window size appropriately since the disparity variation can be estimated from the disparity obtained in the coarse resolution.

The experimental results for random-dot stereograms, synthetic images, indoor and outdoor images are compared with those from other stereo matching algorithms. The results show that the proposed algorithm is effective for the various images, even if the images have too dense or sparse points, noise, and repeating patterns.

2. WAVELET PYRAMID FOR STEREO MATCHING

2.1. Wavelet transform

The wavelet transform is of interest for the analysis of non-stationary signals because it provides an alternative to the classical short-time Fourier transform or Gabor transform.\(^{(37)}\) In the discrete-time case, discrete wavelet transform can be applied to many tasks. By the wavelet transform, signal is decomposed into a set of basis functions called wavelets that are obtained from a prototype wavelet \(\psi(x)\) by dilations, contractions, and shifts.\(^{(38-40)}\) The corresponding wavelets are

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right),
\]

resulting in continuous wavelet transform

\[
\text{CWT}_f(a,b) = \langle \psi_{a,b}(t), f(t) \rangle,
\]

where \(\psi_{a,b}(t)\) is a wavelet with scale factor \(a\) and shift factor \(b\). The wavelet transform uses windows with wide bandwidth at high frequencies and narrow bandwidth at low frequencies according to the constant-Q concept. At resolution \(2^l\), \(f^l[n]\) [the discrete approximation of \(f(x)\)]
and $d'[n]$ [the discrete detail signal of $f(x)$] can be represented as the function of scaling function $\varphi(x)$ and wavelet $\psi(x)$. For the sequence,

$$f^{(0)}[n] = \langle \varphi(t-n), f(t) \rangle, \quad n \in \mathbb{Z},$$  \hspace{1cm} (3a)

$$f(t) = \sum_{n=-\infty}^{\infty} f^{(0)}[n] \varphi(t-n),$$  \hspace{1cm} (3b)

considering first the projection onto $V_0$,

$$f^i[n] = \left< \frac{1}{\sqrt{2}} \varphi(t/n), f^{i-1}(t) \right>.$$  \hspace{1cm} (4)

Substituting $\varphi(t) = \sqrt{2} \sum_k g_0[k] \varphi(2t-k)$, we have

$$f^i[n] = \sum_k g_0[k - 2n] f^{i-1}[k] = \sum_k g_0[2n-k] f^{i-1}[k],$$  \hspace{1cm} (5)

where $g_0[n] = g_0[-n]$. That is, the coefficients of the projection onto $V_1$ are obtained by filtering $f^{(0)}$ with $g_0$ and downsampled by 2. The coefficient of the projection onto $W_1$ is

$$d'[n] = \left< \frac{1}{\sqrt{2}} \psi(\frac{t}{2^n}), f^{i-1}(t) \right>.$$  \hspace{1cm} (6)

Also inserting $\psi(t) = \sqrt{2} \sum_l g_1[l] \varphi(2t-l)$, we have

$$d'[n] = \sum_l g_1[l - 2n] f^{i-1}[l] = \sum_l g_1[2n-l] f^{i-1}[l],$$  \hspace{1cm} (7)

where $g_1[n] = g_1[-n]$. That is, the coefficients of the projection onto $W_1$ are obtained by filtering $f^{(0)}$ with $g_1$ and downsampled by 2. In this paper, orthogonal wavelet is used based on Mallat’s algorithm. The discrete wavelet transform of the 2-D signal can be written by

$$A^d_2 f = ( (f(x,y) * \varphi_2(-x) \varphi_2(-y))(2^{-j}n,2^{-j}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (8a)

$$D^1_2 f = ( (f(x,y) * \varphi_d(-x) \psi_d(-y))(2^{-j}n,2^{-j}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (8b)

$$D^2_2 f = ( (f(x,y) * \psi_d(-x) \varphi_d(-y))(2^{-j}n,2^{-j}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (8c)

and is shown in Fig. 1 by the block diagram that is similar to subband decomposition.

### 2.2. Wavelet pyramidal structure for preserving the translation invariance in stereo image

When the stereo images are filtered and sampled to construct the pyramidal structure, the convolved output must retain translation-invariance, i.e. the translation of input signal must produce simple translation of the transformed coefficients. This characteristic is defined as shiftability and has been addressed by several researchers. \(^{(41)}\) In the frequency domain, the shiftability of sampled signal does not correspond to aliasing. If the discrete signal is resampled by the sampling rate which cannot satisfy Nyquist rate, the sampled signal becomes unstable by aliasing effect. The wavelet transform is unstable with respect to the translation of the input signal because filters have no ideal filter characteristics. Figure 2 illustrates the aliasing effect of decomposed signal by filtering and decimation as shown in Fig. 1. It is, however, possible to achieve a weak form of translation-invariance if the process of wavelet decomposition is modified. The occurrence of aliasing in Fig. 2(d), the signal filtered and decimated by scaling function as shown in Fig. 2(b), is responsible for instability, and cannot be avoided so far as the compactly supported wavelet which has the finite number of filter tabs is used. In this paper, the signal before decimation [Fig. 2(c)] makes use of the matching. Then the signal is decimated to construct a coarse-resolution signal. The shiftability can be preserved in the signal as shown in Fig. 2(c) which has no aliasing. Figure 3 shows the construction process of channels for the proposed stereo matching from finer resolution $(A^d_2,f)$. Wavelet channels obtained by the proposed decomposition method are themselves feature domains consisting of low-varying $(A^d_2,f)$ and detail $(D^1_2 f, D^2_2 f, D^3_2 f)$ features without pre-processing and post-processing:

$$A^d_2 f = ( (f(x,y) * \varphi_d(-x) \varphi_d(-y))(2^{-j+1}n,2^{-j+1}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (9a)

$$D^1_2 f = ( (f(x,y) * \varphi_d(-x) \psi_d(-y))(2^{-j+1}n,2^{-j+1}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (9b)

$$D^2_2 f = ( (f(x,y) * \psi_d(-x) \varphi_d(-y))(2^{-j+1}n,2^{-j+1}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (9c)

$$D^3_2 f = ( (f(x,y) * \psi_d(-x) \psi_d(-y))(2^{-j+1}n,2^{-j+1}m) )_{(n,m) \in \mathbb{Z}^2},$$  \hspace{1cm} (9d)

and is shown in Fig. 1 by the block diagram that is similar to subband decomposition.
Fig. 2. The shiftability in the frequency domain: (a) frequency domain of original signal; (b) Fourier transform of scaling function $\varphi(\omega)$; (c) filtered signal; (d) aliasing effect of downsampled signal.

Fig. 3. The construction of the channels for the proposed stereo matching from $A^{d,j}_{2j+1} f$, where $A_0^d f$, $D_0^2 f$, $D_0^3 f$, and $D_2^3 f$ channels represent the blurred image (LL), the horizontal edges (LH), the vertical edges (HL), and the diagonal edges (HH), respectively.
They are insensitive to image translation and noises and are extracted easily without duplication or loss of information. In fact, feature extraction is a very difficult and complicated process accompanying duplication or loss of information in the conventional hybrid type algorithm which uses low-level and high-level features in stereo matching. Modified wavelet transform as shown in Fig. 3 using Mallat’s wavelet \(31\) constructs the pyramidal structure for matching as shown in Fig. 4. In the proposed algorithm, the signal before decimation is used to satisfy the shiftability of stereo images. These facts are ascertained in the experimental result for a randomly generated 1-D signal. The result is shown in Fig. 5. To confirm the effect of decimation, 1-D signal (128 samples) used in this experiment does not have noises, and the disparity value is determined as the minimum point of mean absolute difference (MAD) without any additional algorithm. Besides, the rectangular window used to calculate MAD considers five samples. Comparing the 32 sample-length signals obtained by conventional wavelet transform [Fig. 5(c) and (d)] with those obtained by modified process [Fig. 5(e) and (f)], the latter is better than the former. Moreover, the disparities obtained by using the same algorithm for the different signals evidently show the advantage of modified process.

In coarse-to-fine strategy, the disparity value (refer to Section A.2) in the current level is initialized by doubling of twice the disparity value obtained in the coarser level. The initial values of the coarsest level are assumed to be zero. Using the pyramidal structure, the searching space in each level is reduced greatly. In the case that the number of levels is \(l\) and the searching space of each level is between \(-s\) and \(s\), the number of the entire image is \(2(2^l - 1)s + 1\). False match and processing time are reduced because maximum searching space can be reduced. In fact, the computational load in the \(N \times N\) size image is \([2(2^l - 1)s + 1]N^2\), but that in the
Fig. 5. (a) Left 1-D signal (128 samples). (b) Right 1-D signal (128 samples). (c) Discrete approximations at the resolution $\frac{1}{2}$ of left signal obtained by conventional wavelet transform (32 samples). (d) Discrete approximations at the resolution $\frac{1}{2}$ of left signal obtained by modified process (32 samples). (e) Discrete approximations at the resolution $\frac{1}{2}$ of left signal obtained by modified process (32 samples). (f) Discrete approximations at the resolution $\frac{1}{2}$ of right signal obtained by modified process (32 samples). (g) Ideal disparity. (h) Disparity obtained for the signal (c) and (d). (i) Disparity obtained for the signal (c) and (f).
Fig. 5. (Continued).
pyramidal structure is
\[ C = (2s + 1)N^2 \left[ 1 + \frac{1}{2^s} + \frac{1}{2^s + 1} \right] \]
\[ = (2s + 1)N^2 \left( \frac{1 - 1/2^s}{1 - 1/2^s} \right) = \left( \frac{4\cdot2^s - 1}{3 \cdot 2^s} \right) (2s + 1)N^2. \] (17)

For example,
\[ C = 10.37\% \text{ of } [2(2l - 1)s + 1]N^2 \text{ for } l = 4, \ s = 1. \] (18)

Although all wavelet channels are considered, the proposed algorithm has less computational load than half the entire one.

3. A STEREO MATCHING ALGORITHM

3.1. The cost function for similarity measure

The cost function for similarity measure, \( E_p \), is defined \( p \)-Hölder norm as follows:

\[ E_p = \sqrt[|p|]{\sum_{j=1}^{N} \sum_{i=1}^{N} |x_{ij} - y_{ij}|^p}, \quad p \geq 1. \] (19)

For \( p = 1 \),
\[ E_1 = \sum_{j=1}^{N} \sum_{i=1}^{N} |x_{ij} - y_{ij}|. \] (20)

Minimizing \( E_1 \) creates some Dirac peaks. On the other hand, for \( p = \infty \)
\[ E_{\infty} = \lim_{p \to \infty} \left( \sqrt[|p|]{\sum_{j=1}^{N} \sum_{i=1}^{N} |x_{ij} - y_{ij}|^p} \right) = \sup |x_{ij} - y_{ij}|. \] (21)

Minimizing \( E_{\infty} \) makes uniform the differences at each point. If the SAD (sum of absolute difference) is selected as the similarity measure for the same SSD (sum of squared difference) value, the difference which includes a Dirac peak reduces to the minimum value as shown in Fig. 6. Taking the example of a two-pixel

![Fig. 6. The difference distribution of the size \( n \) window with (a) uniform distribution and (b) Dirac peak.](image-url)
Stereo matching algorithm 937

problem that the window size is 1 × 2 or 2 × 1, SAD and SSD are

\[ E_1 = |D_{x_1}| + |D_{x_2}|, \]

\[ E_2 = (|D_{x_1}|^2 + |D_{x_2}|^2)^{1/2}, \]

where \( D_{x_1} \) and \( D_{x_2} \) are the differences between the left and right image at each position \( X_1 \) and \( X_2 \), respectively. If SAD is used as the similarity measure, the locus of the \((D_{x_1}, D_{x_2})\) plane with constant difference value, \( D \) must satisfy the following condition:

\[ D = |D_{x_1}| + |D_{x_2}|. \]

Therefore,

\[ D_{x_1} = \begin{cases} -D_{x_1} + D & (D_{x_1} \geq 0, D_{x_2} \geq 0), \\ D_{x_1} - D & (D_{x_1} \geq 0, D_{x_2} < 0), \\ -D_{x_1} + D & (D_{x_1} < 0, D_{x_2} \geq 0), \\ -D_{x_1} - D & (D_{x_1} < 0, D_{x_2} < 0). \end{cases} \]

On the other hand, to consider the constant SSD value, \( D \) is given by

\[ D = (|D_{x_1}|^2 + |D_{x_2}|^2)^{1/2}, \]

that it is a circle with a radius, \( D \).

Considering two cases in the \( n \)-pixel problem (\( n \) pixels in the window):

(i) all difference values are uniformly distributed in the window:

\[ D_{x_1} = \frac{1}{\sqrt{n}}, \quad D_{x_2} = \frac{1}{\sqrt{n}}, \ldots, \quad D_{x_n} = \frac{1}{\sqrt{n}}, \]

(ii) one difference value is particularly large in contrast to the others

\[ D_{x_i} = \begin{cases} 1 & \text{if } i = m, \\ 0 & \text{otherwise}, \end{cases} \]

where \( i = 1, \ldots, n, \quad 1 < m < n. \)

\[ E_1 = 1 \quad E_2 = 1 \quad E_1 = \sqrt{2} \]

\[ D_{x_1} \]

\[ D_{x_2} \]

\[ E_1 \]

\[ E_2 \]

\[ E_1 = \sqrt{2} \]

In both (i) and (ii), SD value \( E_2 \) equals 1. If SSD value is used as the similarity measure, only the minimum SSD point can be selected. However, SAD values \( E_1 \) are \( \sqrt{n} \) and 1 in the cases (i) and (ii), respectively. Therefore, case (ii) is selected as the minimum SAD point.

In the stereo matching problem, Fig. 6(b) shows a more accurately matched case than Fig. 6(a) because the error of unique pixel can be considered as noise. Considering the results so far achieved, it seems that the SAD is a favorable similarity measure for the matching of images with white noise (see Fig. 7).

In the proposed pyramidal structure, WSMAD calculated in the wavelet channels is used to decide the disparities for each layer:

\[ E'_i = \frac{1}{m \times n} \sum_{m} \sum_{n} \omega_{ij}^p |A_i^p(i, j) - A_i^p(i, j + d_{ij})| + \omega_{ij}^p |D_i^p(i, j) - D_i^p(i, j + d_{ij})| + \omega_{ij}^p |O_i^p(i, j) - O_i^p(i, j + d_{ij})|, \]

where \( \omega_{ij}^p, \omega_{ij}^p, \omega_{ij}^p, \) and \( \omega_{ij}^p \) are weights for each wavelet channel in \((i, j)\) and \( m \times n \) denotes the size of window. The disparities are determined as the global energy of WSMAD moves to the minimum state by relaxation. Detailed process for determination of disparities will be explained in Section 3.4.

3.2. The locally adaptive weights for each channel

In the hybrid method using multiple features such as intensity value and various feature information, the fusion of each feature is complicated and difficult. As previously stated, the wavelet channels make up 2-D filter banks. So stereo images are decomposed into four channels that have different characteristics. The SNR of the specific wavelet channel can be improved by bandpass filtering such as wavelet transform. The correlation of left and right channels is increased in the channel that has large SNR value, because such a channel is decorrelated less by noise. On the assumption that the images are decorrelated by Gaussian white noise, the intensity values of left and right images are

\[ I_l(x, y) = I(x, y) + n_l(x, y), \]

\[ I_r(x, y) = I(x + d(x, y), y) + n_r(x, y), \]

\[ I_l(x, y) - I_r(x + d(x, y), y) = I(x, y) + n_l(x, y) - I(x + d(x, y)) - d(x, y), \]

\[ n(x, y) = n_l(x, y), \]

where \( n(x, y) \) is Gaussian white noise such that

\[ n(x, y) \sim N(0, \sigma_n^2). \]

Here, \( N(a, b) \) denotes a normal distribution with mean \( a \) and variance \( b \). In this case, the SNR is increased by the signal power and the disparity estimated in the wavelet channel that has a large SNR value is more reliable.
Therefore, the matching performance can be improved by assigning proper weight to each wavelet channel. To consider the relative importance of each wavelet channel, the weights are calculated based on the similarity between features (value themselves of each wavelet channel) obtained in the stereo image by wavelet transform. First of all, the disparities in each channel, $\hat{d}_{ij}$, are initially estimated independently as the position minimizing MAD between left and right channels, then the weights are obtained by NCC value in each pre-matched position resulting from $\hat{d}_{ij}$:

$$w_q = \frac{M[\sigma_q(i,j)\sigma_r(i,j + \hat{d}_{ij})] - M[\sigma_q(i,j)]M[\sigma_r(i,j + \hat{d}_{ij})]}{\sigma[\sigma_q(i,j)]\sigma[\sigma_r(i,j + \hat{d}_{ij})]}$$

where $M[\cdot]$ denotes the average value, $\sigma[\cdot]$ the standard deviation, and $q(i,j)$ the value of each wavelet channel, $A_q^{h}$, $D_q^{v}$, $D_q^{h}$, and $D_q^{v}$. The weights calculated initially are used in equation (34) to determine the disparity, $d_{ij}$, per pixel. The weights then are updated iteratively by equation (39). In this process, the disparities, $\hat{d}_{ij}$, have same values of $d_{ij}$ in a pixel (the weights were stabilized using 1–5 iterations). If weights for high-pass channels are larger than for low-pass channel, high-level features such as edges or corners become effective. Otherwise, low-level features such as smoothed intensity values become effective. Therefore, low-varying ($A_q^{v}$) and detail features ($D_q^{v}$, $D_q^{h}$, and $D_q^{v}$) are applied adaptively in each pixel according to the weights of each channel. In the experiments, the results with fixed weights and their iteratively updated ones are compared with each other. In methods with fixed weights, the disparities are determined, without relaxation process, as follows:

$$d_{ij} = \arg\min(E_i).$$

### 3.3. The locally adaptive window

In stereo matching, it is important to select an appropriate window size to obtain both precise and stable estimation of correspondences. The window size must be selected by the signal and disparity variances within the window. A large window tends to increase the signal variance, but at the same time it tends to include points of different disparity values. The variance of the signal must be large enough, and the variance of the disparity must be small enough. However, while the signal variation is measurable, disparity variation is not. If the disparities within the selected window are not the same, a false match will occur as shown in Fig. 8. In the case of matching with $3 \times 3$ size window, the real disparity value is “1” at the corner point of disparity pattern, while with the number of matched pixels in the pre-defined window enclosed by the bold line in Fig. 8(b), it is “0” (false match). The error occurs at a position near disparity edge. Therefore, a locally adaptive window is needed to reduce the false matches.

In the algorithm, it is possible to determine the window size appropriately since disparity variation can be estimated from the disparity obtained in the coarse resolution. In this paper, the estimated disparities in the coarse level and the variation of LL channel of wavelet-transformed image are used to decide the variable window shape adaptively. At first, the homogenous and inhomogeneous regions for the disparity estimated in the coarser resolution layer are distinguished. If the estimated disparities in the $5 \times 5$ window are the same, then a pre-defined maximum size window satisfying the homogeneity condition is selected. Otherwise, nine types of windows as shown in Fig. 9 are considered and the type in which MAD of the LL channel has minimum value is selected. Then the size of window is determined while the type of window is retained. The proposed method can estimate an accurate disparity even if the sudden change of disparity in the image exists.

For example, the occlusion exists in the 1-D profile of synthetic image (64th row of randomly generated grayscale image) as shown in Fig. 10. The false matches occurred in disparity edges, i.e. in the occluded regions. Nevertheless, the accurate disparity values can be obtained in the occluded regions using the proposed method.

### 3.4. The relaxation algorithm

In order to improve matching performance, Marr and Poggio, Barnard and Thompson, Kim and Aggarwal, Nasrabadi and Choo have used the cooperative processing to consider the constraints in stereo matching. Marr and Poggio used the neighborhood information of matchable primitives in a simple iterative

![Fig. 8. The occurrence of a false match. In this example, the real disparity value of corner point is “1” (bold italic) and the disparity value estimated by fixed rectangular window (3 x 3) is “0”. Therefore, the locally adaptive window is needed. (a) The pattern of left image. (b) The pattern of right image. (c) The ideal disparity map.](image-url)
Fig. 9. The types of proposed initial window (5 x 5). Here ‘+’ is the center pixel to determine the matching possibility.

Fig. 10. (a) 64th row profile of the left image. (b) 64th row profile of the right image. (c) 64th row profile of the right image with noise. (d) Ideal disparity. (e) The matched disparity value using the fixed size window. (f) The matched disparity value using the proposed adaptive window.
scheme. Barnard and Thompson updated iteratively the probabilities computed from the similarity of feature points obtained by the Moravec operator from each image. Kim and Aggarwal proposed a relaxation scheme that combines continuity of disparity, figural continuity, and smoothness of matching possibility. Nasrabadi and Choo used the Hopfield network for relaxation process. In this paper, a new relaxation algorithm is proposed in order to consider the disparity value of neighboring pixel. The disparity edge blurring can be avoided because only the neighboring node which has the same disparity value affects the new state value of selected node, and the normalization process which is required in the conventional relaxation algorithms can be omitted because the sigmoid function used in the proposed algorithm has the value between 0 and 1.

3.4.1. The initial matching possibility value. Matching possibilities $p_{ij,d}$ are initially calculated for the possible disparities at all pixels from $E_1$ value as equation (29):

$$p_{ij,d} = \frac{1}{1 + \exp[-\lambda E_1 / M[E_1] - 1]}$$

with

$$M[E_1] = \frac{1}{2y \times 2y \times D} \sum E_1'$$

Fig. 10. (Continued).
Here $M[E'_i]$ and $D$ denote the average of all WSMAD values ($E'_i$) and the number of possible disparity in each layer, respectively. The slope of sigmoid function expressed in equation (40) is determined by $\lambda$. As the $\lambda$ value increases, the converted possibility value shows the wide distribution shown in Fig. 11. However, when the $\lambda$ value is too large, the possibility value becomes a binary value divided by the threshold $M[E'_i]$. To determine $\lambda$ in equation (40), it is assumed that the probability distribution function of $E'_i$ is a Gaussian distribution with average $M[E'_i]$ and standard deviation $\sigma(E'_i)$. Under this assumption, to set the possibility value to 0.1 when the value of cumulative distribution function is 0.9, $\lambda$ value is calculated by (see Section A.1)

$$\lambda = \ln 9 \times \frac{M[E'_i]}{1.282 \times \sigma(E'_i)}.$$  

(38)

The initial value of each node is set with the calculated possibility value. Figure 12 is the network for the proposed relaxation algorithm and Fig. 13 is the connection state of each node.

### 3.4.2. The update of the possibility value at each node.

If the determination of disparity in the pixel which includes randomly selected node is ambiguous, then the state value of the node, $p_{i,j,d}$, with position $(i,j)$ and disparity, $d$ is updated. That is, the new value of the node, $p_{i,j,d}^{t+1}$, is

$$p_{i,j,d}^{t+1} = \begin{cases} f(u + \Delta u) & \text{if } \max(p_{i,j,d}) - \min(p_{i,j,d}) < T_p, \\ p_{i,j,d}^t & \text{otherwise}, \end{cases}$$

(39)

where $T_p$ is the threshold value of the possibility difference at the same pixel position to update the value:

$$f(u) = \frac{1}{1 + \exp(-\lambda(u - 0.5))}, \quad \lambda = 4.$$  

(40)

Here $u$ in equation (39) is determined from equation (40) as follows:

$$u = 0.5 - 0.25 \ln \frac{1 - p_{i,j,d}^t}{p_{i,j,d}^t}.$$  

(41)

The difference of possibility is

$$\Delta u = p - p_{i,j,d}^t.$$  

(42)

The average value of the neighboring node, $\bar{p}$, is

$$\bar{p} = \frac{1}{K} \sum_r \sum_s \sum_k p_{r,s,k},$$

if $k = d$, $(r,s) \neq (i,j)$, $(r,s) \in N(i,j)$,  

(43)

where $K$ is the number of neighboring node that has the same disparity as that of center node and $N(i,j)$ is the neighboring node of $(i,j)$ position. The $p_{i,j,d}^{t+1}$ is changed by $\Delta p$ as shown in Fig. 14. The $\lambda$ of equation (40) was determined such that sigmoid function has unit slope when input is 0.5. The value $\lambda$ is given by 4 to satisfy

$$\frac{df(u)}{du} \bigg|_{u=0.5} = \frac{1}{4} \lambda = 1.$$  

(44)

(45)
If probabilities reach the steady state or predetermined number of iterations, the iteration process will be stopped. Then, the disparity of each pixel is decided by

\[ d_{ij} = \text{arg max}(p_{ij,d}). \]  

(46)

The final process can be implemented by MAXNET.\(^{(46,47)}\) Actually, most nodes reached the stable state within 10 epochs. In the experiments, an 8-neighbor node is considered and \( T_p \) is 0.6. The proposed relaxation algorithm reduces the false matches without blurring the disparity edge and nodes are stabilized quickly.

3.5. The overview of the proposed stereo matching algorithm

The proposed algorithm calculates the matching possibility from the WSMAD with a locally adaptive window and weights of each wavelet subband. Then, it decides the disparity values using the new relaxation algorithm by coarse-to-fine strategy. The algorithm is illustrated in Fig. 15.

(i) To begin with, convert the left and right images pyramidal structure by the discrete wavelet transform using Mallat’s wavelet.

(ii) Then, determine the shape and size of the adaptive window in the coarsest resolution.

(iii) By the determined adaptive window, calculate the weights of each wavelet channel for each pixel.

(iv) Using the window and weights, obtain the matching possibility of each disparity.

(v) Conduct the relaxation process by using the matching possibility as initial value.

(vi) Update the previous weights by newly determined disparity.

(vii) If the change of weights does not occur, increase the resolution of images.

(viii) Repeat the previous process until the resolution reaches the original sized imaged with full resolution.

(ix) The distance is calculated from a full resolution disparity map by stereo geometry.

4. RECOVERING DEPTH

Stereopsis geometry has two types: parallel axis and non-parallel axis stereo geometry. In this paper, the general derivation to calculate depth is shown in Section A.2. The result is

\[ x = \frac{4D^2L^2f^2 + 2DL^2(\text{dl} \text{ + dr}) + L^3\text{dldr}}{4DLf^2 + (L^2f - 4D^2f)(\text{dl} \text{ + dr}) - 4DLdldr}, \]  

(47)

where \( D \) is the depth when the disparity is 0, \( L \) the stereo baseline and \( f \) is the focal length of camera. \( \text{dl} \) and \( \text{dr} \) are distances from the center point of left and right images, respectively. In the parallel axis geometry, \( x \) is simplified as

\[ x = -\frac{L \times f}{\text{dl} + \text{dr}}. \]  

(48)

Fig. 15. The proposed stereo matching algorithm.
5. EXPERIMENTS AND DISCUSSION

Various experiments for RDS (random-dot stereogram), synthetic images, indoor and outdoor images, were performed to prove that the proposed algorithm is superior. Three experiments for the proposed algorithm are simulated. They are:

1. pyramidal stereo matching without weights in the wavelet-transformed domain (SMW-FW),
2. the iterative method with locally adaptive weights (SMW-VW),
3. the method with variable window, locally adaptive weights and relaxation process (SMW-RN).

5.1. RDS and synthetic images

The images used for the experiments are 5% RDS without noise, 30% RDS of 10% decorrelated dots, 50% RDS of decorrelated 20% dots, uniformly distributed random gray-dot image with the Gaussian white noise (standard deviation of the noise is 50), and "stripes" image of decorrelated 20% dots as shown in Figs 15-19. The RDS images are composed of white dots (gray level is 255) and black dots (gray level is 0); 5% RDS which has few features is apt to induce the mismatch. The mismatched region is increased by using small window, whereas the blurring effect occurs in the disparity edge by using a large window. In 30% RDS, the randomly selected 10% dots of the right image are decorrelated. That is, if the selected pixel is a white dot, the pixel is converted in to a black dot, and vice versa; 50% RDS shows the image with severe noise and 20% dots or right image are decorrelated. The gray-dot image with Gaussian noise has uniform distribution between 0 and 255 and the variance of the noise is 2550. The image intensity distribution is similar to that of the real image. "Stripes" image has a repeated pattern. The same patterns easily generate the mismatching due to noise and large searching space. The searching space is from 0 to 15 pixels in the experiments and the "stripes" image has much noise. Under this condition, the image always has the possibility of duplicated match in all pixels. Therefore, using the conventional stereo matching algorithm, the accurate result cannot be obtained. The disparity of all RDSs and the synthetic images is a wedding-cake type with four layers. To evaluate the performance of the proposed algorithms, SAE (sum of absolute error) and MSE (mean squared error) are used:

\[
\text{SAE: } s_1 = \sum_{i} \sum_{j} |d_{ij} - \tilde{d}_{ij}|
\]

\[
\text{MSE: } s_2 = \frac{1}{N^2} \sum_{i} \sum_{j} (d_{ij} - \tilde{d}_{ij})^2
\]

where \( N \times N \) is the size of the image. The searching space of all algorithms is set from 0 to 15 pixels. In the proposed hierarchical algorithm, the number of total layers is four and the searching space in each layer is from 0 to 1 pixel. The maximum size of the locally adaptive window is 7 x 7. The proposed algorithms were compared with conventional intensity-based stereo
Table 1. Results of the proposed stereo matching algorithm (SWM-RN) for RDS and synthetic images

<table>
<thead>
<tr>
<th>Images</th>
<th>Noise</th>
<th>MSE</th>
<th>SAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5% RDS</td>
<td>None</td>
<td>0.045</td>
<td>730</td>
</tr>
<tr>
<td>30% RDS</td>
<td>10% Decorrelated</td>
<td>0.031</td>
<td>511</td>
</tr>
<tr>
<td>50% RDS</td>
<td>20% Decorrelated</td>
<td>0.039</td>
<td>643</td>
</tr>
<tr>
<td>Gray-dot</td>
<td>Gaussian (σ=50)</td>
<td>0.026</td>
<td>421</td>
</tr>
<tr>
<td>Stripes</td>
<td>20% Decorrelated</td>
<td>0.055</td>
<td>907</td>
</tr>
</tbody>
</table>

Fig. 20. "Stripes" image (128 × 128) (20% dots of right image are decorrelated). (a) Left image. (b) Right image.

Fig. 21. The experimental results for RDS and synthetic images. Various methods are compared by SAE. The methods used in these experiments are SAD, SSD, NCC, SMW-FW, SMW-VW, and SMW-RN. SAD, SSD, and NCC are classical intensity-based methods. On the other hand, SMW-FW, SMW-VW, and SMW-RN are the proposed stereo matching algorithm. (a) 5% RDS as shown in Fig. 16. (b) 30% RDS as shown in Fig. 17. (c) 50% RDS as shown in Fig. 18. (d) Gray-dot image as shown in Fig. 19. (e) "Stripes" image as shown in Fig. 20.
Fig. 22. The disparity map for RDSs and synthetic images. The results are obtained using the proposed algorithm (SMW-RN). (a) 5% RDS as shown in Fig. 16. (b) 30% RDS as shown in Fig. 17. (c) 50% RDS as shown in Fig. 18. (d) Gray-dot image as shown in Fig. 19.
matching algorithms that use the SAD, SSD, and NCC as similarity measures. The experimental results are shown in Figs 20–22 and Table 1. According to the results, the proposed algorithms exhibit good performance even for the images of unfavorable conditions, for example, when the images have too many or too few feature points, severe noise, and repeating patterns. Especially, the false match is reduced greatly even without post-processing by the method using relaxation.

5.2. Real images

The indoor image ("bear") and outdoor image ("pentagon"), as shown in Figs 23 and 24, were used for the test. The "bear" image pair was acquired in the laboratory by moving one camera. In this image, the toy, "bear" is posed in front of the "ball" on the bar. The searching space for the "bear" image and "pentagon" image is from −15 to 15 and from −31 to 31, respectively. The results of the experiment which uses the

Fig. 23. The experimental results for "stripes" image as shown in Fig. 20. They are obtained by various methods. Each result shows the best one obtained for each method. (a) NCC. The result is obtained by classical intensity-based method. (b) SMW-FW. The window size is 11 × 11, in which the result has best performance. (c) SMW-VW. The window size is 9 × 9. (d) SMW-RN. Variable window is used and the maximum window size is 7 × 7.
Fig. 23. (Continued).

Fig. 24. "Bear" image (200 x 200). (a) Left image. (b) Right image.
(a) (b)

Fig. 25. "Pentagon" image (512 x 512). (a) Left image. (b) Right image.

Fig. 26. The experimental result for "bear" image. The proposed algorithm (SMW-RN) is used to obtain the result.

Fig. 27. The experimental result for "pentagon" image. The proposed algorithm (SMW-RN) is used to obtain the result.
relaxation process (SMW-RN) are shown in Figs 25–27. The results show that the proposed algorithm is well applied to the real images without blurring the disparity edge.

6. CONCLUSIONS

This paper describes a new stereo matching algorithm using pyramidal structure obtained by modified wavelet transform. By the proposed method, stereo images are converted into multiple directional features without loss of information. The features are quite adequate to stereo matching because they have the characteristics of translation invariance. The proposed stereo matching method includes new algorithms to determine locally adaptive weights and variable windows. The weights decide the relative importance of wavelet channels composed of blurred image, horizontal, vertical, and diagonal edge. False matches in the disparity edges are reduced by the variable window. Moreover, a new relaxation algorithm and the characteristics of similarity measures are discussed in this paper. The relaxation algorithm reduces noise effect without the blurring of disparity edges. Matching possibility calculated by the locally adaptive weights and variable window in wavelet-transformed stereo images is used for the relaxation process. Experimental results for random-dot stereograms, synthetic images, indoor and outdoor images show that the proposed algorithm shows good performance for various images even if the images have severe noise, repeated pattern, and a too sparse or dense texture. Further study for the hardware implementation of the proposed algorithm is required.

APPENDIX A. DETERMINATION OF $\lambda$ VALUE AND RECOVERING DEPTH

A.1. Determination of $\lambda$ value

If $f(x)$ has Gaussian distribution

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right), \quad \sigma > 0,$$

(A.1)

the cumulative distribution function (cdf) $\Phi$ is

$$\Phi\left(\frac{c - \mu}{\sigma}\right) = p(X \leq c) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{c} \exp\left(-\frac{1}{2} \left(\frac{\nu - \mu}{\sigma}\right)^2\right) d\nu,$$

(A.2)

where $\mu$ is the average and $\sigma$ the standard deviation.

When the cdf value becomes 0.9 as shown in Fig. 28, value $c$ has to satisfy

$$p(X > c) = 1 - p(X \leq c) = 0.1$$

(A.3)

and

$$\Phi\left(\frac{c - \mu}{\sigma}\right) = 0.9.$$ (A.4)

Therefore [see reference (48)],

$$\frac{c - \mu}{\sigma} = 1.282,$$ (A.5)

$$c = 1.282\sigma + \mu.$$ (A.6)

From equation (36)

$$\left\{1 + \exp\left(\frac{\lambda}{1.282\sigma} - 1\right)\right\} = 0.1.$$ (A.7)

$$\lambda = \ln 9 \times \frac{\mu}{1.282\sigma} = 1.713903726 \times \frac{\mu}{\sigma}.$$ (A.8)

A.2. Recovering depth

In the stereo geometry as shown in Fig. 29

$$\tan \theta_i = \frac{dL_i}{f}$$

(A.9)

and

$$\tan(\theta - \theta_i) = \frac{dL}{f}.$$ (A.10)

Therefore,

$$\tan^{-1} \frac{dL}{f} = \theta_i,$$ (A.11)

and

$$\tan^{-1} \frac{dL_i}{f} = \theta - \theta_i = \theta - \tan^{-1} \frac{dL}{f}.$$ (A.12)

By equations (A.11) and (A.12),

$$\theta - \tan^{-1} \frac{dL}{f} = \tan^{-1} \frac{dL_i}{f},$$ (A.13)

$$\tan \left(\theta - \tan^{-1} \frac{dL}{f}\right) = \frac{dL}{f} = \frac{\tan \theta - dL/f}{1 + \tan \theta dL/f}.$$ (A.14)

$$dL = f \times \frac{\tan \theta - dL/f}{1 + \tan \theta dL/f}.$$ (A.15)

Fig. 28. The Gaussian possibility distribution.
Fig. 29. The stereo geometry.

Similarly,

\[ dr' = f \times \frac{\tan \theta - dr/f}{1 + \tan \theta dr/f} \]  

The distance \( x \) is calculated by

\[ x = \frac{L \times f}{dl' \times dr'} \times \frac{(1 + \tan \theta dl/f)(1 + \tan \theta dr/f)}{\left(\tan \theta dl/f\right)(1 + \tan \theta dr/f) + (\tan \theta dr/f)(1 + \tan \theta dl/f)} \]  

(A.17)

Substituting

\[ \tan \theta = \frac{L}{2D} \]  

(A.18)

into equation (A.17),

\[
x = \frac{4D^2L^2 + 2DfL^2(dl + dr) + L^2 dl dr}{4DLf^2 + (L^2f - 4D^2f)(dl + dr) - 4DLdL dr}.
\]  

(A.19)

Since

\[ \tan \theta = 0, \]  

(A.20)

in the parallel axis stereo geometry, \( x \) is simplified as

\[
x = \frac{L}{(-dl/f) + (-dr/f)} = \frac{L \times f}{dl + dr}, \]  

(A.21)

where \( dl + dr \) is the disparity value.

REFERENCES

joined the Department of Computer Engineering of Dongseo University as a full-time instructor. His main interests are in image processing, computer vision, and neural networks.

About the Author—YEONG-HO HA received the B.S. and M.S. degrees in Electronic Engineering from Kyungpook National University, Taegu, S. Korea, in 1976 and 1978, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Texas at Austin, Texas, 1985. In March 1986 he joined the Department of Electronic Engineering of Kyungpook National University as an Assistant Professor, and is currently a Professor. His main interests are in digital signal processing, and computer vision. He served as TPC co-Chair of 1994 IEEE International Conference on Intelligent Signal Processing and Communication Systems. He is a member of IEEE, Pattern Recognition Society, Korea Institute of Telematics and Electronics, and Korean Institute of Communication Sciences.