

A Robust Stereo Vision with Confidence Measure Based on Tree Agreement

Jeongmok Ha^(✉) and Hong Jeong

Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Cheongam-ro 77, Pohang, South Korea
{jmokha,hjeong}@postech.ac.kr

Abstract. We present an improved non-learning-based confidence measure based on tree agreement for the stereo matching problem. To use confidence information for accurate matching, we propose an improved method to assemble a cost aggregation table with confidence measure. The proposed confidence measure and cost aggregation method were evaluated using the KITTI and HCI datasets. Compared to other non-learning-based confidence measures, the proposed confidence measure showed the best ability to detect wrongly-estimated pixels. By using estimated confidence measure, we showed that the proposed cost aggregation method improved disparity map quality compared to previous methods. The proposed algorithm could estimate disparity relatively accurately even in some very challenging outdoor scenes.

Keywords: Markov random fields · Automotive stereo vision · Confidence measure · Tree agreement · Cost aggregation table

1 Introduction

The stereo vision problem is usually solved by an approximate inference on a *Markov random fields* (MRFs) model [1]. An MRF model may have difficulty matching pixels when the data are corrupted or the smoothness constraints do not satisfy the assumed conditions. Especially, numerous factors can cause mismatches in outdoor scenes, which are important in automotive vision systems; occluded areas [2], textureless areas [3], and repetitive patterns are examples [4]. Although very effective stereo algorithms have been proposed [5,6], these problems are still difficult to solve.

To overcome this limitation, several authors have proposed methods to evaluate the accuracy with which stereo image pairs are matched; i.e., the *confidence* [4]. The confidence measure represents the accuracy with which labels on each pixels are estimated without any ground truth data. Various confidence measures have been proposed [7,8]; they can be categorized into, non-learning-based and learning-based confidence measures.

Non-learning-based confidence measures use the shape of a cost curve, consistency between left and right disparities, and other local information as clues to

find mismatches. Among these methods, *left and right consistency check* (LRC) has been widely used for disparity refinement [9]. Generally, non-learning-based confidence measures are easy to implement, but are not accurate enough to be used. LRC cannot not detect textureless areas, especially in outdoor scenes.

Learning-based confidence measures find the best combination of non-learning-based confidence measures by training them. After various confidence measures have been determined, the best combination of them is obtained. Usually, ensemble learning methods (e.g., random forest [10] and regression forest [11]) are used for the training process. Learning-based methods detect mismatched disparity more accurately than do non-learning-based confidence measures. However, to use learning algorithms for various confidence measures, all confidence measures must be obtained before the training process. This is a troublesome task.

An alternative approach is to use a tree-shaped MRF, and to develop a confidence measurement based on tree agreement [12]. The method is based on the assumption that the full graph is divided into subgraphs, and exploits the property that the convex sum of energy optimized on each subgraph would be the same as the energy optimized on the full graph. The confidence measure is defined as the difference between full graph energy and the convex sum of subgraph energies. This confidence measure works on a *union jack* tree structure with *semi-global matching* (SGM) [13], which is a multidirectional 1D cost-aggregation stereo-matching algorithm.

In this paper, we propose a confidence measure that is based on tree agreement and that uses *cost aggregation table* (CAT) [14–16]. The CAT is a 2D cost aggregation method which is an improved version of SGM; CAT estimates each disparity by aggregating all costs in the image in a short time. Because CAT optimizes the MRF problem by dividing the full graph into four subgraphs, the tree agreement property can be applied easily. To use a confidence measure for stereo matching, we also propose an improved cost aggregation method. The CAT algorithm is used because its pixel selection can be effectively applied to the optimization process.

The contributions of this paper are summarized as follows:

- A tree agreement based confidence measure on four subgraphs is proposed.
- A new cost aggregation method with confidence measure is proposed.
- The proposed confidence measure and cost aggregation method were evaluated with previous methods using the KITTI [17] and HCI [18] datasets.

2 Related Work

Several good survey papers present and evaluate various kinds of confidence measures. One of them presents excellent evaluation of 17 commonly-used confidence measures, and categorizes them well based on the confidence properties [7]. Another paper analyzes various confidence measures on real road scene using the KITTI dataset [8].

Some papers tried to improve disparity results by identifying mismatched or well-matched pixels [19]. A non-parametric scheme was used to learn the probability that errors will match by classifying matches into three cases: correct, nearby foreground, or other wrong depth [20]. To propagate information in the valid pixels into invalid pixels by adaptive filtering, an asymmetric consistency check has been proposed [21]. A credibility map was developed to fuse a high-resolution image with a low-resolution depth map from a *time-of-flight* (ToF) sensor [22]. An algorithm was presented that propagates confidence information by Bayesian approach [23], and the concept of Stixel that extend pixel-level representation into medium-level representation was proposed.

Many approaches have been used to apply machine learning techniques to improve disparity estimation. A combination of learning algorithm and feature selection was proposed to construct a reliable feature set for effective matching process [24]. The perceptron concept that uses prior knowledge in the image has been used for stereo matching [25]. Stability property related to confidence measure have been used to identify subsets unambiguously [26]. A statistical background model for image blocks learned from the image itself has been proposed [27].

Recently, ensemble learning based confidence measures have been proposed. The random forest supervised learning technique has been used to improve depth maps obtained from ToF cameras by improving per-pixel confidence measure with real-world data [28]. A random decision forest framework was used to train confidence measures, and the importance of each confidence measure was ranked by the metric of Gini and permutation [10]. Another learning-based confidence measure [29] differs from previous methods in that it presented how to leverage estimated confidence to increase the accuracy of the disparity estimate by using detected ground control points on an MRF framework. To select an effective confidence measures, the regression forest was used for the training phase [11]; a confidence-based matching cost modulation was also proposed.

Some efforts have been dedicated to modulating cost functions that use confidence measures. An operator was proposed to improve matching reliability by modulating matching cost using confidence measure [30]. A self-aware matching measure [31] was proposed by measuring the correlation coefficient between a cross-matching curve and self-matching curve; this method does not need any additional parameters, and it can be applied without modification to other stereo matching algorithms.

3 Problem Statement

MRFs have been widely used to solve stereo problems. The four-neighbor-connected graph is commonly used because of its simplicity and convenience [1]. A four-neighbor-connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set \mathcal{V} of pixels, and a set \mathcal{E} of edges. Each pixel $\mathbf{p} \in \mathcal{V}$ is a random variable that can have a discrete disparity $d_{\mathbf{p}} \in \mathcal{D}$, where $d_{\mathbf{p}}$ is a disparity of pixel \mathbf{p} and \mathcal{D} is a set of possible disparities. An edge is a pair of two pixels $(\mathbf{p}, \mathbf{q}) \in \mathcal{E}$, where $\mathbf{q} \in \mathcal{V}$; the edge encodes the disparity relationship between the pixels.

The stereo problem is solved by *maximum a posteriori* (MAP) estimation on joint distribution: $p(\mathbf{d}|\mathbf{o}) = p(\mathbf{o}|\mathbf{d})p(\mathbf{d})$, where $p(\cdot)$ represents a probability distribution, \mathbf{d} is a set of all disparities in \mathcal{V} , and \mathbf{o} is a set of observation data. The Gibbs distribution [32] can be used to change the *a posteriori* distribution to an energy function. The likelihood and prior terms become data and smoothness terms of the energy function:

$$E(\mathbf{d}) = \sum_{\mathbf{p} \in \mathcal{V}} \phi_{\mathbf{p}}(d_{\mathbf{p}}) + \sum_{(\mathbf{p}, \mathbf{q}) \in \mathcal{E}} \psi_{(\mathbf{p}, \mathbf{q})}(d_{\mathbf{p}}, d_{\mathbf{q}}), \quad (1)$$

where $E(\cdot)$ represents energy function, a unary term $\phi_{\mathbf{p}}(\cdot)$ is a data term on the pixel \mathbf{p} , and the pairwise term $\psi_{(\mathbf{p}, \mathbf{q})}(\cdot, \cdot)$ represents the smoothness between pixels \mathbf{p} and \mathbf{q} . Eq. (1) is the standard energy formulation of MRF models.

By minimizing this energy term, the a set of optimum disparity solution \mathbf{d}^* is estimated:

$$\mathbf{d}^* = \arg \min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} E(\mathbf{d}). \quad (2)$$

The goal of the stereo problem is to find disparities that minimize the energy.

4 Confidence Measure with Tree Agreement

In this paper, we use a confidence measure related to *tree agreement* [12] to determine whether or not the disparity of each pixel is well estimated.

The confidence measure related to tree agreement was determined based on *tree-reweighted message passing* (TRW) [33], an representative MAP inference algorithm to find optimum solutions of general labeling problems. TRW optimizes energy function iteratively on each subgraph separately by decomposing a full graph into subgraphs. Then TRW re-parameterizes the estimated results to make the same energy function.

TRW divides a full graph \mathcal{G} into a convex sum of subgraphs $\mathcal{G}^t = (\mathcal{V}^t, \mathcal{E}^t)$, where $t \in \mathcal{T}$ and $\mathcal{T} = \{0, \dots, T-1\}$ is a set of subgraphs. Each subgraph has its weight ω^t so that $\sum_{t \in \mathcal{T}} \omega^t = 1$. The energy of full graph $E^{\mathcal{G}}(\cdot)$ is weighted convex sum of energy of subgraphs:

$$E^{\mathcal{G}}(\mathbf{d}) = \sum_{t \in \mathcal{T}} \omega^t E^t(\mathbf{d}), \quad (3)$$

where $E^t(\cdot)$ represents the energy of the subgraph t . The optimum result of stereo problem is obtained by finding the minimum energy:

$$\min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} E^{\mathcal{G}}(\mathbf{d}) = \min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} \sum_{t \in \mathcal{T}} \omega^t E^t(\mathbf{d}). \quad (4)$$

The optimal disparity solution of the stereo problem is found by identifying the disparity set that minimizes the energy function. The minimum energy function

has a lower bound because the full graph is divided into subgraphs. On the subgraphs, the estimate on each subgraph is also a solution of the stereo problem.

In the ideal case (in which all disparity estimates are correct), the estimated results on a full graph are the same as the results on the local subgraphs. The full graph consists of a convex sum of subgraphs, so the energy optimized on the full graph must be the same as the sum of energies optimized on each subgraphs. However, in most cases, the energy on the full graph is larger than the sum of energies on subgraphs; the difference can be defined as the confidence of each pixel.

Because the minimum of the sum is always greater or equal to the sum of minima, the minimum energy of the full graph is always greater or equal to the sum of minimum energies of the subgraphs:

$$\min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} \sum_{t \in \mathcal{T}} \omega^t E^t(\mathbf{d}) \geq \sum_{t \in \mathcal{T}} \min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} \omega^t E^t(\mathbf{d}). \quad (5)$$

The inequality becomes an equality when the optimum disparities of each subgraph are the same as the optimum disparity of the full graph; i.e., in the ideal case in which the disparity of each pixel is well estimated.

In this manner, the difference between left and right hand sides of (5) can be a measurement of the confidence in the estimate; the confidence is defined as the difference between the energy of the full graph and the sum of energies of the subgraphs:

$$\gamma = \min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} \sum_{t \in \mathcal{T}} \omega^t E^t(\mathbf{d}) - \sum_{t \in \mathcal{T}} \min_{\mathbf{d} \in \mathcal{D}^{|\mathcal{V}|}} \omega^t E^t(\mathbf{d}), \quad (6)$$

where γ is an uncertainty measure. Because bigger γ increases with the size of the gap between global minimum energy and sum of local minimum energies, γ is actually an uncertainty measure rather than a confidence measure.

5 Proposed Method with Confidence Measure

5.1 Cost Aggregation Table

CAT is an efficient algorithm or structure for the stereo matching problem [14]. CAT divides the full MRF graph into four subgraphs, $\mathcal{T} = (0, 1, 2, 3)$, i.e., north-east, north-west, south-east, and south-west respectively, then initializes the weight of each subgraph as $\omega^t = 1/4$. The subgraphs of CAT are denoted as \mathcal{G}^0 , \mathcal{G}^1 , \mathcal{G}^2 , and \mathcal{G}^3 . This structure produces approximate MAP inference results in a short time by dividing the full graph into four subgraphs. Because the full graph is divided into four subgraphs, the uncertainty measure based on tree agreement can be used.

We denote a reference pixel as $\mathbf{k} \in \mathcal{V}$; the reference pixel is a pixel whose disparity is to be estimated. The CAT structure is defined differently according to the reference pixel. The subgraphs with the reference node \mathbf{k} is denoted as $\mathcal{G}_{\mathbf{k}}^0$,

$\mathcal{G}_{\mathbf{k}}^1$, $\mathcal{G}_{\mathbf{k}}^2$, and $\mathcal{G}_{\mathbf{k}}^3$, which represent north-east, north-west, south-east, and south-west sides relative to the reference pixel \mathbf{k} , respectively.

The joint distribution of the disparity set assumes to consist of the product of distribution of disparities of each pixel and each subgraph:

$$p(\mathbf{d}|\mathbf{o}; \mathcal{G}) = \prod_{\mathbf{k} \in \mathcal{V}} p(d_{\mathbf{k}}|\mathbf{o}; \mathcal{G}) = \prod_{\mathbf{k} \in \mathcal{V}} \prod_{t \in \mathcal{T}} p(d_{\mathbf{k}}|\mathbf{o}; \mathcal{G}_{\mathbf{k}}^t). \quad (7)$$

Although each pixel is estimated independently, estimation of the optimum disparity for each pixel does not require aggregation of all costs in the image every time. With the help of the parents system, all costs can be aggregated in just four scans as follows.

For example, during the first scan, only subgraph \mathcal{G}^0 is used to aggregate costs (in the south-east direction). The pixel on the top left corner is set as the seed pixel \mathbf{s}^0 . The parent pixels $\mathcal{P}_{\mathbf{p}}^t = \{\mathbf{p}_h^t, \mathbf{p}_v^t\}$ of pixel \mathbf{p} are set as their horizontal and vertical adjacent neighbors, e.g., $\mathcal{P}_{\mathbf{p}}^0 = \{(y, x-1), (y-1, x)\}$. Then the costs are aggregated from parent pixels to child pixels. This aggregation step is performed on all four subgraphs independently.

When aggregating costs, a message-passing algorithm based on dynamic programming is used. This algorithm is similar to the cost aggregation method of SGM, but expands the 1D cost aggregation of SGM to 2D. The message is computed considering its smoothness constraint from its parent pixels. The minimum costs of two parent pixels are compared, and the smaller cost is delivered to the child pixel.

The message from parent pixels $\mathcal{P}_{\mathbf{p}}^t$ to pixel \mathbf{k} on subgraph t is defined as:

$$m_{\mathbf{p}}^t(d_{\mathbf{p}}) = \phi_{\mathbf{p}}(d_{\mathbf{p}}) + \min_{\mathbf{q} \in \mathcal{P}_{\mathbf{p}}^t} \left[\min_{d_{\mathbf{q}}} [\psi_{(\mathbf{p}, \mathbf{q})}(d_{\mathbf{p}}, d_{\mathbf{q}}) + m_{\mathbf{q}}^t(d_{\mathbf{q}})] \right], \quad (8)$$

where $m_{\mathbf{p}}^t(\cdot)$ is a message from parent pixels $\mathcal{P}_{\mathbf{p}}^t$ to pixel \mathbf{p} . The message is delivered from the parent pixels with smoothness term $\psi_{(\mathbf{p}, \mathbf{q})}(\cdot, \cdot)$ to pixel \mathbf{p} . This processing is performed iteratively from seed pixel \mathbf{s}^t to the opposite corner pixel. Usually, a census transform [34] is used as the data term, and a two-step penalty function [13] is used as the smoothness term.

After all messages have been computed on every subgraph, belief in the reference pixel \mathbf{k} is defined as the sum of all messages computed on pixel \mathbf{p} :

$$b_{\mathbf{p}}(d_{\mathbf{p}}) = \sum_{t \in \mathcal{V}} m_{\mathbf{p}}^t(d_{\mathbf{p}}), \quad (9)$$

where $b_{\mathbf{p}}(\cdot)$ is a belief on pixel \mathbf{p} . After beliefs on every pixel are computed, a winner takes all method is used to find the optimum disparity on each pixel:

$$d_{\mathbf{p}}^* = \arg \min_{d_{\mathbf{p}}} b_{\mathbf{p}}(d_{\mathbf{p}}). \quad (10)$$

Algorithms based on subgraph division can use the confidence measure based on tree agreement. With the subgraph division of CAT structure, the proposed

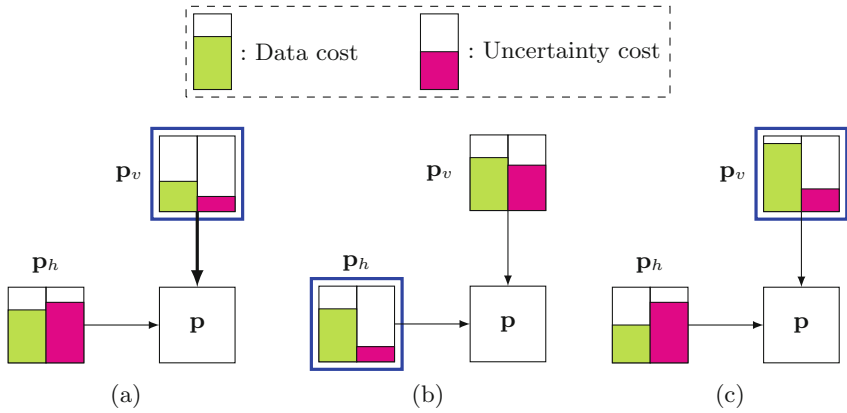


Fig. 1. Diagrams of cost aggregation step with the confidence term for north-east table. (a) Both data and uncertainty costs are lower than the other cost. (b) Uncertainty costs are compared when data costs are same. (c) Both data cost and uncertainty cost of horizontal and vertical pixels are summed and lower cost is propagated to the child pixel.

confidence measure is defined as the difference between the global minimum energy of the full graph and the sum of local minimum energies of the subgraphs. This uncertainty measure can be defined on each pixel \mathbf{p} in the CAT structure:

$$\gamma_{\mathbf{p}} = \min_{d_{\mathbf{p}} \in \mathcal{D}} \sum_{t \in \mathcal{T}} \omega^t E^t(d_{\mathbf{p}}) - \sum_{t \in \mathcal{T}} \min_{d_{\mathbf{p}} \in \mathcal{D}} \omega^t E^t(d_{\mathbf{p}}), \quad (11)$$

where $\gamma_{\mathbf{p}}$ is a uncertainty measure on pixel \mathbf{p} . After all costs in the image are aggregated in the direction of each table \mathcal{T} , the confidence of each pixel is computed.

5.2 Cost Aggregation with Confidence Term

The proposed algorithm is based on 2D cost aggregation algorithm on CAT. Because the algorithm is based on message passing, the confidence of the pixel can be propagated with the messages.

The previous CAT algorithm computes messages by finding the minimum cost between horizontal and vertical adjacent pixels. Only the smallest cost in the horizontal and vertical pixels is propagated to the direction of the subgraph. By adding a confidence term to the energy formulation, the costs of pixels for which confidence is high can be chosen while aggregating costs. When one horizontal and one vertical pixel have high confidence, the cost the pixel for which confidence is higher should be aggregated. We propose a new message that identify the pixel for which confidence is highest.

We define a new message from parent pixels to pixel \mathbf{p} by adding a confidence term to (8):

$$m_{\mathbf{p}}^t(d_{\mathbf{p}}) = \phi_{\mathbf{p}}(d_{\mathbf{p}}) + \gamma_{\mathbf{q}} + \min_{\mathbf{q} \in \mathcal{P}_{\mathbf{p}}^t} \left[\min_{d_{\mathbf{q}}} [m_{\mathbf{q}}^t(d_{\mathbf{q}}) + \psi_{\langle \mathbf{p}, \mathbf{q} \rangle}(d_{\mathbf{p}}, d_{\mathbf{q}})] \right]. \quad (12)$$

During each aggregation step, the minimum costs with respect to disparities on each parent pixel are computed, then the data term, the minimum cost of the parent pixels, and the confidence measure $\gamma_{\mathbf{p}}$ are added (Fig. 1). During the next aggregation step, two aggregated costs with the uncertainty measures are compared to each other, then the smaller of these costs is selected for message passing. If a pixel has a low confidence measure, it does not propagate its information to the child pixel because this information is likely to be wrong. The belief and optimum disparity are determined in the same way as in the standard CAT algorithm.

6 Experimental Results

In this section, we present qualitative and quantitative experimental evaluation of confidence measure and stereo accuracy. We used the KITTI dataset [17] for quantitative and qualitative evaluation, and the HCI dataset [18] only for qualitative evaluation because the dataset does not provide ground truth. Empirically determined parameters used; SGM – $P1$: 0.25, $P2$: 5, CAT and proposed algorithm – $P1$: 0.6, $P2$: 12. Our PC simulation environment was Intel i7-3770 processor 3.4 GHz with CPU, 8 GB memory, Microsoft Windows 8.1, Microsoft Visual studio 2013 C++, and OpenCV 3.0.

6.1 Confidence Measure Comparison

We used the KITTI dataset to evaluate the proposed confidence measure and the proposed algorithm. The KITTI dataset is widely used to evaluate disparity estimation algorithms because it provides various stereo image pairs of road scene scenarios; in this trial, the 194 training image pairs with ground truth were used to evaluate the confidence measure and proposed algorithm.

To evaluate the proposed confidence measure, sparsification curve and *area under curve* (AUC) were used [8]. The sparsification curve shows how *pixel error rate* (PER) is affected by removal of pixels that have the lowest confidence measure. Lower AUC decreases when the confidence measure detects mismatched pixels accurately.

Popular non-learning-based confidence measurements for stereo problem were tested [7]; *curvature* (CUR), *naive pear ratio* (PKRN), *maximum margin* (MMN), *naive winner margin* (WMNN), *left right difference* (LRD), and the proposed confidence measure. Cost curves used for the confidence measure were aggregated costs of CAT.

Table 1. The comparison tables of the SGM, CAT, and proposed algorithm on non-occlusion areas and all areas.

Window	Non-occlusion areas (%)				All areas (%)			
	Metrics	SGM	CAT	Proposed	Metrics	SGM	CAT	Proposed
3×3	PER-2px	16.03	13.75	11.42	PER-2px	17.87	15.41	12.94
	PER-3px	11.35	9.85	7.93	PER-3px	13.25	11.52	9.41
	PER-4px	8.82	7.69	6.08	PER-4px	10.75	9.36	7.51
	PER-5px	7.23	6.29	4.91	PER-5px	9.17	7.94	6.30
	APE (px)	2.07	1.81	1.56	APE (px)	2.95	2.45	2.07
5×5	PER-2px	11.58	10.06	8.91	PER-2px	13.44	11.67	10.32
	PER-3px	7.94	7.04	6.10	PER-3px	9.82	8.61	7.42
	PER-4px	6.05	5.43	4.67	PER-4px	7.93	6.97	5.91
	PER-5px	4.88	4.42	3.78	PER-5px	6.75	5.93	4.96
	APE (px)	1.60	1.46	1.33	APE (px)	2.43	2.05	1.78
7×7	PER-2px	9.99	8.76	7.96	PER-2px	11.81	10.28	9.27
	PER-3px	6.77	6.03	5.43	PER-3px	8.60	7.50	6.63
	PER-4px	5.14	4.64	4.16	PER-4px	6.94	6.06	5.27
	PER-5px	4.14	3.78	3.39	PER-5px	5.91	5.16	4.42
	APE (px)	1.45	1.34	1.26	APE (px)	2.24	1.89	1.66
9×9	PER-2px	9.27	8.17	7.57	PER-2px	11.05	9.64	8.22
	PER-3px	6.25	5.59	5.14	PER-3px	8.01	6.98	6.27
	PER-4px	4.75	4.31	3.96	PER-4px	6.48	5.65	4.99
	PER-5px	3.82	3.53	3.25	PER-5px	5.52	4.81	4.20
	APE (px)	1.39	1.29	1.24	APE (px)	2.14	1.81	1.61

In most of the sparsification curves of confidence measurements, the proposed confidence measure showed the lowest curve shape among tested confidence measures. AUC curve of all KITTI training images were depicted in the ascending order according to AUC of the proposed confidence measure (Fig. 2). LRD and other confidence measures occasionally achieved lower AUC than the proposed confidence measure, but the proposed confidence measure achieved the lowest AUC in most images.

6.2 Quantitative Evaluation on KITTI Dataset

Three cost aggregation algorithms (SGM, CAT, and the proposed algorithm) were compared to each other using KITTI dataset. Based on KITTI stereo evaluation metrics, PER with threshold 2px, 3px, 4px, and 5px and *average pixel error* (APE) were used for disparity accuracy measurement. All PERs and APEs

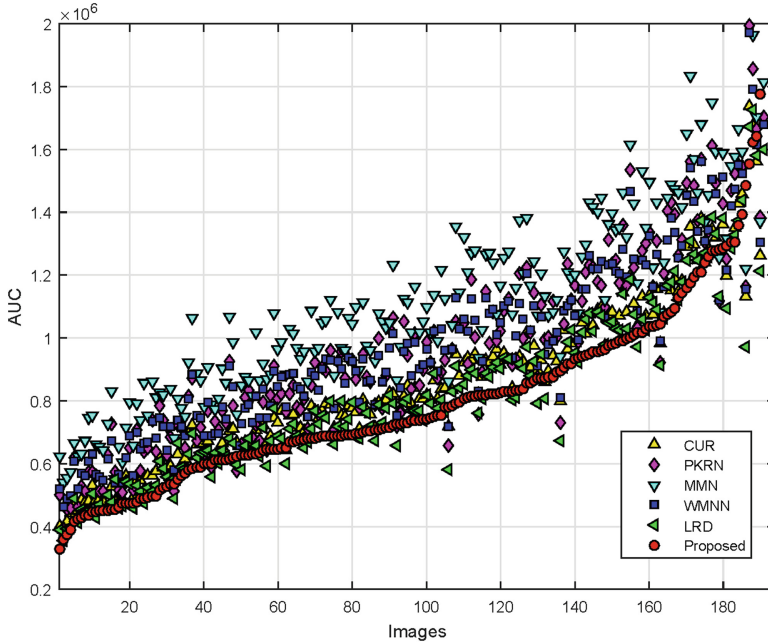


Fig. 2. AUC curve of CUR, PKRN, MMN, WMNN, LRD, and the proposed confidence measure using 194 KITTI training images.

were determined by averaging error results of 194 disparity maps. To determine the effect of window size of census transform, algorithms were operated with 3×3 , 5×5 , 7×7 , and 9×9 window sizes. No disparity refinement technique was used to evaluate the effect of confidence term.

Regardless of window size of census transform, every metric showed that the proposed algorithm was the best among the three algorithms (Table 1). The amount of the error rate reduction on all areas is higher than that on non-occlusion areas; it means that proposed method works effectively in occlusion areas. Disparity maps of CAT and the proposed algorithm showed that some high uncertainty areas on disparity map of CAT were corrected on disparity map of the proposed algorithm (Fig. 3).

6.3 Qualitative Evaluation on HCI Dataset

We used the HCI dataset to show the effect of the confidence term in very challenging stereo scenarios. The HCI dataset provides 11 different kinds of harsh environment scenarios. Because this dataset does not provide ground truth, it was only used to show that proposed algorithm gives stable disparity results

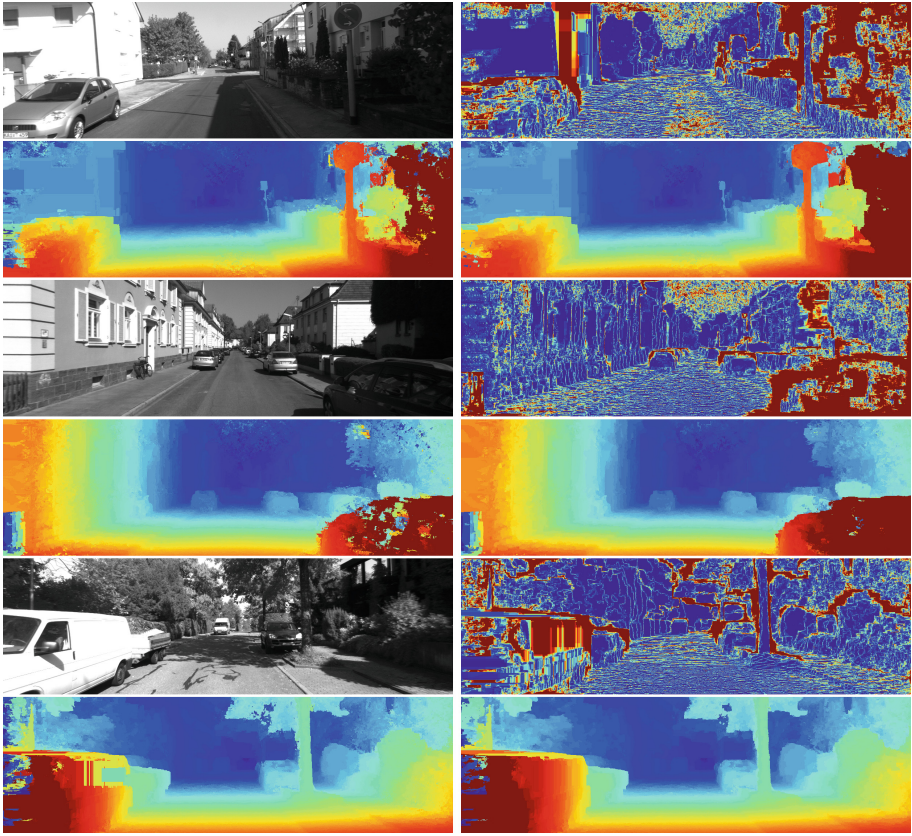


Fig. 3. Three examples of uncertainty maps, and disparity maps of the CAT and CAT with confidence term using KITTI dataset. Top left: left image. Top right: uncertainty map; red: high uncertainty; blue: low uncertainty. Bottom left: disparity maps computed by CAT; red: close to camera; blue: far from camera. Bottom right: disparity maps computed by CAT with confidence term (Color figure online).

even in some harsh environments. Especially, pixel matching in images of night and snow, rain blur, reflecting cars, and sunflare scenarios can be ambiguous and cause false stereo matching results. Disparity maps of CAT and the proposed algorithm were obtained with its confidence measures (Fig. 4). The uncertainty map shows high values in areas in which a harsh environment affects the camera. The proposed algorithm can estimate disparity relatively stable in areas that CAT estimates wrongly.

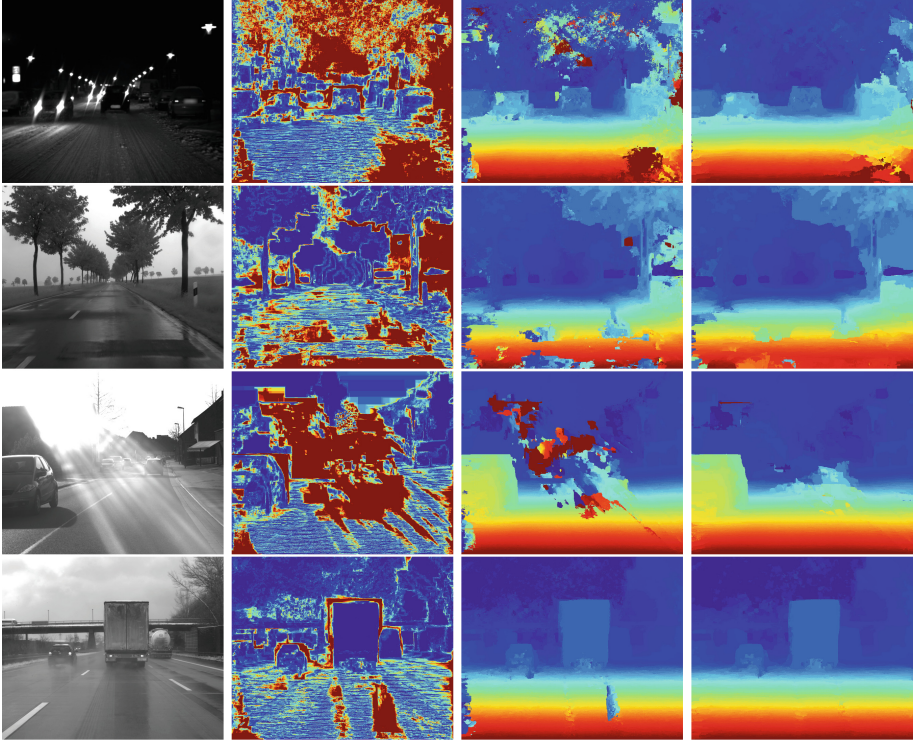


Fig. 4. Examples of uncertainty map (second column, red: high uncertainty, blue: low uncertainty), disparity map of the CAT (third column, red: close to camera, blue: far from camera), and disparity map of CAT with confidence term (fourth column) using night and snow (first row), rain blur (second row), sunflare (third row), and wet autobahn (fourth row) scenario image pairs of HCI dataset (Color figure online).

7 Conclusions

We proposed a non-learning-based confidence measure that uses the tree agreement property. The proposed confidence measure is based on the fact that convex sum of each subgraph energy must be the same as the energy of full graph if pixel is well matched. To use the new confidence measure, we proposed a new CAT-based cost aggregation method with the confidence term. It could select the cost among horizontal and vertical pixels that has low uncertainty values.

We evaluated the proposed confidence measure and cost aggregation method on the KITTI and HCI datasets. The proposed confidence measure showed lower AUC than all other non-learning-based confidence measures. The proposed cost aggregation with the confidence term achieved more accurate disparity results than did methods that do not use this term. Even on very challenging scenarios provided by the HCI dataset, the proposed combination of confidence measure and cost aggregation method showed reliable disparity results.

Acknowledgments. This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (NRF-2014H1C1A1066380).

References

1. Li, S.Z.: Markov Random Field Modeling in Image Analysis. Springer Science and Business Media, London (2009)
2. Egnal, G., Wildes, R.P.: Detecting binocular half-occlusions: empirical comparisons of five approaches. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 1127–1133 (2002)
3. Manduchi, R., Tomasi, C.: Distinctiveness maps for image matching. In: *ICIAP*, p. 26. IEEE (1999)
4. Egnal, G., Mintz, M., Wildes, R.P.: A stereo confidence metric using single view imagery with comparison to five alternative approaches. *Image Vis. Comput.* **22**, 943–957 (2004)
5. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**, 7–42 (2002)
6. Jeong, H.: *Architectures for Computer Vision: From Algorithm to Chip with Verilog*. Wiley, USA (2014)
7. Hu, X., Mordohai, P.: A quantitative evaluation of confidence measures for stereo vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2121–2133 (2012)
8. Haeusler, R., Klette, R.: Analysis of KITTI data for stereo analysis with stereo confidence measures. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) *ECCV 2012 Ws/Demos, Part II*. LNCS, vol. 7584, pp. 158–167. Springer, Heidelberg (2012)
9. Fua, P.: A parallel stereo algorithm that produces dense depth maps and preserves image features. *Mach. Vis. Appl.* **6**, 35–49 (1993)
10. Haeusler, R., Nair, R., Kondermann, D.: Ensemble learning for confidence measures in stereo vision. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 305–312. IEEE (2013)
11. Park, M.G., Yoon, K.J.: Leveraging stereo matching with learning-based confidence measures. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 101–109 (2015)
12. Drory, A., Haubold, C., Avidan, S., Hamprecht, F.A.: Semi-global matching: a principled derivation in terms of message passing. In: Jiang, X., Hornegger, J., Koch, R. (eds.) *GCPR 2014*. LNCS, vol. 8753, pp. 43–53. Springer, Heidelberg (2014)
13. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 328–341 (2008)
14. Ha, J.M., Jeon, J.Y., Bae, G.Y., Jo, S.Y., Jeong, H.: Cost aggregation table: cost aggregation method using summed area table scheme for dense stereo correspondence. In: Bebis, G., et al. (eds.) *ISVC 2014, Part I*. LNCS, vol. 8887, pp. 815–826. Springer, Heidelberg (2014)
15. Ha, J., Jeon, B., Jun, W., Lee, J., Jeong, H.: An improved 2D cost aggregation method for advanced driver assistance systems. In: *2015 IEEE Intelligent Vehicles Symposium (IV)*, pp. 89–94. IEEE (2015)
16. Ha, J., Jeon, B., Jeon, J., Jo, S.Y., Jeong, H.: Cost aggregation table: a theoretic derivation on the markov random field and its relation to message passing. In: *22th IEEE International Conference on Image Processing (ICIP)*. IEEE (2015)

17. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the KITTI vision benchmark suite. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354–3361. IEEE (2012)
18. Meister, S., Jähne, B., Kondermann, D.: Outdoor stereo camera system for the generation of real-world benchmark data sets. *Opt. Eng.* **51**, 021107-1 (2012)
19. Kostková, J., Šára, R.: Stratified dense matching for stereopsis in complex scenes. In: BMVC, vol. 5, p. 6. Citeseer (2003)
20. Kong, D., Tao, H.: A method for learning matching errors for stereo computation. In: BMVC, vol. 1, p. 2 (2004)
21. Min, D., Sohn, K.: An asymmetric post-processing for correspondence problem. *Sig. Process. Image Commun.* **25**, 130–142 (2010)
22. Garcia, F., Mirbach, B., Ottersten, B., Grandidier, F., Cuesta, A.: Pixel weighted average strategy for depth sensor data fusion. In: 17th IEEE International Conference on Image Processing (ICIP), pp. 2805–2808. IEEE (2010)
23. Pfeiffer, D., Gehrig, S., Schneider, N.: Exploiting the power of stereo confidences. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 297–304. IEEE (2013)
24. Lew, M.S., Huang, T.S., Wong, K.: Learning and feature selection in stereo matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 869–881 (1994)
25. Cruz, J., Pajares, G., Aranda, J., Vindel, J.: Stereo matching technique based on the perceptron criterion function. *Pattern Recogn. Lett.* **16**, 933–944 (1995)
26. Šára, R.: Finding the largest unambiguous component of stereo matching. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 900–914. Springer, Heidelberg (2002)
27. Sabater, N., Almansa, A., Morel, J.M.: Meaningful matches in stereovision. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 930–942 (2012)
28. Reynolds, M., Dobos, J., Peel, L., Weyrich, T., Brostow, G.J.: Capturing time-of-flight data with confidence. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 945–952. IEEE (2011)
29. Spyropoulos, A., Komodakis, N., Mordohai, P.: Learning to detect ground control points for improving the accuracy of stereo matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1621–1628. IEEE (2014)
30. Gherardi, R.: Confidence-based cost modulation for stereo matching. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4. IEEE (2008)
31. Kindermann, R., Snell, J.L., et al.: Markov Random Fields and Their Applications. American Mathematical Society, vol. 1. Providence, RI (1980)
32. Kindermann, R., Snell, J.L., et al.: Markov Random Fields and Their Applications. American Mathematical Society Providence, vol. 1. RI (1980)
33. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1568–1583 (2006)
34. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Eklundh, J.O. (ed.) Computer Vision ECCV 1994, pp. 151–158. Springer, Heidelberg (1994)