Confidence-based iterative efficient large-scale stereo matching

Hao Ma¹, Shunyi Zheng¹,²*, Yingsong Li¹, Li Gui¹,²,³, Rongyong Huang⁴,⁵ and Haitao Wei⁶

Abstract: In this study, we integrate confidence into efficient large-scale stereo (ELAS) matching to produce a more accurate approach to binocular stereo for high-resolution image matching. Elas ensures good performance in the presence of poorly textured and slanted surfaces, but one of its deficiencies is its unsatisfactory ability to capture disparity discontinuities. Our formulation explicitly models the effects of confidence as a likelihood term in a principled manner using the Bayes rule. Because it is an iterative method, we associate each point with a variable confidence value and update this value based on a given confidence updating rule. Meanwhile, complementary support points are selected from stable points whose confidence value exceeds a predefined threshold, which differs from ELAS, whose support points are matched in advance and kept unchanged in the subsequent process. Confidence also plays a vital role in avoiding expensive computation, and the adjustment of support points makes disparity estimation more flexible. Quantitative evaluation demonstrates the effectiveness and efficiency of the proposed formulation in improving the accuracy of disparity estimation.

ABOUT THE AUTHORS

Hao Ma is a PhD candidate at Wuhan University, Wuhan, PR China. Her main research direction is dense stereo matching in computer vision.

Shunyi Zheng is a professor at the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, PR China. His research interests include photogrammetry, remote sensing, and computer vision.

Yingsong Li is a PhD candidate at Wuhan University, Wuhan, PR China, with research interests including photogrammetry, remote sensing, image processing, and handheld scanners.

Li Gui is a postdoctoral fellow at the School of Electronic Information, Wuhan University, Wuhan, PR China. His research directions include photogrammetry and image mosaic.

Rongyong Huang is a teacher at the School of Marine Sciences, Guangxi University, Nanning, PR China. His research interests include photogrammetry, remote sensing, and image matching.

Haitao Wei is a teacher at the Institute of Smart City, Zhengzhou University, Zhengzhou, PR China. His research interests include photogrammetry, remote sensing, and image processing.

PUBLIC INTEREST STATEMENT

As an important technique of three-dimensional reconstruction, binocular stereo matching is undergoing rapid development. ELAS is an efficient stereo matching method for large-scale images. To improve the performance of ELAS, CI_ELAS is proposed in this paper by integrating the concept of confidence. High confidence means high reliability. After an evaluation on data-set from Middlebury and KITTI benchmarks, the proposed method outputs a more reliable disparity map, which can be further used to generate three-dimensional information and reconstruct the scene.
1. Introduction
The estimation of disparity maps from binocular imagery has played a fundamental role in computer vision and multimedia processing for decades. Many researchers are dedicated to the solution of stereo matching with high accuracy and efficiency. Related algorithms mainly include local algorithms and global algorithms (Facciolo, de Franchis, & Meinhardt, 2015; Scharstein & Szeliski, 2002; Tombari, Mattoccia, & Di Stefano, 2007). Local algorithms have an advantage over the global algorithms in terms of inexpensive computation. Disparity maps are estimated in a pixel-wise fashion by comparing features over a support region (a concatenation of the features of pixels in a window centered in the pixel) of the reference and target images. However, they are susceptible to image noise and are largely ambiguous in poorly textured or repetitive regions. Global algorithms integrate prior constraints into a correlation-based stereo model to decrease the matching ambiguities (Boykov, Veksler & Zabih, 1998; Cheng & Caelli 2007; Felzenszwalb & Huttenlocher, 2006; Kolmogorov & Zabih, 2001; Woodford, Torr, Reid, & Fitzgibbon, 2009). A compatibility function that expresses the compatibility between neighboring disparities is typically introduced to aggregate support (Tappen & Freeman, 2003), giving rise to the Markov random field (MRF) based model. Optimization of the resulting MRF based energy function is generally considered to be NP-hard. Many techniques such as graph cut (Boykov et al., 1998; Kolmogorov & Zabih, 2001; Taniai, Matsushita, & Naemura, 2014) or belief propagation (Felzenszwalb & Huttenlocher, 2006) have been proposed to solve them effectively, but these algorithms are all too slow for images of a reasonable size (Szeliski et al., 2008).

To deal efficiently with large images, efficient large-scale stereo (ELAS) is proposed as a generative probabilistic model for stereo matching that allows for dense matching with small aggregation windows by reducing ambiguity in the correspondences (Geiger, Roser, & Urtasun, 2011). A piecewise linear prior is constructed over the disparity space by forming a triangulation on a set of robustly matched correspondences known as support points. This method can achieve state-of-the-art performance and enables real-time stereo matching at resolutions greater than 1 megapixel. The proposed prior of ELAS assures good performance in the presence of poorly textured and slanted surfaces, but its deficiencies include unsatisfactory performance in capturing disparity discontinuities (Geiger et al., 2011).

To address this problem, we incorporate into ELAS the concept of confidence (Hu & Mordohai, 2010), which can describe the correctness of the estimated disparity assigned by the matching strategy. We named the resulting algorithm confidence-based iterative ELAS (CI_ELAS). Our formulation explicitly models the effect of confidence as a likelihood term in a principled manner using the Bayes rule. We associate each point with a variable confidence value and update this value in an iterative manner based on a predefined confidence updating rule. In the first iteration, the disparity estimation process is the same as that of ELAS; during the next iteration, the support points, which are used to construct a new triangulation, are those selectively obtained in the preliminary and previous steps with a high level of confidence. Other stable points with high confidence levels are kept unchanged during subsequent iterations. The entire process is halted once the maximum number of iterations is achieved or the number of unstable points shows no obvious change, which means that the disparities of most image pixels are stable and reliable. Depth estimation can be improved by introducing the notion of confidence. As demonstrated in our experiments, the proposed method can achieve satisfactory behavior and outperforms ELAS with a tolerable sacrifice in efficiency because of its iteration scheme.

This paper is organized as follows. Section 2 briefly describes ELAS. Section 3 introduces the proposed method in detail. Section 4 gives the qualitative and quantitative results of the proposed
algorithms and makes comparisons with other well-known stereo methods. Section 5 includes conclusions and directions for future work.

2. ELAS

In this section, we briefly describe the general process of ELAS before introducing our method. ELAS (Geiger et al., 2011) is an efficient approach to binocular stereo that was inspired from the observation that many stereo correspondences are highly ambiguous, whereas some are support points and can be robustly matched due to their texture and uniqueness. Specifically, ELAS makes use of a generative model to estimate the disparity maps. The input images are assumed to be rectified such that the correspondences are restricted to the same horizontal line in both images. The support points are first obtained to construct a triangulation, and every triangular mesh corresponds to a plane defined by the disparity value and location in an image plane of its vertices. The mean disparity value of each pixel located in a triangular mesh can be computed using the plane model. The seeking range for possible disparities of each pixel is then either an interval centered in its mean disparity value or in a set of its surrounding support point disparities in a small 20 × 20-pixel neighborhood.

The support points are matched on a regular grid using the l₁ distance between vectors formed by concatenating the horizontal and vertical Sobel filter responses of the 9 × 9-pixel windows. The Sobel masks are 3 × 3 pixels, and the grid has a fixed step-size of 5 pixels. The support points are consistent with the following parameters: 1) support points should be matched from left to right and from right to left; 2) their ratio between the best and second-best matches is less than a threshold; and 3) their disparity values are consistent with the following parameters: 1) support points should be matched from left to right and 3) their disparity values are consistent with all surrounding support points. Let S = {s₁, ..., s_n} be a set of robustly matched support points, where s_n = (u_n, v_n, d_n) and the first two elements indicate the location and the third indicates the disparity. Let O = (o₁, ..., o_N) be a set of image observations with each observation o_n = (u_n, v_n, f_n) formed as the concatenation of its image coordinates and a feature vector, and o_n(₁) and o_n(₂) are denoted as the observations in the left and right images, respectively. Here, N is the number of pixels in the image, M is the number of matched support points, and f_n is the concatenation of the horizontal and vertical Sobel filter responses. The disparity is computed based on maximum a posteriori (MAP) estimation, and o₁(₁), ..., o_N(₁) denotes all observations in the right image that are located on the epipolar line of o₁(₁):

\[
d_{n}^{c} = \arg \max P(d_{n}|o_{n}^{(1)}, o_{1}^{(2)}, ..., o_{N}^{(2)}, S),
\]

where d_n is a disparity value and

\[
P(d_{n}|o_{n}^{(1)}, o_{1}^{(2)}, ..., o_{N}^{(2)}, S) \propto p(d_{n}|o_{n}^{(1)}, S)p(o_{1}^{(2)}, ..., o_{N}^{(2)}|o_{n}^{(1)}, d_{n}) \propto p(d_{n}|o_{n}^{(1)}, S)(\sum_{n=1}^{N} p(o_{i}^{(2)}|o_{n}^{(1)}, d_{n})),
\]

the conditional probability of d_n given o_n(₁) and S is formulated as

\[
P(d_{n}|o_{n}^{(1)}, S) \propto \begin{cases} 
\gamma + \exp \left( -\frac{(d_{n} - \mu(o^{(1)}_n))^{2}}{2\sigma^{2}} \right) & \text{if } |d_{n} - \mu| < 3\sigma \sqrt{\text{var}(S, o^{(1)}_n)}, \\
0 & \text{otherwise}
\end{cases}
\]

where \(\mu(S, o_{n}^{(1)})\) is a mean function that links the support points and the observations and \(N_k\) is the set of the disparities of its surrounding support points, which are contained in a small 20 × 20 neighborhood of \(o_{n}^{(1)}\). Specifically, for the r th triangulation, the mean disparity value of an observation \(o_{n}^{(1)}\) in this triangulation can be calculated with the following formula:

\[
\mu_{t}(o^{(1)}_n) = a_t u_n + b_t v_n + c_t.
\]

The conditional probability of \(o_{n}^{(2)}\) given \(o_{n}^{(1)}\) and \(d_{n}\) is formulated as
\( p(\alpha_n^{(r)}|\alpha_n^{(l)}, d_n) \propto \begin{cases} 
\exp(-\beta \|\mathbf{f}_n^{(l)} - \mathbf{f}_n^{(r)}\|_1) & \text{if } \left( \frac{u_n^{(l)}}{v_n^{(l)}} \right) = \left( \frac{u_n^{(r)} + d_n}{v_n^{(r)}} \right), \\
0 & \text{otherwise} \end{cases} \) \tag{5}

where \( \mathbf{f}_n^{(l)} \) and \( \mathbf{f}_n^{(r)} \) are the feature vectors of \( \alpha_n^{(l)} \) and \( \alpha_n^{(r)} \), respectively, and \( \beta \) is a constant. Finally, taking the negative logarithm yields an energy function that can be easily minimized only if \( |d - \mu| < 3\sigma \) or if \( d \) is an element of the neighboring support point disparities:

\[
E(d) = \beta \|\mathbf{f}_n^{(l)} - \mathbf{f}_n^{(r)}(d)\|_1 - \log \left[ \gamma + \exp \left( -\frac{|d - \mu(S, \alpha_n^{(l)})|^2}{2\sigma^2} \right) \right]. \tag{6}
\]

The algorithm is pixel-wise and can normally be done in parallel because the disparity value of any pixel that requires estimation in an image plane depends only on the support points around it and its corresponding triangular mesh, which is constructed based on the locations and disparity values of the support points and remains unchanged once the support points are given.

ELAS can reduce the disparity search space by considering the already-constructed triangular mesh and its neighboring disparities, compute accurate disparity maps of high-resolution images at frame rates close to real time, and decrease the stereo matching ambiguities by introducing a prior distribution estimated from robust support points. A high level of performance is obtained in the presence of poorly textured and slanted surfaces, but its ability to capture disparity discontinuities is not satisfactory and can be further improved. To address this problem, in our paper, each pixel whose disparity is to be estimated is associated with a confidence value, which gives a determination for updating its disparity. In addition, the set of support points is adjusted to produce a better result. Without the adjustment of support points, the disparity value assigned by ELA cannot be changed. A detailed description of our improved method is given in Section 3.

3. Confidence-based iterative ELAS

This section describes the proposed stereo matching method. Given the reference and target images, disparity maps are estimated with CI_ELAS. Without loss of generality, in the following section, we consider the left image as the reference image.

3.1. Confidence calculation and updating

To begin, we must give the definition of confidence for each pixel. Confidence is defined based on the ratio between the second-best match cost and the best match cost obtained by an arbitrary disparity estimation method using the Peak Ratio (PKR) rule (Egnal, Mintz, & Wildes, 2004; Hu & Mordohai, 2010):

\[
C_{\text{PKR}}(x, y) = \frac{\min_{d \neq D(x,y)} \text{Cost}(x, y, d)}{\min_d \text{Cost}(x, y, d)}, \tag{7}
\]

\( \text{Cost}(x, y, d) \) is a cost value obtained by assigning a disparity hypothesis \( d \) for a pixel \((x, y)\). \( D(x, y) \) indicates the disparity value at \((x, y)\) with the minimum matching cost. To limit the confidence value within the range from 0 to 1, the following formula is adopted in this paper instead:

\[
\text{Conf}(x, y) = 1 - \frac{\min_d \text{Cost}(x, y, d)}{\min_{d \neq D(x,y)} \text{Cost}(x, y, d)}. \tag{8}
\]

Inspired by the process of obtaining the support points of ELAS, we can define the confidence value using the same match criterion, and the primary confidence value for each pixel can then be calculated using Equation (8). The confidence updating rule is described in Section 3.2.
3.2. Formulation of the proposed method

For clarity, we use the same notation as ELAS; in addition, \(d_n^{(i)}\), \(cf_n^{(i)}\), and \(S_i\) indicate the disparity values, confidence values, and the set of support points in the \(i\)th iteration, respectively. Assuming that known information \((S_i, d_n^{(i)}, cf_n^{(i)})\) and observations \((o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)})\) are conditionally independent given the disparity \(d_n^{(i+1)}\) of \(o_n^{(i)}\), and \(S_i\) is independent of \((d_n^{(i)}, cf_n^{(i)})\). Based on ELAS, with the introduction of confidence values, we obtain a new posterior of \(d_n^{(i+1)}\), whose maximum is equivalent to that of the product of three probabilistic components:

\[
P(d_n^{(i+1)}|o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}, S_i, d_n^{(i)}, cf_n^{(i)}) \propto P(d_n^{(i+1)}|o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}, S_i, d_n^{(i)}, cf_n^{(i)})
\]

\[
= P(o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}, S_i, d_n^{(i)}, cf_n^{(i)}|d_n^{(i+1)}) P(d_n^{(i+1)}) P(d_n^{(i)}|o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}, S_i, d_n^{(i)}, cf_n^{(i)})
\]

\[
\propto P(d_n^{(i+1)}|o_n^{(i)}, S_i) P(o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}|d_n^{(i)}|d_n^{(i+1)}) P(d_n^{(i+1)}) P(d_n^{(i)}|o_n^{(i)}, S_i, d_n^{(i)}, cf_n^{(i)})
\]  

(9)

the fourth line in (9) is derived from the assumption that \(o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}, S_i\) and \(d_n^{(i)}, cf_n^{(i)}\) are independent, and the third component of the last line is a likelihood term for updating disparity; we model this term as

\[
P(d_n^{(i)}|o_n^{(i)}, S_i) P(o_n^{(i)}, o_1^{(i)}, o_2^{(i)}, \ldots, o_N^{(i)}|d_n^{(i)}|d_n^{(i+1)})
\]

\[
= \exp \{-\Psi(d_n^{(i)}, d_n^{(i+1)}, cf_n^{(i)})\},
\]

(10)

where function \(\Psi(d_n^{(i)}, d_n^{(i+1)}, cf_n^{(i)})\) penalizes the divergence of disparity assignment \(d_n^{(i+1)}\) from the previous one \(d_n^{(i)}\). Compared with ELAS, because confidence values are introduced, disparities are calculated based on ELAS and are updated according to the constraint of the penalty function \(\Psi\) in each iteration.

For the construction of function \(\Psi\), we use the robust penalty function derived from the total variation model (Rudin, Osher, & Fatemi, 1992; Wang & Yang, 2011).

\[
\Psi(x, y) = -\ln \left( (1 - \eta) \exp \left( \frac{|x - y|}{\gamma_d} \right) + \eta \right),
\]

(11)

where \(\eta\) and \(\gamma_d\) are two parameters of the total variation model that control the sharpness and upper-bound of \(\Psi\), respectively, for our likelihood term. To make it more reasonable, we substitute \(\gamma_d\) with \(cf\). \(\Psi\) is then defined as

\[
\Psi(d_n^{(i)}, d_n^{(i+1)}, cf_n^{(i)}) = -\ln \left( (1 - \eta) \exp \left( -|d_n^{(i)} - d_n^{(i+1)}| \cdot cf_n^{(i)} + \eta \right) \right).
\]

(12)

To gain some intuition for Equation (12), we set \(\eta\) to zero. Then,

\[
\Psi(d_n^{(i)}, d_n^{(i+1)}, cf_n^{(i)}) = |d_n^{(i)} - d_n^{(i+1)}| \cdot cf_n^{(i)}.
\]

(13)

which penalizes the new disparity value of a pixel that deviates from a previous one according to its confidence value. If the value of confidence is sufficiently high, the difference between \(d_n^{(i)}\) and \(d_n^{(i+1)}\) should not be excessive; otherwise, the penalty will also be large, which apparently is not the correct direction for disparity updating. If the value of confidence is low, a disparity assignment that diverges from the current one is allowable and acceptable because the current one (with a low confidence value) is not sufficiently reliable, and a larger change is definitely needed.

Above all, the new disparity value is updated by minimizing the following energy function:
\[
E(d^{(i+1)}) = \beta \left\| \nabla d^{(i)} - \nabla d^{(i+1)} \right\|_1 - \log \left( \gamma + \exp \left( -\frac{(d_n^{(i+1)} - d_n^{(i)})^2}{2\sigma^2} \right) \right) \\
- \alpha \ln \left( 1 - \eta \exp \left( -\left| d_n^{(i)} - d_n^{(i+1)} \right| \cdot cf_n^{(i)} \right) + \eta \right). \tag{14}
\]

Parameter \( \alpha \) is added to control the strength of the confidence-based term, which is the third term of Equation (14). The minimization process is iterative and is the same as ELAS for each iteration under another energy function. Note that the set of support points and the confidence value of each unstable point must be updated in an iterative manner. Support points \( S^{(0)} \) (\( i \geq 1 \)), which are used to construct a new triangulation, are those selectively obtained in the preliminary and previous steps with a high confidence level. \( S^{(0)} \) obtained at the beginning of the first iteration is the same as with ELAS and is applied to obtain the primary disparity \( d^{(0)} \). Other stable points with high confidence levels are kept unchanged during the succeeding iterations.

When one iteration is finished, every point’s confidence level, with the exception of the stable ones, must be updated according to the confidence updating formula in Equation (15), which is defined based on not only its previous and current disparity, but also its corresponding point’s current disparity (Shi, Wang, Yin, & Pei, 2015):

\[
cf_n^{(i+1)} = cf_n^{(i)} + \Delta cf_n^{(Update)} + \Delta cf_n^{(LRC)}, \tag{15}
\]

where

\[
\Delta cf_n^{(Update)} = \begin{cases} 
\lambda \cdot \text{Conf} & \text{if } \left| d_{n,j}^{(i+1)} - d_{n,j}^{(i)} \right| \leq T^{(Update)}_D \\
\text{Conf} & \text{otherwise}
\end{cases}, \tag{16}
\]

\[
\Delta cf_n^{(LRC)} = \begin{cases} 
-\text{Conf} & \text{if } \left| d_{n,j}^{(i+1)} - d_{n,z}^{(i+1)} \right| > T^{(LRC)}_D \\
0 & \text{otherwise}
\end{cases}, \tag{17}
\]

when taking the left image as the reference image. Specifically, the confidence updating process obeys three rules: (1) the updated confidence level should be based on the current level; (2) updating will increase the confidence level if the newly updated disparity is close to the previous disparity in the last iteration, and reduce the confidence level otherwise; and (3) the LRC operation on disparities gives an indication of whether to decrease the confidence. \( \text{Conf} \) is the confidence increment, and \( \lambda \) is the positive factor. \( d_{n,j}^{(i+1)} \) and \( d_{n,z}^{(i+1)} \) represent disparity values in the left image, and \( d_{n,z}^{(i+1)} \) represents disparity values in the right image. Two parameters, \( T^{(Update)}_D \) and \( T^{(LRC)}_D \), indicate the maximum tolerated discrepancy between the last and current disparity values and the left-right consistency check threshold, respectively. \( \text{Conf}, T^{(Update)}_D \), and \( T^{(LRC)}_D \) are set to 0.25, 2, and 2, respectively, in our experiments. The parameter settings are discussed specifically in Section 3.3.

### 3.3. Updating of support points

Updating of support points is the kernel part of CI ELAS. Based on the fundamental and theoretical description in Sections 3.1 and 3.2, the pipeline of updating support points is illustrated in Figure 1.

#### 3.3.1. Initialization

First, \( S^{(0)} \) is obtained in the same manner as ELAS. Among the support points in \( S^{(0)} \), the value of a pixel with high confidence is set to 1.1 in \( CF^{(i)} \), and that of other pixels are set to 0 in \( CF^{(i)} \). Here, the value 1.1 functions as an indicator of a stable point regardless of the range of confidence. In a similar manner, \( D_D^{(0)} \) and \( D_S^{(0)} \) are initialized as one negative value besides stable points in \( S^{(0)} \). Once \( (S^{(0)}, D_D^{(0)}, CF^{(0)}) \) is calculated, disparity maps \( D(i+1) \) can be derived from Equation (14), and \( D_S^{(i+1)} \) is calculated in the same way. For clarity, \( CF^{(i)} \) is the set of confidence maps of the reference image and the target image in the \( i \)th iteration.
3.3.2 Stable points set
A stable point meets three conditions: (1) its confidence value is greater than a predefined threshold (set to 0.95 in our experiments); (2) its correspondence’s confidence exceeds this threshold; and (3) the left-right consistency meets with a threshold of 2.

3.3.3. Remove redundant points
Not every stable point must be a support point; an excessive density of support points will result in the unsatisfactory construction of a triangular mesh. It is thus necessary to remove redundant points when selecting the support points from stable points. The principle behind this step is that a stale point is viewed as a redundant point once more than one support point in a window is centered at this pixel, and the size of the window is the step-size mentioned in Section 2. The entire process is halted once the maximum number of iterations is achieved or the number of unstable points shows no obvious change.

4. Experiments

4.1. Parameter settings
The parameters of CI_ELAS contained in the energy function are mainly composed of two parts:

(1) One from ELAS: $\beta = 0.03$, $\sigma = 3$, $\gamma = 15$, and $\tau = 0.9$ are suggested as the best values for good empirical performance, and we keep them unchanged for the evaluation of all image pairs, where $\tau$ is used to find support points in the preliminary step (Geiger et al., 2011) mentioned in Section 2; and

(2) Another from the confidence-based term in Equation (14) and the confidence update formula: $\eta$, $\alpha$, $\lambda$, and $\text{Conf}$. To find an appropriate setting for the three parameters, given an image pair Cones, we evaluated the proposed algorithm for all possible combinations of $\eta$, $\alpha$, $\lambda$, and $\text{Conf}$, with $\eta = 0.005, ..., 0.05$ and $\eta$ incremented in steps of 0.005 and including an additional value 0.001; the reciprocal of $\alpha$ ranges from 0.02 to 1 at intervals of 0.02 and an additional value 0.01, and
\( \lambda \) from 1 to 10 at intervals of 0.5. Conf defined in the confidence update formula is set to a constant of 0.25, and a pixel becomes stable when the confidence value exceeds 0.95. The three parameters \( \eta \), \( \alpha \), and \( \lambda \) are analyzed in the next paragraph.

We first set \( \eta \) to two widely different values, 0.03 and 0.001, and obtain Figure 2(a) and (b). Visual comparison of Figure 2(a) and (b) shows that the proposed method’s performance is slightly influenced by \( \eta \).

Figure 2. Performance under different parameter settings.

Notes: The vertical axis represents an error greater than 2%. Parts a to f show the behaviors of the proposed algorithm CI-ELAS by setting \( \eta = 0.03, \eta = 0.001, \lambda = 1.5, \lambda = 8, \alpha = 2.5, \) and \( \alpha = 1 \), respectively. Best viewed in color.
By setting $\lambda$ to 1.5 and 8, we reach a similar conclusion on $\lambda$. In Figure 2(e) and (f), two close values of $\alpha$ (2.5 and 1) lead to some change. The parameter $\alpha$, which is a sensitive one for the proposed algorithm, actually controls the strength of the confidence-based term in Equation (14), which plays a vital role in CI_ELAS; it is not surprising that a small fluctuation in $\alpha$ results in an obvious variation in the error rate caused by the strength. For Cones, the best behavior of CI_ELAS is found when $\eta = 0.03$, $\alpha = 2.5$, and $\lambda = 1.5$ among all kinds of parameter settings. Therefore, throughout our experiments, we use the same parameter settings, which simultaneously perform well for most image pairs compared with ELAS.

The percentage of non-converged pixels and the running time at each iteration are shown in Figure 3, and the corresponding experimental image is taken from the KITTI data-set with a resolution of $1,226 \times 370$. Figure 3(a) shows that disparities of plenty of pixels become stable at the first three iterations, and as the iteration proceeds, the number of the converged pixel increases to a much lesser degree and tends to remain steady. After five iterations, the procedure is convergent. In Figure 3(b), the running time is spent to reconstruct the triangulation and to estimate the unstable pixels’ disparity; thanks to the dramatic decrease in the number of unstable points, fewer and fewer pixels’ disparities require estimation during the iteration, and additional time consumption is not obvious.

For efficiency, the number of iterations can be set to 4, and the corresponding running time is one time longer than ELAS by 5 s in terms of the computer’s setting (dual-core CPU E5200 running at 2.50 GHz; only one core is used in our experiments), which could be further accelerated on a more advanced computer and is actually acceptable among many stereo matching methods that spend more than 1 min when operating on large images.

In addition, as already validated by the authors of ELAS’s paper, those compared algorithms are at least two times slower than ELAS, and most of them are unable to achieve a satisfactory balance between accuracy and computational efficiency; our method, the improved ELAS based on confidence, still runs faster than those stereo matching methods, which means that the proposed method accounts for both accuracy and efficiency.

4.2. Experimental results

In this section, we evaluate our method and obtain some quantitative results for a variety of stereo images from the Middlebury data-set, including Cones, Teddy, Art, Aloe, Dolls, Baby3, Cloth3, Lamp2, and Rock2 (Scharstein & Szeliski, 2002), the 2014 Middlebury data-set (Scharstein et al., 2014), and the KITTI data-set (Geiger, Lenz, & Urtasun, 2012). It can be concluded from the experimental results that the performance of our method is competitive with that of ELAS.
In Figure 4, the three rows from top to bottom list the performance for three image pairs: Rock2, Cones, and Art. The first column depicts the left camera images, the second column is the disparity map estimated by ELAS, and the last column is that estimated by CI_ELAS. The red regions indicate erroneous areas. Table 1 gives information regarding the image size, and the percentages of error estimated pixels among the non-occluded ones are shown in Tables 2 and 3. Note that non-occluded pixels are used to compute error rates, and occluded pixels are obtained using left-right consistency (Cochran & Medioni, 1992; Fua, 1993). For highly textured objects such as those in Rock2 and Cones, ELAS and CI_ELAS achieved significantly small errors among all other image pairs, mainly due to the piecewise linear prior in ELAS. In addition, the disparities of pixels near object boundaries and in poorly textured regions become much smoother than with ELAS, as shown in the blue and green ellipses, respectively, in Figure 4. The best result improves almost 0.6% (Teddy), which is equivalent to 4,050 pixels for 0.675-megapixel images. For other image pairs, accuracy is also enhanced over that of ELAS, as shown in Tables 2 and 3.

For the Middlebury 2014 data-set, in version 3 of the Middlebury stereo evaluation code, 15 training image pairs are used to evaluate proposed stereo matching algorithms; our evaluation was therefore conducted with these images. In addition, half-size views at resolutions above 1 megapixel are adopted here. The evaluation metric is from the KITTI vision benchmark. Out-Noc is percentage of erroneous pixels in non-occluded areas, Out-All is percentage of erroneous pixels in total, Avg-Noc is average disparity error in non-occluded areas and Avg-All is average disparity error in total. Tables 4 and 5 show the evaluation results for ELAS and CI_ELAS, with error thresholds of 1 and 2, respectively. These quantitative results validate that our method outperforms ELAS on the Middlebury 2014 data-set.
For comparisons with other state-of-the-art algorithms, newly proposed methods together with classical semi-global matching (SGM) (Hirschmuller, 2008) are considered here. Specifically, line-segment-based ELAS (LS ELAS) (Ait-Jellal, Lange, Wassermann, Schilling, & Zell, 2017), dense stereo-matching using a local adaptive multi-cost approach (LAMC DSM) (Stentoumis, Grammatikopoulos, Kalisperakis, & Karras, 2014), SGM, and recursive edge-aware filters for stereo matching (REAF)
(Cigla, 2015) are introduced. LS ELAS is an extension of the ELAS algorithm that extracts edges and samples candidate support points along them. For every two consecutive valid support points, a (straight) line segment is created. The triangulation is forced to include the set of line segments (constrained Delaunay) for better preservation of the disparity discontinuity at the edges. LAMC DSM is an adaptive local stereo method. It is integrated into a hierarchical scheme that exploits adaptive windows. By integrating reverse directions and rate calculation, new approaches for REAFs are presented with comprehensive analyses via computational complexity and filter characteristics.

Figure 5. Comparisons with other matching methods on 2014 Middlebury data-set with error threshold 1.

Figure 6. Comparisons with other matching methods on 2014 Middlebury data-set with error threshold 2.

Figure 7. Results for one image pair in KITTI data-set: (a) original image, (b) ground truth map, (c) result with ELAS, and (d) our result. Black pixels in (c) and (d) indicate erroneous pixels, where the absolute disparity error is 3.

Note: Best viewed in color.
The quantitative data of the compared methods are taken from the online Middlebury evaluation system. Because Avg-Noc and Avg-All are not provided, we omit the two metrics for them. In the first columns of Tables 4 and 5, “F” represents the full-size images to be tested, “H” represents half size, and “Q” represents quarter size. The superscripts in the second and third columns of Tables 4 and 5 indicate the rank among all of the compared methods. Our method does not seem competitive with the state-of-the-art methods such as LS ELAS and SGM. However, newly proposed methods may not outperform our method either.

In addition, more detailed information can be acquired from Figures 5 and 6. CI_ELAS was not an excellent performer, but not the worst overall. On “Jadeplant,” CI_ELAS had the lowest error percentage, and it outperformed SGM on “Recycle.” Interestingly, CI_ELAS and LAMC DSM had similar trends on these 15 training images. Although the proposed method is less effective than the state-of-the-art algorithms, it provides a fresh perspective to the field of stereo matching, just like the emerging newly proposed methods such as LAMC DSM.

For the KITTI data-set, 30 randomly selected image pairs were evaluated in our experiments. This data-set contains 194 training image pairs and 195 test image pairs for evaluation of stereo matching algorithms. Figure 7 provides visual comparisons of ELAS and CI_ELAS and indicates that the bad image of ELAS is much denser than that of CI_ELAS. The data in Table 6 show that CI_ELAS showed slightly better performance than ELAS on the KITTI data-set, but the values of the four evaluation metrics were much smaller than those of the Middlebury data-set, which suggests that ELAS and CI_ELAS behave better on the KITTI data-set. To compare with other methods, SGM, Deep-Raw (Chen, Xun, Liang, Yinan, & Chang, 2015), SymST-GP (Rahla et al., 2016), and HLSC mesh (Hadfield, Lebeda, & Bowden, 2017) are considered here. Deep-Raw, trained via the Convolutional Neural Network on a large set of stereo images with ground truth disparities, is a new measure of pixel dissimilarity that outperforms the traditional matching cost. SymST-GP creates a pipeline capable of generating three-dimensional volumes in real time for high- and low-resolution images using dense stereo-based photosymmetry. HLSC mesh uses high-level cues to improve the stereo matching performance and allows standard stereo reconstruction to be unified with a wide range of classic top-down cues from urban scene understanding. Table 6 shows that CI_ELAS has the best performance among all of the methods compared except for SGM. Actually, the conclusions from the comparison with the KITTI data-set are similar to those with the Middlebury benchmark: CI_ELAS outperforms ELAS and some of the newly proposed methods, but it is still outperformed by the state-of-the-art methods.

5. Conclusions and future work
We propose an improved stereo matching method based on ELAS. ELAS exhibits excellent performance in the presence of poorly textured and slanted surfaces owing to the novel design of piecewise linear prior, but its ability to capture disparity discontinuities is not satisfactory. CI_ELAS has reserved the advantage and improved the defect in some ways. By introducing the concept of confidence, we associate each point with a changeable confidence level and successfully update the disparity and confidence levels in an iterative manner. In addition, support points are selectively

| Method    | Out-Noc | Out-All | Avg-Noc | Avg-All |
|-----------|---------|---------|---------|---------|
| ELAS      | 8.14%   | 9.31%   | 1.51 px | 1.67 px |
| CI_ELAS   | 8.09%   | 9.22%   | 1.51 px | 1.65 px |
| SGM       | 5.76%   | 7.00%   | 1.2 px  | 1.3 px  |
| Deep-Raw  | 8.93%   | 11.07%  | 3.9 px  | 4.9 px  |
| SymST-GP  | 9.79%   | 11.66%  | 2.5 px  | 3.3 px  |
| HLSC mesh | 11.22%  | 12.82%  | 2.3 px  | 2.9 px  |
obtained at each iteration from these stable points with sufficiently high confidence, which naturally leads to a more accurate result for disparity estimation. The performance of the proposed algorithm can still be improved. In future studies, we plan to assign a weight to each support point to make an appropriate estimation of the pixels that surround them to further improve the model’s overall behavior.

Funding
This research was supported by National Natural Science Foundation of China (NSFC) [grant number 41671452], [grant number 41701532]; China Postdoctoral Science Foundation funded project [grant number 2017M612510]; and Fundamental Research Funds for the Central Universities [grant number 2042016kf0012].

Author details
Hao Ma1
E-mail: 904292234@qq.com
Shunyi Zheng1,2
E-mail: syzheng@whu.edu.cn
Yingsong Li3
E-mail: yingsong-li@whu.edu.cn
Li Gui4
E-mail: whuguili@whu.edu.cn
ORCID ID: http://orcid.org/0000-0002-1357-6338
Rongyong Huang1
E-mail: rongyonghuang@gxu.edu.cn
Haitao Wei5
E-mail: zzu_wei@whu.edu.cn
1 School of Remote Sensing, Wuhan University, No. 129 Luoyu Road, Wuhan 430079, PR China.
2 Collaborative Innovation Center of Geospatial Technology, No. 129 Luoyu Road, Wuhan 430079, PR China.
3 School of Electronic Information, Wuhan University, No. 129 Luoyu Road, Wuhan 430079, PR China.
4 Coral Reef Research Centre of China, Guangxi University, Nanning 530000, PR China.
5 School of Marine Sciences, Guangxi University, Nanning 530000, PR China.
6 Institute of Smart City, Zhengzhou University, Zhengzhou 450000, PR China.

Citation information
Cite this article as: Confidence-based iterative efficient large-scale stereo matching, Hao Ma, Shunyi Zheng, Yingsong Li, Li Gui, Rongyong Huang & Haitao Wei, Cogent Engineering (2018), 5: 1427676.

References
Alt-Jellal, R., Lange, M., Wassermann, B., Schilling, A., & Zell, A. (2017). LS-ELAS: Line segment based efficient large scale stereo matching. Paper presented at the International Conference on Robotics and Automation (ICRA), Singapore.
Boykov, Y., Veksler, O., & Zabih, R. (1998). Markov random fields with efficient approximations. Paper presented at the 1998 IEEE computer society conference on Computer vision and pattern recognition.
Chen, Z., Xun, S., Liang, W., Yinon, Y., & Chang, H. (2015). A deep visual correspondence embedding model for stereo matching costs. Paper presented at the IEEE International Conference on Computer Vision.
Cheng, L., & Caelli, T. (2007). Bayesian stereo matching. Computer Vision and Image Understanding, 106(1), 85–96. https://doi.org/10.1016/j.cviu.2005.08.009
Cigla, C. (2015). Recursive edge-aware filters for stereo matching. Paper presented at the Computer Vision and Pattern Recognition, Embedded Vision Workshop.
Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., … Rother, C. (2008). A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6), 1068–1080. https://doi.org/10.1109/TPAMI.2007.70844

Taniai, T., Matsushita, Y., & Noemura, T. (2014). Graph cut based continuous stereo matching using locally shared labels. Paper presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH.

Tappen, M. F., & Freeman, W. T. (2003). Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. Paper presented at the IEEE International Conference Proceedings on Computer Vision, Nice.

Tombori, F., Mattoccia, S., & Di Stefano, L. (2007). Segmentation-based adaptive support for accurate stereo correspondence. In D. Mery & L. Rueda (Eds.), Advances in image and video technology (pp. 427–438). Berlin: Springer. https://doi.org/10.1007/978-3-540-77129-6

Wang, L., & Yang, R. (2011). Global stereo matching leveraged by sparse ground control points. Paper presented at the IEEE Conference on Computer Vision & Pattern Recognition, Colorado Springs, CO.

Woodford, O., Torr, P., Reid, I., & Fitzgibbon, A. (2009). Global stereo reconstruction under second-order smoothness priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12), 2115–2128. https://doi.org/10.1109/TPAMI.2009.131