EFFECTIVE PARAMETERS IN SEARCH SPACE REDUCTION
USED IN A FAST EDGE-BASED STEREO MATCHING

PAYMAN MOALLEM

Electronic Engineering Department, Faculty of Engineering, Isfahan University, Hezarjarib Avenue, Isfahan, Iran
P_Moallem@eng.ui.ac.ir

KARIM FAEZ

Electrical Engineering Department, Amirkabir University of Technology, Hafez Avenue, Tehran, Iran
kfaez@aut.ac.ir

Received 23 November 2002
Revised 11 June 2004

Reduction of search region in stereo correspondence can increase performances of the matching process, in the context of execution time and accuracy. For an edge-based stereo matching, we establish relationships between the search space and the parameters like relative displacement of edges, disparity under consideration, image resolution, CCD (Charge-Coupled Device) dimension and focal length of the stereo system. Then, we propose a novel matching strategy for the edge-based stereo. Afterward, we develop a fast edge-based stereo algorithm with combination of the obtained matching strategy and a multiresolution technique using Haar wavelet. Considering the conventional multiresolution technique using Haar wavelet, the execution times of our proposed method are decreased between 26% to 47% in the feature matching stage. Moreover, the execution time of the overall algorithms (including feature extraction and feature matching) is decreased between 15% to 20%. Theoretical investigation and experimental results show that our algorithm has a very good performance; therefore this new algorithm is very suitable for fast edge-based stereo applications like stereo robot vision.

Keywords: Stereo matching; edge; search reduction.

1. Introduction

Stereo vision refers to the ability to infer information on 3D structures and distances of a scene from at least two images (left and right), taken from different viewpoints. A stereo system must solve two essential problems: correspondence and reconstruction. The correspondence consists of determining which item in left image corresponds to which item in right image. It is usually not a good practice to try to find the corresponding points for all pixels. For example, a point in a uniform region in one image may correspond to many points in a corresponding region in other image. Thus, feature points or matching primitives are selected so that an
unambiguous match could be resulted.\(^2\) Depth in a stereo system is related to the
inverse of the disparity, which is the difference between position of the corresponding
points in left and right images. Disparity of image points forms a disparity map,
which can be displayed as an image. If geometry of the stereo system is known, the
disparity map can be converted to a depth map of viewed scene. This process is
called the \textit{reconstruction} step in the stereo algorithms.

Edge-based stereo matching is a popular method in some fast stereo vision
applications.\(^3\),\(^4\) Finding corresponding edges is considered being the most difficult
part of the edge-based stereo matching algorithms. Usually, correspondence for a
feature point in first image is obtained by searching a predefined region of second
image, based on \textit{epipolar line} and \textit{disparity range} constraint.\(^1\)

Corresponding points must lie on corresponding epipolar lines. \textit{Epipolar lines}
are defined by intersection of epipolar planes and an image plane. For each point
in a 3D scene, the corresponding epipolar plane is a plane that contains point and
stereo base line, which is a line connecting the two optical centers of the cameras.\(^1\)
With refer to the maximum and minimum of depth and geometry of a stereo system,
\textit{disparity range} can be estimated\(^5\) and this range is used to limit search region. In
some applications, the disparity range is large. For example, consider a robot stereo
vision that should be active in a wide range of depth.

Reduction of search region can increase performances of the matching process, in
the context of execution time and accuracy. \textit{Coarse to fine strategy} is proposed
by many to reduce the search space. In this strategy, information obtained at a
coscale is used to guide and limit the search for the matching of finer scale
primitives or feature points. In this approach, an initial matching begins at a coarse
scale, where feature density is low due to scale change. This reduction in the feature
density reduces the search space, which in turn makes the matching easier and
coscale, but not necessarily more accurate, because localization at coarse scale is
less accurate. Such multistage strategy can be used with scale specific primitive
representation. These techniques are classified into two categories: \textit{multiresolution}
and \textit{block matching}. In the \textit{multiresolution} strategies, initial matching begins at
the lowest scale and then extended to finer ones.\(^2\) Usage of the hierarchical Gaussian
basis functions\(^6\) and wavelet transform\(^7\) are common techniques in stereo vision. In
the \textit{block matching} methods, the disparity estimation of a large block in the fine
level is used as initial estimation of disparity for the entirety of block, and then the
disparity is calculated within each block accurately.\(^8\)

Considering a threshold on directional derivative of disparity,\(^9\) we could restrict
the search space in an edge-based stereo correspondence and develop some fast
algorithms.\(^10\) We showed that relative displacement of edges have effect in search
space. In this paper, we establish a relationship between search space for edge
matching and geometric parameters of a stereo system like CCD dimension, image
resolution and focal length. Since these effective parameters are related to geometry
of a stereo system, our proposed relationships are system-oriented. Therefore,
we will focus on CIL\textsuperscript{a} stereo image database\textsuperscript{11} at Carnegie Mellon University as an example, since its parameters are well known. We use directional derivative of disparity as an auxiliary concept to conclude the relationship.

In the next section, we briefly discuss the concept of directional derivative of disparity.\textsuperscript{10} In Sec. 3, we represent a relationship between DDD\textsuperscript{b} range and displacement of feature points (like edges) in left and right images.\textsuperscript{9} Considering the pdf\textsuperscript{c} of DDD, we already obtained a probabilistic relationship between the maximum of DDD and some geometric constraints.\textsuperscript{12} Then, we develop those for CIL stereo image database\textsuperscript{13,14} and establish a probabilistic relationship between the DDD range and geometric parameters of stereo system like CCD dimension, image resolution and focal length. Therefore in Sec. 4, the search space effectively reduces and relates to the mentioned effective parameters, including relative displacement of primitives, maximum disparity, image resolution, CCD dimension and finally focal length of stereo system. Considering our proposed search space and the hierarchical multiresolution technique using Haar wavelet,\textsuperscript{7} we suggest a fast edge-based stereo matching algorithm,\textsuperscript{13,14} in Sec. 5. Finally in Sec. 6, we discuss the implementation results.

2. Directional Derivative of Disparity

Figure 1 shows a camera geometry for a basic stereo system where the camera’s optical axes are parallel to each other and perpendicular to the baseline connecting the two cameras L and R. OL and OR are the optical centers of the left and right cameras and O\textsubscript{l} and O\textsubscript{r} are the origins of the left and right image planes. For a point \(P(X,Y,Z)\) in 3D scene, its projections onto the left and right images are \(p_l(x_l,y_l)\) and \(p_r(x_r,y_r)\). Considering this camera geometry, it can be shown that 
\[
y_l = y_r \text{ and the disparity } d \text{ is inversely proportional to the depth } Z:\]
\[
d = x_l - x_r = \frac{b \cdot f}{Z}, \tag{1}
\]
where \(f\) is focal length of the camera lens and \(b\) is the separation of two cameras or baseline.\textsuperscript{1} Given two points \(P_1\) and \(P_2\) in 3D scene, DDD can be defined as the difference in disparities divided by the cyclopean separation, where cyclopean separation is the average distance between \((p^l_1,p^l_2)\) and \((p^r_2,p^r_2)\).\textsuperscript{14} Assume that a virtual camera is placed in the middle of the cameras L and R, i.e., at position of the origin, O\textsubscript{c}, we have:
\[
d_2 = x^2_l - x^2_r, \quad d_1 = x^1_l - x^1_r, \quad p^c_2 = \frac{p^2_l + p^2_r}{2}, \quad p^c_1 = \frac{p^1_l + p^1_r}{2}. \tag{2}
\]
\textsuperscript{a}Calibrated Imaging Laboratory.
\textsuperscript{b}DDD is an abbreviation that we use for Directional Derivative of Disparity.
\textsuperscript{c}probability density function.
Fig. 1. Defining directional derivative of disparity (DDD) in the stereo system with parallel cameras.

With these definitions, DDD or $d_\alpha$ can be defined as:

$$d_\alpha = \frac{|d_2 - d_1|}{r} = \frac{|d_2 - d_1|}{\|p_2^c - p_1^c\|},$$  \hspace{1cm} (3)

where $\|\cdot\|$ denotes the vector norm. The value of $d_\alpha$ can be used to define various stereo matching constraints. A brief summary is shown below:\textsuperscript{9,10}:

- $|d_\alpha| > 2$ — Violation of nonreversal order constraint.
- $|d_\alpha| = 2$ — Violation of uniqueness constraint.
- $|d_\alpha| < 1.1$ or $1.2$ — Empirical limit of DDD.
- $|d_\alpha|$ $\ll$ 1 — Figural continuity constraint.

### 2.1. Applications of DDD in the stereo correspondence

Bult and Julesz\textsuperscript{15} provided evidence supporting the claim that, for binocular fusion of random dot stereograms by human visual system, DDD must not exceed the unity. Pollard, Mayhew and Frisby\textsuperscript{16} suggested that for most natural scene surfaces, including jagged one, DDD value between correct matches is usually less than 1. Discarding the ambiguity in the correspondence problem, they impose a threshold on directional derivative of disparity constraint among the candidate matches. Pollard \textit{et al.}\textsuperscript{17} derived the intrinsic relationship between the directional derivative of disparity, surface orientation, and the depth in 3D scenes. Moallem, Faez and Haddadnia\textsuperscript{12} used the probability density function of DDD to establish some relations between search space and geometric parameters of stereo system. They also used an empirical threshold on DDD to restrict the search region for matching edge points.\textsuperscript{9,10} Li and Hu\textsuperscript{14} used DDD as a basis for a unified cooperative stereo matching. They selected some families of neighborhood support functions based on DDD.
3. DDD and the Displacements of the Edges

Assume that the feature points in the left image are some edges in successive scan lines. We want to find their correspondences in the right image. Considering the points P1 and P2 in Fig. 1, we define the displacements of P1 and P2 in the left and right images, \( \Delta x_l, \Delta x_r \) and \( \Delta y \) as:

\[
\begin{align*}
\Delta x_l &= x_l^2 - x_l^1, \\
\Delta x_r &= x_r^2 - x_r^1, \\
\Delta y_l &= \Delta y_r = \Delta y = y^2 - y^1.
\end{align*}
\]  

We showed that DDD between P1 and P2 could be defined as\(^9,10\):

\[
d_\alpha = \frac{2(\Delta x_l - \Delta x_r)}{\sqrt{(\Delta x_l + \Delta x_r)^2 + (\Delta y_l + \Delta y_r)^2}}. \tag{5}
\]

For the edge points in the successive scan lines, we have \( \Delta y = 1 \), so Eq. (5) can be simplified as:

\[
d_\alpha = \frac{\Delta x_l - \Delta x_r}{\sqrt{(\Delta x_l + \Delta x_r)^2 + 1}}. \tag{6}
\]

Suppose \( x_l^1 = x_l^2 \), it means \( \Delta x_l = 0 \). Considering Eq. (6), we can find an upper and lower limits of \( d_\alpha \) in some ranges of \( \Delta x_r \). Some results are shown in Table 1.

While the \( \Delta x_r \) range is increased, the \(|d_\alpha|\) limit is neared to 2.0, which violates the uniqueness constraint in the stereo system.\(^10\) For other values of \( \Delta x_l \) and \( \Delta x_r \), the ranges of \( d_\alpha \) can be computed. Some results for \( \Delta x_l = +1 \) are shown in Table 2. The other cases (\( \Delta x_l = -1, \Delta x_l = +2 \) and \( \Delta x_l = -2 \)) can be simply computed.

| Table 1. Range of \( d_\alpha \) for some values of \( \Delta x_r \) in case of \( \Delta x_l = 0 \). |
|---|---|
| Some values of \( \Delta x_r \) | Range of \( d_\alpha \) |
| -1, 0, +1 | -0.89 < \( d_\alpha \) < +0.89 |
| -2, -1, 0, +1, +2 | -1.41 < \( d_\alpha \) < +1.41 |
| -3, -2, -1, 0, +1, +2, +3 | -1.66 < \( d_\alpha \) < +1.66 |
| -4, -3, -2, -1, 0, +1, +2, +3, +4 | -1.79 < \( d_\alpha \) < +1.79 |

| Table 2. Range of \( d_\alpha \) for some values of \( \Delta x_r \) in case of \( \Delta x_l = +1 \). |
|---|---|
| Some values of \( \Delta x_r \) | Range of \( d_\alpha \) |
| 0, +1, +2 | -0.55 < \( d_\alpha \) < +0.89 |
| -1, 0, +1, +2, +3 | -0.89 < \( d_\alpha \) < +2.0 |
| -1, 0, +1, +2, +3, +4 | -1.11 < \( d_\alpha \) < +2.0 |
| -1, 0, +1, +2, +3, +4, +5 | -1.26 < \( d_\alpha \) < +2.0 |
In Table 2, $\alpha$ is +2 in the case of $\Delta x_l = +1$ and $\Delta x_r = -1$. Since the limit of $|\alpha|$ is 2, there is no need to decrease $\Delta x_r$ lower than $-1$. On the other hand, while $\Delta x_r$ can be increased toward positive infinity, $\alpha$ is also nearer to $-2$, which is the lower limit of $\alpha$.

4. DDD Ranges, Disparity Values and Search Space

In the previous section, DDD range was related to $\Delta x_l$ value and $\Delta x_r$ range. When $\Delta x_l$ is known, the $\Delta x_r$ range can be selected if the range of $\alpha$ is already determined. In this section, we will show that the range of $\alpha$ is a probabilistic phenomenon and it is dependent on the stereo system parameters and the disparity value. Probability density function (pdf) of DDD for a point $(x_l, y)$ in the left image with disparity $d$ may be found by mapping DDD of Eq. (2) into a tangent direction at $[X, Y, Z]$ in the 3D space of the scene, and using the relationship to transform the pdf of the scene coordinate tangent to the pdf of DDD. Performing these calculations and approximating the results for simplification results in:

$$P_{\alpha}(\alpha) = \frac{f/d}{2(1 + (f/d)^2\alpha^2)^{3/2}}.$$  \hspace{1cm} (7)

The approximations performed in Eq. (7) for simplification purpose is only valid for proper ranges of camera parameters and geometry of the stereo system. Typically, $f$ is in the range 10 to 60 millimeters (mm) and $b$ may be 5 to 20 centimeters (cm).

The general shape of this pdf is similar to a Gaussian function. Moreover, this pdf depends explicitly on the disparity under consideration ($d$) and the focal length of the cameras ($f$), or equivalently on the depth and baseline (since $f/d = Z/b$). As an example, consider the CIL images database. Some geometric parameters of the CIL images are listed in Table 3.

Therefore, the focal length of the CIL is about 2478 pixels ($57 \text{mm}/23 \mu\text{m} \approx 2478$). Figure 2 shows the pdfs for the various disparity values as a function of DDD values. These functions have some sharp peaks near zero DDD and become wider when the disparity increases. Assume the range of $\alpha$ is $R\alpha = \{d^L_{\alpha} < \alpha < d^H_{\alpha}\}$, therefore the $P(\alpha \in R\alpha)$ can be obtained by integrating Eq. (7) over $R\alpha$ as:

$$P(\alpha \in R\alpha) = \int_{d^L_{\alpha}}^{d^H_{\alpha}} P_{\alpha} \cdot d(\alpha).$$  \hspace{1cm} (8)

When $R\alpha$ is a subset of $[-2, +2]$, then we have a restricted search region and we should allow some errors. Assume that this error is less than 0.5% that

<table>
<thead>
<tr>
<th>Table 3. Some of the CIL geometric parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal cells in CCD Array</td>
</tr>
<tr>
<td>Horizontal pixels in computer frame buffer</td>
</tr>
<tr>
<td>Distance between sensor elements (mm)</td>
</tr>
<tr>
<td>Focal length (mm)</td>
</tr>
</tbody>
</table>
$P(d_\alpha \in Rd_\alpha) > 0.995$. In a typical stereo algorithm, sometimes up to 10% error in the matching stage can be acceptable, so that the condition $P(d_\alpha \in Rd_\alpha) > 0.995$ is suitable,\textsuperscript{13} since the error due to the search region selection is less than 0.5%.

We are interested in finding a maximum of the disparity (or $d_{\text{max}}$) so that $P(d_\alpha \in Rd_\alpha) > 0.995$. For example, suppose $\Delta x_l = +2$, $\Delta y = +1$ and $\Delta x_r = \{0, +1, +2, +3\}$, so $d_\alpha$ is varied between $d_\alpha^L = -0.37$ and $d_\alpha^R = +2.0$. We want to find the disparity $d_{\text{max}}$ so that $P(d_\alpha \in Rd_\alpha) > 0.995$. We increase $d$ from zero and compute $P(d_\alpha \in Rd_\alpha)$ for each value of $d$ considering Eq. (8). The maximum value of $d$ that satisfies the condition $P(d_\alpha \in Rd_\alpha) > 0.995$ is proposed as $d_{\text{max}}$. In this example, $d_{\text{max}}$ is 130 pixels.

Tables 4 and 5 show the relations between $\Delta x_l$ and $\Delta x_r$ by context $P(d_\alpha \in Rd_\alpha) > 0.995$; for other values of $\Delta x_l$ and $\Delta x_r$, the computations are similar.

5. Fast Algorithm

In the calibrated stereo system with parallel optical axes, area-based or feature-based algorithms consist of two stages: feature point (or primitives) extraction and

<table>
<thead>
<tr>
<th>Some values of $\Delta x_r$</th>
<th>$d_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1, 0, +1$</td>
<td>230</td>
</tr>
<tr>
<td>$-2, -1, 0, +1, +2$</td>
<td>360</td>
</tr>
<tr>
<td>$-3, -2, -1, 0, +1, +2, +3$</td>
<td>420</td>
</tr>
<tr>
<td>$-4, -3, -2, -1, 0, +1, +2, +3, +4$</td>
<td>450</td>
</tr>
</tbody>
</table>
Table 5. $d_{\text{max}}$ for some values of $\Delta x_r$ in case of $\Delta x_l = +1$.

<table>
<thead>
<tr>
<th>Some values of $\Delta x_r$</th>
<th>$d_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, +1, +2</td>
<td>170</td>
</tr>
<tr>
<td>-1, 0, +1, +2, +3</td>
<td>290</td>
</tr>
<tr>
<td>-1, 0, +1, +2, +3, +4</td>
<td>340</td>
</tr>
<tr>
<td>-1, 0, +1, +2, +3, +4, +5</td>
<td>380</td>
</tr>
</tbody>
</table>

**stereo matching.** Most of the fast stereo algorithms use low-level primitives like edges, those that do not require sophisticated semantic analysis in their extraction.\(^4\) Matching the horizontal edges in the stereo system with parallel optical axes is an open problem,\(^20\) so some authors suggested nonhorizontal edge points as feature points.\(^4,9,10,13,14,19,21,22\) In a correlation-based framework, stereo matching for a pixel in a reference image (left) is obtained by searching in a predefined region of the second image (right).

Our matching algorithm has two stages: edge extraction and edge matching. The edge extraction stage consists of identifying nonhorizontal thinned edges, extracted in the left and right images. The former, the nonhorizontal edge extraction, is performed and thinned edge points are classified into two groups, positive and negative, depending on the gray level difference between the two sides of the edge points in the $x$ direction. The left and right images are convolved with a proper gradient mask. To detect nonhorizontal edge points, we use a gradient of two-dimensional Gaussian function with $\sigma = 1$ in $x$ direction as the gradient mask. This mask is truncated to $5 \times 5$. A nonhorizontal thinned positive edge in the left image is localized to a pixel that the filter response or $\rho_l(x, y)$ has to exceed a positive threshold $\rho_0^+$ and has to obtain a local maximum in $x$ direction. Therefore,\(^23\)

\[
\begin{align*}
\rho_l(x, y) &> \rho_0^+ \quad \text{← Threshold} \\
\rho_l(x, y) &> \rho_l(x-1, y) \quad \text{← Local Maximum} \\
\rho_l(x, y) &> \rho_l(x+1, y) 
\end{align*}
\]

The positive threshold $\rho_0^+$ is considered pre-defined or can be selected based dynamically on the filter response. Consider $\rho_{\text{mean}}^+$ is the mean of the positive values of the filter response, we choose $\rho_0^+ = 1.5 \rho_{\text{mean}}^+$, the coefficient of 1.5 is empirically selected. The extraction of nonhorizontal negative thinned edge points is similar to the positive ones. Consider $\rho_{\text{mean}}^-$ is the mean of the negative values of the filter response, the negative threshold value is $\rho_0^- = 1.5 \rho_{\text{mean}}^-$. Before posing the matching strategy, we should define a successive connected edge set and its properties.

**5.1. Successive connected edge points**

Some connected edge points in the left image can be grouped together as some sets that we call successive connected edges (SCE) sets.\(^10\) Each SCE set $\Psi$ consists of
n successive edges, \( \Psi = \{ p^l_1, p^l_2, p^l_3, \ldots, p^l_n \} \). Its coordinates and type identify each edge point as \( p^l_k = (x^l_k, y^l_k, \text{type}) \). The set \( \Psi \) is said a successive connected edges set if these three conditions are met:

1. The types of all \( p^l_k \) are the same.
2. The successive \( p^l_k \) are in successive scan lines. On the other hand, we have, \( y^l_{k+1} - y^l_k = 1 \).
3. The absolute difference between \( x \) values of the two successive points is less than 3, so we have, \( x^l_{k+1} - x^l_k = \{-2, -1, 0, +1, +2\} \).

### 5.2. Search space in the edge matching

Suppose that we want to find the correspondences of the edge points of a sample SCE set in left image. Considering \( d_{\text{max}} \) and the connectivity of the edge points in a SCE set, if value of \( \Delta x_l \) is known, \( \Delta x_r \) range can be computed directly by previous discussion in Sec. 4 (see Tables 4 and 5).

For example consider \( p^l_{i-1} \) and \( p^l_i \) are two successive edge points in a sample SCE set and \( A_R \) is the correspondence of \( p^l_{i-1} \) in the right image (see Figs. 3(a) and 3(b)). Considering Eq. (4), \( \Delta x_l \) is the difference between the \( x \) position of \( p^l_i \) and \( p^l_{i-1} \) in the left. For finding \( B_R \) (the correspondence of \( p^l_i \)) in the right image, the search space region or \( \Delta x_r \) range can be computed by the mentioned tables. This search region in the right image is shown in Fig. 3(b). \( X_{\text{Range}}^R \) is considered as the search space in the right image coordinate system, can be computed as:

\[
\begin{align*}
    d &= x^l_{i-1} - x^r_{A} \\
    \Delta x_l &= x^l_i - x^l_{i-1} \Rightarrow X_{\text{Range}}^R = \Delta x_r + x^l_{i-1} - d. \\
    \Delta x_r &= x^r_A - X_{\text{Range}}^R
\end{align*}
\]  

(10)

In Fig. 3, \( \Delta x_l \) is considered to be \(-1\) and if \( d_{\text{max}} < 380 \) then \( \Delta x_r \) range is \( \{-5, -4, -3, -2, -1, 0, +1\} \), therefore the search region is restricted to seven pixels only. The search region is also shown in Fig. 3(b). If \( d_{\text{max}} < 170 \) then \( \Delta x_r \) range
is only \{-2, -1, 0\}. See Table 5 for the case of $\Delta x_l = +1$, the case $\Delta x_l = -1$ is similar in the negative direction.

5.3. **SCE set construction and matching strategy**

In the process of edge matching, after finding a correspondence point, the pair matched edge is deleted from the left and right edge images. Therefore each matched edge point is examined only one time. In the following edge matching strategy, the SCE set constructing the edge points of the left image and establishing the correspondence are implemented concurrently. The proposed matching strategy can be expressed briefly in two phases as follows:

**Phase I.**

Find the next edge in left image by systematic scan, from the left to right and from the top to bottom. The found edge is considered as $p^l_1$, the first point of the corresponding SCE set. Consider $X^r_{\text{Range}}$ in the right image coordinate as the full search region based on the epipolar line and other constraints, and set $i = 1$. Go to phase II.

**Phase II.**

1. Find the correspondence of $p^l_i$ in the right image, $B_R$ (see Fig. 3(b)) based on $X^r_{\text{Range}}$. If $B_R$ is found, delete $p^l_i$ and $B_R$ from the edge images and go to step 2, else go to phase I.
2. Compute $d = x^l_i - x^r_B$ (Eq. (1)), and go to step 3.
3. Find $p^l_{i+1}$, the next point of $p^l_i$ in the corresponding SCE set based on the three conditions of SCE sets in Sec. 5.1 and go to step 4. If $p^l_{i+1}$ is not found, go to phase I.
4. Compute $\Delta x_l = x^l_{i+1} - x^l_i$ (Eq. (4)) and go to step 5.
5. Compute $\Delta x_r$ range, considering previous discussion in Sec. 4 and then go to step 6.
6. Compute $X^r_{\text{Range}} = \Delta x_r + x^l_{i-1} - d$ (Eq. (10)) and go to step 7.
7. Set $i = i + 1$ and got to step 1.

5.4. **The overall algorithm**

The search space reduction is very important to decrease the processing time of the matching. In the proposed SCE sets construction and matching strategy, from the second point up to end point of SCE sets, we could effectively reduce the search space for establishing the correspondence. On the other hand, considering a SCE set in the left image as $\Psi = \{p^l_1, p^l_2, p^l_3, \ldots, p^l_n\}$, the search space can be reduced for the points of $p^l_i$, $i = 2, 3, \ldots, n$ but for the first point that is found in the first phase of proposed matching strategy, i.e., $p^l_1$, we use the conventional multiresolution technique using Haar wavelet\(^7\) to reduce the search region. We use three levels of Haar wavelet (original + two lower resolutions). We use the normalized cross
correlation (NCC) with the window size of $5 \times 5$, $3 \times 3$ and $3 \times 3$ for the coarse, medium and fine level. In the coarsest level, the search space is selected by considering the maximum of the disparity. At the higher resolution levels, the search space is cut around each found maximum correlation location in the previous level, $\pm 7$ pixels along the epipolar line. At the highest resolution, only the similar edges on the epipolar line of other image are examined.

For the second point up to end point of SCE sets, considering the relationship between disparities, $\Delta x_l$ and $\Delta x_r$, we showed that the search region could be reduced. If the disparity search range could be automatically reduced to an effective range (about 10 pixels or less), then several local maximum would stay out of the selection process and therefore, the disparities found would be correct, even if the size of the matching block is small. Therefore, we can use NCC with a window size of even $3 \times 3$ in the restricted search region. Hence our algorithm uses multiresolution technique using Haar wavelet to reduce the search region for the first point of a SCE set and then use the restricted search region based on the mentioned relationships.

6. Implementation Results

As test images, three different stereo scenes are selected from CIL database at CMU, we called them Texture, Home and Castle. In this set, maximum disparity is less than 130 pixels, therefore we consider $d_{\max} = 130$ for our computations. Regarding our previous discussion for CIL stereo scenes, Table 6 shows the restricted search region for the values $\Delta x_l = \{-2, -1, 0, +1, +2\}$, when $\Delta y = 1$. Therefore $\Delta x_r$ range in part 5 of phase II of the SCE set construction and matching algorithm is selected based on Table 6.

All the scenes used are gray levels, $576 \times 384$ and their real disparity ranges are shown in Table 7. The maximum disparity, the interested areas in these scenes are

<table>
<thead>
<tr>
<th>Scene</th>
<th>$x_{min}$</th>
<th>$x_{max}$</th>
<th>$y_{min}$</th>
<th>$y_{max}$</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture</td>
<td>188</td>
<td>450</td>
<td>44</td>
<td>332</td>
<td>90</td>
<td>115</td>
</tr>
<tr>
<td>Home</td>
<td>140</td>
<td>560</td>
<td>20</td>
<td>370</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td>Castle</td>
<td>140</td>
<td>560</td>
<td>20</td>
<td>370</td>
<td>60</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 6. The relationship between $\Delta x_l$ and the search region for CIL stereo scenes when $\Delta y = 1$ and $\Delta x_l$ is between $-2$ to $+2$.

<table>
<thead>
<tr>
<th>$\Delta x_l$</th>
<th>$-2$</th>
<th>$-1$</th>
<th>$0$</th>
<th>$+1$</th>
<th>$+2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min $\Delta x_r$</td>
<td>$-3$</td>
<td>$-2$</td>
<td>$-1$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>Max $\Delta x_r$</td>
<td>$0$</td>
<td>$0$</td>
<td>$+1$</td>
<td>$+2$</td>
<td>$+3$</td>
</tr>
</tbody>
</table>

Table 7. The disparity ranges and interested areas of the tested scenes.
limited and are also shown in Table 7. Table 3 also shows the geometric parameters of CIL stereo image database. To determine the error in the matching stage of the algorithms, we compute the accurate disparity map for the nonhorizontal edge points of the left images by searching in the predefined region of the right image based on Table 7 and NCC with window size of $15 \times 15$. The errors in these computed disparity maps are corrected manually.

We compare the results of our algorithm that we called RS (Restricted Search) with two fast edge-based stereo methods. The first one that we called HM, is a hierarchical multiresolution method based on Haar wavelet. This method is exactly based on the matching of the first point of SCE set in the proposed algorithm. The other method that we called FS (Full Search), finds the correspondence of nonhorizontal edge points of the left image by examining all of the similar edge points in the epipolar line of the right image in the disparity range 0 to 130 pixels. As the similarity criterion, NCC with window size of $11 \times 11$ and the threshold value of 0.8 were selected. The feature points used in the reference algorithms as well as our algorithm are nonhorizontal edge points.

The codes of the algorithms were written by WatCom C and implemented by a PC under Windows operating system, with a Pentium II 450 MHz processor. As we mentioned before, our stereo algorithm has two stages, feature extraction and feature matching. Therefore at first the results of implementation of these two stages are discussed separately and then the execution time of the overall algorithm is investigated.

6.1. Feature extraction

The feature extraction stage of all investigated algorithms is the same and includes edge detection and multiresolution computing, for the left and right images. The multiresolution computing includes constructing image pyramids both in gray level and edge points. The feature extraction is done for both of the left and right images. Table 8 shows the results. In this table, the execution time of both edge detection in fine level of pyramid and multiresolution computing are shown separately. For FS algorithm, the multiresolution computing is not needed. Moreover, the left edge column shows the total number of extracted edges from the interested area of the left image, including positive and negative edges. In fact, this column shows the number and type of edges should be matched. The execution time column is in millisecond scale.

The extracted edge points for Texture, Home and Castle stereo scenes are shown in Figs. 4 to 6, respectively (parts (c) and (d)). In these figures, the red and blue color are used to show the positive and negative edge points, respectively.

6.2. Feature matching

For Texture scene, the results of the investigated algorithms are shown in Table 9. Considering the total number of extracted features from the left image that are
Table 8. The results of feature extraction on the stereo scenes, left edge column shows the number of extracted edges from the interested areas of left image.

<table>
<thead>
<tr>
<th>Scene</th>
<th>Total</th>
<th>Positive</th>
<th>Negative</th>
<th>Edge extraction</th>
<th>Multiresolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture</td>
<td>8354</td>
<td>4054</td>
<td>4500</td>
<td>87</td>
<td>56</td>
</tr>
<tr>
<td>Home</td>
<td>3261</td>
<td>1523</td>
<td>1738</td>
<td>86</td>
<td>56</td>
</tr>
<tr>
<td>Castle</td>
<td>5793</td>
<td>2731</td>
<td>3062</td>
<td>90</td>
<td>53</td>
</tr>
</tbody>
</table>

Table 9. The implementation results on Texture scene, where 8354 features are extracted from the interested area of left image.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Matched (%)</th>
<th>Failed (%)</th>
<th>Error (%)</th>
<th>Execution time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>69.8</td>
<td>30.2</td>
<td>1.1</td>
<td>1942</td>
</tr>
<tr>
<td>HM</td>
<td>65.8</td>
<td>34.2</td>
<td>0.5</td>
<td>340</td>
</tr>
<tr>
<td>RS</td>
<td>64.2</td>
<td>35.8</td>
<td>0.6</td>
<td>288</td>
</tr>
</tbody>
</table>

Table 10. The implementation results on Home scene, where 3261 features are extracted from the interested area of left image.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Matched (%)</th>
<th>Failed (%)</th>
<th>Error (%)</th>
<th>Execution time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>91.5</td>
<td>8.5</td>
<td>7.4</td>
<td>386</td>
</tr>
<tr>
<td>HM</td>
<td>95.1</td>
<td>4.9</td>
<td>5.6</td>
<td>221</td>
</tr>
<tr>
<td>RS</td>
<td>94.1</td>
<td>5.9</td>
<td>4.5</td>
<td>188</td>
</tr>
</tbody>
</table>

Table 11. The implementation results on Castle scene, where 5793 features are extracted from the interested area of left image.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Matched (%)</th>
<th>Failed (%)</th>
<th>Error (%)</th>
<th>Execution time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>88.9</td>
<td>12.0</td>
<td>11.3</td>
<td>992</td>
</tr>
<tr>
<td>HM</td>
<td>85.9</td>
<td>14.1</td>
<td>10.6</td>
<td>251</td>
</tr>
<tr>
<td>RS</td>
<td>86.4</td>
<td>13.6</td>
<td>9.3</td>
<td>200</td>
</tr>
</tbody>
</table>

shown in Table 8, the percentage of the matching, failed and error in the matching stage are shown in the matched, failed and error columns, respectively. Considering the accurate disparity map for the feature points, the error of the matching stage could be computed. The error column shows the percentage of this error with respect to the total extracted left features. The execution time column is in millisecond scale. Tables 10 and 11 show the results of matching and the execution times of the algorithms on Home and Castle stereo scenes, respectively.
Figure 4\textsuperscript{d} shows the implementation results on Texture stereo scene. This figure shows the left and right images (a) and (b) and the extracted features including nonhorizontal edge points (c) and (d). The negative and positive type of edges are shown with blue and red color, respectively. Moreover, images (e) through (g) show the disparity maps of the interested areas of the left image obtained by the algorithms. In the disparity map figures (e) through (g), the failed edges and the error in the matching are shown with magenta and black color, respectively. Figures 5 and 6\textsuperscript{d} show the corresponding disparity maps and other results derived by the algorithms, on Home and Castle stereo scenes, respectively.

Fig. 4. The implementation results on Texture stereo scene. The positive and negative edges are shown with red and blue color, respectively, in figures (c) and (d). The disparity range is between 90 to 115 pixels. The disparity maps are also shown in figures (e) through (f), respectively.

Fig. 5. The implementation results on Home stereo scene. The disparity range is between 50 to 70 pixels.

\textsuperscript{d}To see color figures, please write to sales@wspc.com.sg.
7. Discussion

The execution time, the failed and error rates are some essential parameters in the proposed algorithms. Therefore we first discuss these and then investigate some effective parameters about execution time of our proposed algorithm.

7.1. Execution times

In a complete matching algorithm, the processing time includes the feature extraction time plus the matching time. Table 12 shows the execution time of the algorithms including the execution time of the edge extraction, multiresolution and matching shown separately.

Comparing the execution time of the stages to each other, it can be seen that the edge extraction time is nearly constant but the edge matching time is highly dependent on both the scenes and algorithms. This phenomenon has a great effect on the performances of our proposed algorithm. On the other hand, the execution time of the edge matching in the proposed algorithm is highly reduced with respect to the others. Considering HM as the fastest method, RS reduces the matching time between 26% to 47%, and the overall execution time between 15% to 20%.

Table 12. Execution time of the algorithms in milliseconds, the first item in parentheses is the edge matching time and the second one is the edge extraction time plus multiresolution (if needed).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Texture</th>
<th>Home</th>
<th>Castle</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>1942 (1855 + 87)</td>
<td>386 (299 + 87)</td>
<td>992 (905 + 87)</td>
</tr>
<tr>
<td>HM</td>
<td>340 (197 + 143)</td>
<td>221 (79 + 142)</td>
<td>251 (108 + 143)</td>
</tr>
<tr>
<td>RS</td>
<td>288 (145 + 143)</td>
<td>188 (46 + 142)</td>
<td>200 (57 + 143)</td>
</tr>
</tbody>
</table>
Table 13. Comparing the failed and error rates of algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Failed (%)</th>
<th>Error (%)</th>
<th>Failed (%)</th>
<th>Error (%)</th>
<th>Failed (%)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture</td>
<td>30</td>
<td>1.1</td>
<td>34</td>
<td>0.5</td>
<td>35.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Home</td>
<td>8.5</td>
<td>7.4</td>
<td>4.9</td>
<td>5.6</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Castle</td>
<td>12.0</td>
<td>11.3</td>
<td>14.1</td>
<td>10.6</td>
<td>13.6</td>
<td>9.3</td>
</tr>
</tbody>
</table>

7.2. Failed and error rates

Table 13 shows the failed and error rate of the algorithms. The failed and error rate of RS algorithm is close to the reference algorithms. On Castle scene, the failed and error rates of all algorithms are high. This scene has a $10 \times 16$ grid of black dots over a white background that is used for camera calibration. Since the considered maximum disparity is higher than the distance between calibration dots in an image, for each dot in the left image, there are some matching candidate dots in the right image. Therefore the error rate on Castle scene is higher than others (see Figs. 6(e)–6(g)). On Texture scene where the interested area is selected out of the dot grids, the error rate is lower than others.

7.3. Important factors in the processing time

Conventionally, for the nonfirst points of SCE sets, our proposed matching strategy works faster than the first point, since the search spaces for the nonfirst points are more restricted. Therefore, when the number of first point of SCE sets are decreased relative to nonfirst points, we expect that the processing time should be lower relatively. In fact, the connectivity of edge points in successive scan lines are important factor in decreasing the matching time. If the number of edge points in each SCE set is increased, the matching time will be decreased, especially in complex scenes. In Castle and Home, the connectivity of edge points is higher than Texture. Therefore considering HM algorithm, the reduction of matching time of our algorithm in Castle (47.2%) and Home (41.8%) scenes is higher than Texture (26.4%).

In a complete matching algorithm, the processing time includes the feature extraction time plus the matching time. Therefore other important parameter in the processing time is the feature extraction time. Generally, the feature extraction time is nearly independent of edge point density, because the main computation of this stage is the convolution of the images with a predefined proper mask. Table 8 shows that the execution time of feature extraction is nearly constant. On the other hand, the execution time of matching is mostly dependent on parameters like connectivity of edge points.

8. Conclusion

Edge-based stereo matching is a popular method in some fast stereo vision applications. Reducing the search space in the matching can increase the execution speed
and accuracy. In this paper, we introduced the concept of search space reduction in edge-based stereo by limiting the range of directional derivative of disparity, $d_\alpha$, considering the geometric parameters of stereo system like, focal length of lens, CCD and image dimensions and maximum disparity. Then, we posed a fast matching strategy for the successive connected edge points, SCE sets in two phases, based on that restriction. In fact, we highly restricted the search space region in the matching of the SCE sets for nonfirst points (phase II). We used the hierarchical multiresolution scheme to reduce further the search space in phase I of the proposed algorithm. Finally, we developed a fast edge-based stereo matching algorithm, RS. The matching execution time of our algorithm is lower than other algorithms, but the matched and error rates are close to the reference algorithms.

Connectivity of edge points are very important factor to decrease the execution time of our algorithm, therefore our method might not work well in a scene like random dot, but work well in a real scene where the connectivity of edge points is guaranteed. Therefore this algorithm is very suitable for real time applications like robotics, where the effective parameters in the search space are well known and the connectivity of edge points are guaranteed.

The execution time of edge extraction stage that is implemented separately for left and right images is nearly independent of scene complexity. Therefore to decrease the execution time, we suggest using two separate processors for implementation of edge extraction for left and right images. Another processor can be used in pipeline manner to implement the matching stage. Therefore the mentioned parallel architecture can be used to implement the proposed edge-based stereo algorithms in real time applications.

References