AMBIGUITY CONCEPT IN STEREO MATCHING PIPELINE

E. Sarrazin[1],[*] M. Cournet[1], L. Dumas[2], V. Defonte[2], Q. Fardet[2], Y. Steux[2], N. Jimenez Diaz[2], E. Dubois[1], D. Youssefi[1], F. Buffe[1]

[1] Centre National d'Etudes Spatiales (CNES), 18 avenue E. Belin, Toulouse, France – emmanuelle.sarrazin@cnes.fr
[2] CS, 6 rue Brindejonc des Moulinais, Toulouse, France

Commission II, WG II/4

KEY WORDS: Ambiguity, Confidence, Stereo Matching, Disparity Map Denoising, Semi Global Matching Optimization

ABSTRACT:

In a 3D reconstruction pipeline, stereo matching step aims at computing a disparity map representing the depth between image pair. The evaluation of the disparity map can be done through the estimation of a confidence metric. In this article, we propose a new confidence metric, named *ambiguity integral metric*, to assess the quality of the produced disparity map. This metric is derived from the concept of ambiguity, which characterizes the property of the cost curve profile. It aims to quantify the difficulty in identifying the correct disparity to select. The quality of ambiguity integral metric is evaluated through the ROC curve methodology and compared with other confidence measures. In regards to other measures, the ambiguity integral measure shows a good potential. We also integrate this measure through various steps of the stereo matching pipeline in order to improve the performance estimation of the disparity map. First, we include ambiguity integral measure during the Semi Global Matching optimization step. The objective is to weight, by ambiguity integral measure, the influence of points in the SGM regularization to reduce the impact of ambiguous points. Secondly, we use ambiguity as an input of a disparity refinement deep learning architecture in order to easily locate noisy area and preserve details.

1. INTRODUCTION

Among the stages that constitute a 3D reconstruction pipeline (Michel et al., 2020), the stereo-matching step is one of the most crucial steps, as it strongly impacts the computed elevation surface quality. The principle of stereo-matching step is to match homologous points between left and right images, in order to estimate a disparity map, which reflects the apparent motion between the image pair. Depending on the scene, the kind of texture or the brightness, the stereo matching is more or less challenging and correctly performed. It is therefore relevant to assess the quality of a disparity map. For instance, the 3D reconstruction downstream steps could ignore or reduce the impact of the least confident points. This disparity map assessment can be done through the computation of a confidence metric. According to (Hu and Mordohai, 2012), a confidence metric is characterized by high values for correct disparities and low values for errors. They also emphasize an important property of a good confidence: if matched pixels are sorted in descending order of confidence value, all bad pixels (mismatch, occlusion) should end up in last positions.

A review of existing confidence metrics in the literature is presented section 2. Then, the ambiguity concept and an new associated confidence metric are defined in section 3. Section 4 shows therefore the corresponding results of this new confidence metric. And finally, stereo matching steps using this metric are developed in section 5.

2. RELATED WORK

In (Hu and Mordohai, 2012), a large number of confidence metrics has been reviewed and divided into several categories based on: local properties of the cost curve, the analysis of the entire cost curve, the consistency between the left and right disparity maps, the distinctiveness-based confidence measures. This set of metrics has been completed by Poggi (Poggi et al., 2017) especially with machine learning-based metrics. In ensemble learning-based approaches, confidence metrics are estimated through random forests (Haeusler et al., 2013; Spyropoulos et al., 2014; Min-Gyu Park and Yoon, 2015; Gouveia et al., 2015). The features used for random forests correspond to a selection of confidence metrics mostly defined in (Hu and Mordohai, 2012). With the rise of deep learning, methods based on convolutional neural networks have been developed. Some of these CNN approaches focused on the disparity map learning (Poggi et al., 2017), other methods worked directly with the cost volume in order to take more information into account (Mehltretter and Heipke, 2019; Kim et al., 2019).

Before using these metrics in a stereo pipeline, it is necessary to evaluate them. As mentioned in the introduction, if matched points were sorted in confidence decreasing order, all bad points should be ranked at last. To evaluate this property, a methodology is proposed in (Hu and Mordohai, 2012). This approach is based on the computation of the Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve, which represents the error rate as a function of sorted matched points in confidence decreasing order. To compute the ROC curve, pixels are sorted in confidence decreasing order. A subset of $p\%$ first pixels is extracted and the error rate for this subset is computed. It represents the percentage of pixels whose absolute disparity difference with ground truth is greater than a threshold value. The process is reiterated until all pixels have been taken into account. From the ROC curve, the AUC can be derived. The AUC measures the ability to identify errors in disparity map with the help of a confidence metric. For a disparity map with an error rate $\epsilon \in [0, 1]$, an ideal confidence measure can reach an error rate of 0 for the first $1 - \epsilon$ pixels. (Hu and Mordohai, 2012) provides the formulation for an ideal confidence

---

measure. The ideal *ROC* is:

$$R_{opt}(p) = \begin{cases} 0 & p \leq 1 - \epsilon \\ \frac{p-(1-\epsilon)}{p} & p > 1 - \epsilon \end{cases} \qquad (1)$$

where p is the percentage of considered pixels. So, the ideal *AUC* can be formulated as:

$$AUC_{opt} = \int_{1-\epsilon}^{1} \frac{p-(1-\epsilon)}{p} dp = \epsilon + (1-\epsilon)\ln(1-\epsilon) \quad (2)$$

Using this methodology, (Hu and Mordohai, 2012) identify the following measures as good confidence measures: Naive Peak Ratio (PKRN), Naive Winner Margin (WMNN), Left/Right Difference (LRD), Attainable Maximum Likelihood (AML), Distinctive Similarity Measure (DSM) and Self-Aware Matching Measure (SAMM). It must be emphasized that the confidence metric results depend on the similarity measure used to compute cost volume. Moreover, some methods perform better near discontinuities. (Poggi et al., 2017) show that machine learning based metrics offer better results compared to classical metrics, and especially deep learning approaches. But these are very time consuming and they required a large dataset with ground truth. Another difficulty is their ability to generalize depending on the training dataset. For remote sensing images, at the present time, there are few datasets (Bosch et al., 2016; Bosch et al., 2019) and they focuses on urban areas. This can be a limitation for the usage of these methods.

3. AMBIGUITY CONCEPT

3.1 Definition

In the (Scharstein and Szeliski, 2002) taxonomy, a stereo matching pipeline is composed of the following steps:

1. matching cost computation;

2. cost (support) aggregation;

3. disparity computation / optimization; and

4. disparity refinement.

The matching cost computation stage consists in computing a matching cost measure for a given pixel of the left image and each possible disparity within the disparity range. Matching cost measures set forms the cost curve for the given pixel. All of the cost curves are gathered in the cost volume matrix. During the disparity computation step, the pixel of the right image, for which the correlation score is the highest (or the lowest depending on the matching cost measure chosen), is selected as being the homologous pixel of the pixel from the left image. The column difference between the peer points is called the disparity.

The ambiguity notion, introduced in (Hu and Mordohai, 2012) is based on the characteristics of the cost curve for a pixel (x, y). Indeed, the ambiguity aims to quantify the difficulty to identify the appropriate disparity to select. The more minima the curve has, the more difficult it is to decide that a disparity value should prevail over the others. The figure 1 shows non ambiguous and ambiguous curve profiles.
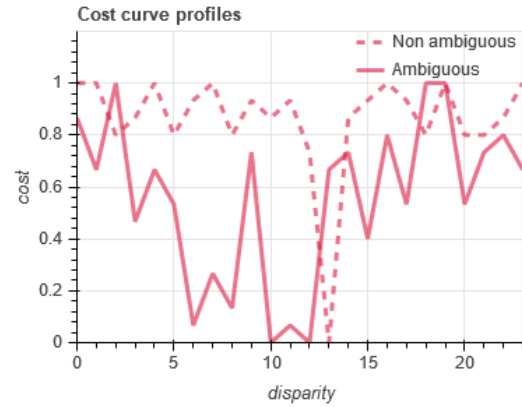

Figure 1. Cost curve profiles.

We propose a mathematical formulation for ambiguity:

$$Amb(x, y, \eta) = Card\big(\{d \in [d_{min}, d_{max}] \\ | cv(x, y, d) < \min_d(cv(x, y, d)) + \eta\}\big) \qquad (3)$$

where $cv(x, y, d)$ is the cost value at pixel (x, y) for disparity d in disparity range $[d_{min}, d_{max}]$. This is a local formulation, since the neighborhood of a pixel is not taken into account. The ambiguity is related to a vertical gap between cost curve points. Figure 2 represents the methodology of the ambiguity curve creation using equation 3. The figure 3 shows ambiguity curve for non ambiguous and ambiguous profiles. The faster the curve increases, the more it demonstrates the ambiguity of a pixel. In the best cases, the curve should increase only for high η values, as for non ambiguous profile in figure 3.
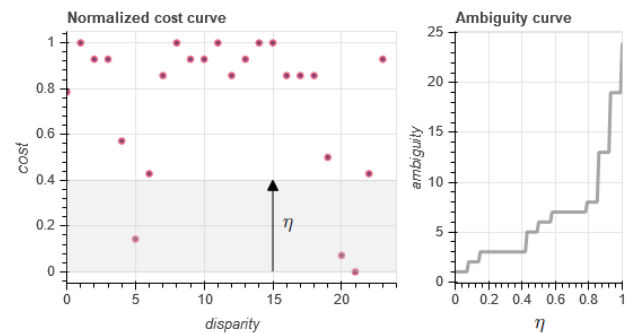

Figure 2. Methodology to compute ambiguity curve from cost curve (left), Ambiguity curve (right).

3.2 Ambiguity integral metric

From the ambiguity curve, we derived a new confidence metric to assess the quality of disparity map, called *Ambiguity integral measure*. It is defined as the area under the ambiguity curve:

$$Amb_{int}(x, y) = \int Amb(x, y, \eta) d\eta \qquad (4)$$

The more ambiguous a point is, the larger the ambiguity integral measure. For implementation, the integral is discretized:

$$Amb_{int}(x, y) = \sum_{k=0}^{N} Amb(x, y, k\delta\eta)\delta\eta \qquad (5)$$
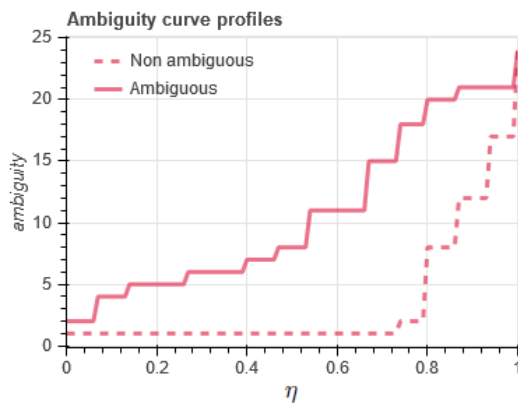
Figure 3. Ambiguity curve for ambiguous and non ambiguous profiles.

The discretization step $\delta\eta$ must be smaller than the quantification step of the similarity measure. This metric enables to keep a reasonable computation time. It can be considered as the opposite of a confidence metric as defined by (Hu and Mordohai, 2012) and can be normalized to keep values between $[0, 1]$. Consequently, a confidence metric can be derived with the following formula:

$$Conf_{Ambint}(x, y) = 1 - Amb_{int}(x, y) \qquad (6)$$

This metric is closed to the Probabilistic Measure (PRB) defined by (Hu and Mordohai, 2012). But, PRB is limited to a measure that reaches its minimum and maximum values.

4. EVALUATION

The quality of *ambiguity integral metric* is evaluated using the ROC curve methodology described in 2. We compare this metric with several confidence measures described in (Hu and Mordohai, 2012; Poggi et al., 2017). We have chosen to include measures with moderate computation time and robust enough to be integrated in a space-industry image processing ground segment. Therefore, machine learning-based confidence measures are excluded in the comparison. To study this metric, Pandora[1], an open-source stereo matching framework (Cournet et al., 2020), is used. Inspired by the (Scharstein and Szeliski, 2002) taxonomy, Pandora is a modular pipeline that allows to configure pipeline steps by testing various algorithms. Pandora is not only a tool dedicated for prototyping as it will be included in the 3D reconstruction pipeline for the CO3D mission (Lebègue et al., 2020). For the evaluation, Pandora's stereo pipeline is described in figure 4.

4.1 Middlebury case

Figure 5 illustrates results on Middlebury Cones and shows that ambiguity integral measure is a good confidence measure. Indeed, the curve is located below the other confidence measure curves. Moreover, we can mention that errors points are well located to the right of the curve. This therefore implies that the most ambiguous points are mainly the error points. Table 1 presents AUC value for confidence measures. The AUC value for ambiguity integral measure is lesser than other confidence measures and for this example, it is very close to the ideal confidence value. In figure 6, we can notice that ambiguous areas
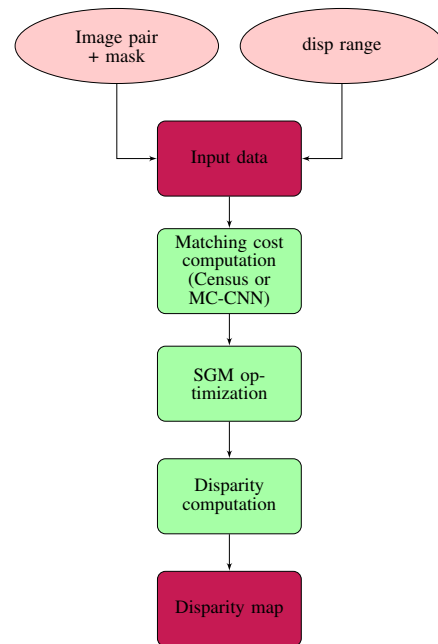
[1] https://github.com/CNES/Pandora

Figure 4. Pandora's stereo pipeline.

are located in occlusion areas. The figure 7 presents the histogram of the distribution of ambiguity measure. The majority of points have a low ambiguity integral value.
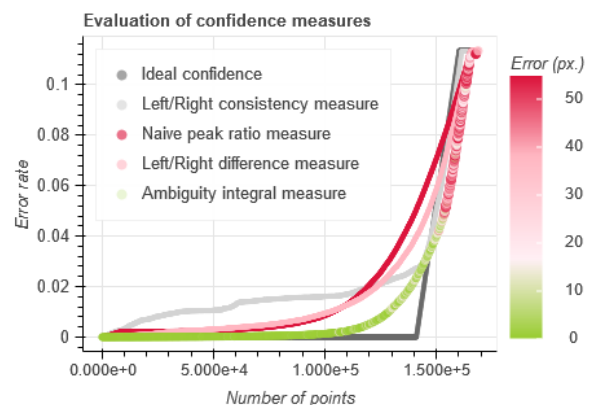

Figure 5. ROC curve comparison for Middlebury Cones. Circle colors correspond to error difference in pixels between ground truth (Stereo matching pipeline with Census and SGM, threshold at 3 pixels with ground truth).

Confidence measure	AUC value
Ideal	*1032.3*
Ambiguity integral	**1051.5**
Left/Right difference	1652.8
Naive Peak Ratio	1823.5
Naive winner margin	1906.0
Left/Right consistency	1986.8
Naive maximum margin	2526.6
SGM number of paths	8107.6

Table 1. AUC comparison for Middlebury Cones case (Stereo matching pipeline with Census and SGM, threshold at 3 pixels with ground truth).

4.2 Influence of similarity measure used

In figure 8, we compare the results between matching costs computed with Census and with MC-CNN (Žbontar and
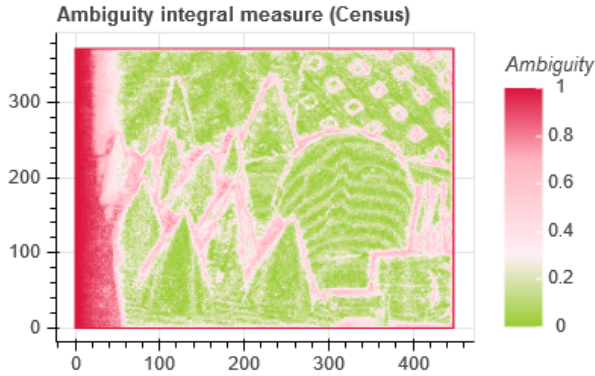
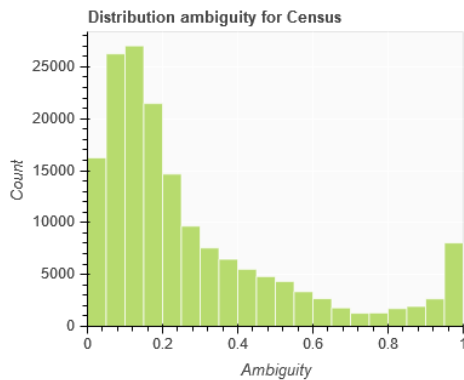Figure 6. Ambiguity integral measure for Middlebury Cones (Stereo matching pipeline with Census and SGM).

Figure 7. Histogram of ambiguity integral measure for Middlebury Cones (Stereo matching pipeline with Census and SGM).

LeCun, 2015). We can notice that MC-CNN measure is less ambiguous than Census. The use of SGM optimization has less impact on AUC value for MC-CNN than it has for Census. This is one limitation of the ambiguity integral measure: if the measure is itself very ambiguous, the metric is not able to well discriminate pixels.
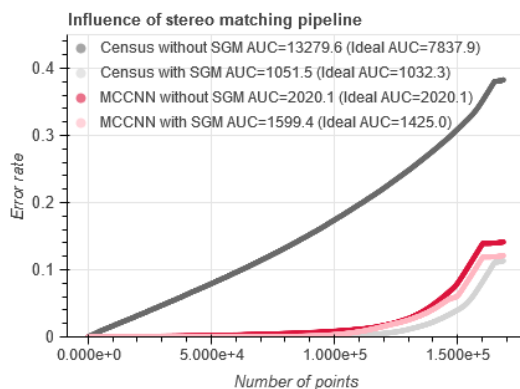

Figure 8. ROC curve comparison for Middlebury Cones with ambiguity integral measure for different stereo matching pipelines (Threshold at 3 pixels with ground truth).

4.3 Remote sensing case

We apply the ambiguity integral measure on two remote sensing cases: World-View 3 images of Buenos Aires (Argentina),

Pléiades images of Montpellier (France). Figure 9 presents ROC comparison between confidence measures. Table 2 shows comparison of *AUC* results for satellite images. As for the Middlebury Cones case, ambiguity integral metric offers better results than other confidence measures.
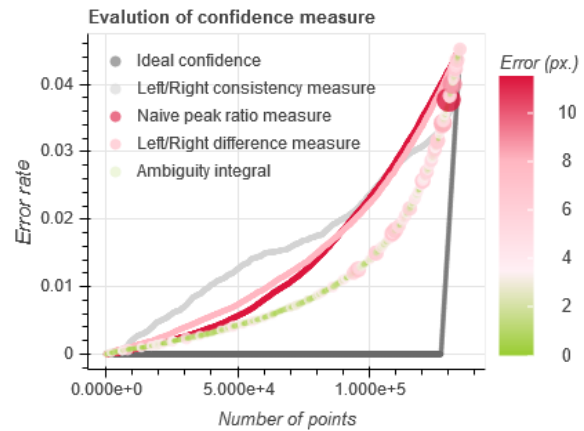

Figure 9. ROC curve comparison for Montpellier images.

5. APPLICATIONS

Ambiguity integral metric can be used to assess the quality of the produced disparity map. But it can also be directly exploited in the stereo matching pipeline to improve stereo matching performances.

5.1 Improve SGM optimization

The objective of optimization step is to add a smoothness constraint on the cost volume. This constraint can result in the minimization of energy:

$$E = E_d + E_s \tag{7}$$

where E_d contains the local information of cost volume

$$E_d = \sum_{\mathbf{p}=(x,y)} cv(\mathbf{p}, d(\mathbf{p})) \tag{8}$$

and E_s is the term that imposes the regularization and carry the global information. Semi-Global Matching (*SGM*) method is an algorithm that approximates global 2D smoothness constraint by combining 1D constraint along particular directions (Hirschmuller, 2008). For one direction $\mathbf{r}$, energy to minimize is declined from equation 7:

$$L_{\mathbf{r}} = E_d(\mathbf{p}) + E_s(\mathbf{p}, \mathbf{r}) \tag{9}$$

$$
\begin{aligned}
L_{\mathbf{r}}(\mathbf{p}, d) = & cv(\mathbf{p}, d) + \min(L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d), \\
& L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d - 1) + P_1, \\
& L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d + 1) + P_1, \\
& \min_i L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, i) + P_2) \\
& - \min_k L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, k) \tag{10}
\end{aligned}
$$

where $\mathbf{p}$ is the pixel at position (x, y), $\mathbf{r}$ is the direction, and P_1 and P_2 are constant penalties for disparity change respectively of one and more. Then, the optimized cost volume is obtained

Confidence measure	AUC value for Buenos Aires images	AUC value for Montpellier images
Ideal	*84.8*	*284.1*
Ambiguity integral	**286.6**	**1346.3**
Left/Right difference	418.3	1901.0
Naive Peak Ratio	410.8	1846.1
Naive winner margin	459.4	2239.8
Left/Right consistency	489.3	2009.6
Naive maximum margin	507.0	2187.5
SGM number of paths	1403.1	3671.9

Table 2. Confidence measure comparison on satellite images (Threshold at 3 pixels with ground truth).

by combining all directions.

$$S(\mathbf{p}, d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d) \qquad (11)$$

In this section, our goal is to reduce the impact of more ambiguous pixels, through the usage of the new confidence measure in optimization step. (Höllmann et al., 2020) has also presented another methodology to take into account confidence measure and geometrical constraints into SGM optimization.
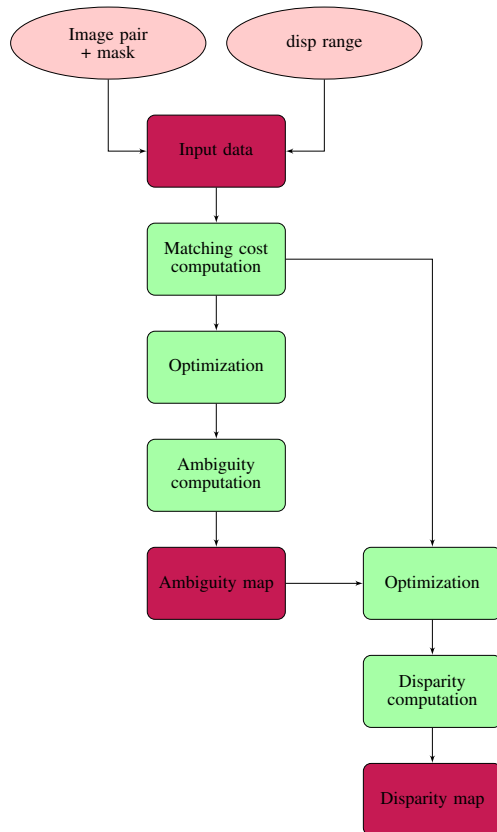

Figure 10. Pipeline with ambiguity integration in optimization step.

5.1.1 Methodology In SGM, same weight is given to E_d and E_s terms. But, some pixels might have a badly estimated disparity and they might degrade the disparities on the neighborhood after the optimization step. The idea is to identify non ambiguous points in order to rely on them during the optimization step. This can be done by weighting, by confidence measure, the influence of points in the SGM regularization in order to reduce impact of ambiguous points. Consequently, depending on whether a pixel is more reliable than the previous pixel

along direction $\mathbf{r}$, we might want to give more weight to E_d than in E_s, or vice versa. In our approach, for one direction $\mathbf{r}$, equation 9 is derived as follow:

$$L_{\mathbf{r}}(\mathbf{p}, d) = E_d^{Conf}(\mathbf{p}, d) + E_s(\mathbf{p}, d) \qquad (12)$$

where E_d^{Conf} is defined by:

$$E_d^{Conf}(\mathbf{p}, d) = Conf(\mathbf{p}, d) \times cv(\mathbf{p}, d) \qquad (13)$$

and $Conf$ is the confidence measure set in equation 6. By multiplying E_d by confidence measure value, the influence of E_s fluctuates. It enables, in extreme case where $\mathbf{p}$ is very confident and $\mathbf{p} - \mathbf{r}$ is not, to avoid optimizing $\mathbf{p}$ by $\mathbf{p} - \mathbf{r}$. On the contrary, in a case where $\mathbf{p} - \mathbf{r}$ is very confident and $\mathbf{p}$ is not, it allows to reduce the influence of E_d term against E_s term for point $\mathbf{p}$. Equation 10 becomes:

$$
\begin{aligned}
L_{\mathbf{r}}^{conf}(\mathbf{p}, d) = & Conf(\mathbf{p}, d) \times cv(\mathbf{p}, d) \\
& + \min(L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d), \\
& L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d - 1) + P_1, \\
& L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d + 1) + P_1, \\
& \min_i L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, i) + P_2) \\
& - \min_k L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, k) \qquad (14)
\end{aligned}
$$

As shown in figure 10, the stereo pipeline optimization step is modified to include confidence accordingly. A first pipeline is executed to generate a confidence map. Then, the initial cost volume and the confidence map are used to compute a new optimization step.

5.1.2 Evaluation The figure 11 presents the results of the modified SGM algorithm. The method shows improvements on area with constant disparity values. Nevertheless, the approach adds errors at buildings borders where there are high disparity gaps. There is a degradation at disparity discontinuities, because these are not confident areas. To work correctly, this method still needs to rely on confident points. So, it can remove correct areas if they are ambiguous. We can compare the use of ambiguity with the use of high P_2 penalty. The method with high P_2 penalty behaves like a majority vote. The most represented values are propagated and it is more difficult to move from one disparity to another. It delays disparity jumps and can sometimes eliminate some of them. On areas with mostly wrong disparity or noise, the method can smooth disparities with noise. On the contrary, optimization with confidence is not influenced by the majority.

5.1.3 Conclusion Results are promising but have to be improved especially at object borders, where there are disparity gaps that correspond to ambiguous areas. We could use a

(a) Left image   (b) Ground truth

(c) Disp map using   (d) Disp map using
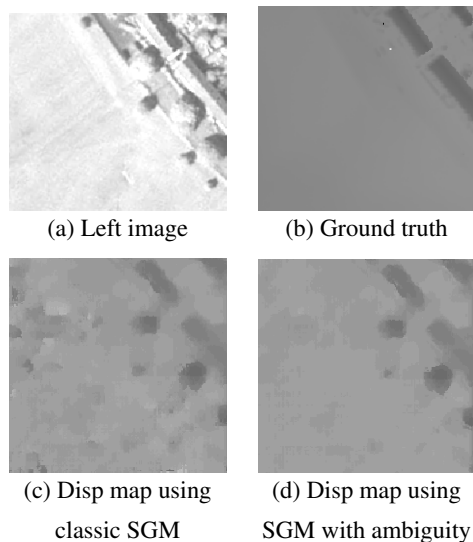
classic SGM   SGM with ambiguity

Figure 11. Usage of ambiguity in SGM Optimization.

segmentation to perform a piecewise optimization step. The objective would be to use segmentation information in order to cancel the historic along a direction and ease the disparity change, where we cross a segment.

5.2 Disparity map denoising

There exist classical approaches to fill invalid pixels identified as mismatch or occlusion (Hirschmuller, 2008; Žbontar and LeCun, 2015). In (Hirschmuller, 2008), the principle is to detect occlusions and mismatches during cross-check validation and then, to use the median of the valid disparity value in each direction to fill mismatches and second lowest value to fill occlusions.

In last few years, denoising with neural network have been proposed. Recent methods (Stucker and Schindler, 2020) use encoder-decoder architectures for denoising like U-Net (Ronneberger et al., 2015).

More complex architectures have been proposed (Gidaris and Komodakis, 2017). In this section, we propose to evaluate the contribution of ambiguity in an encoder-decoder architecture for denoising the disparity map. For this purpose, two encoder-decoder neural networks are trained. The first one, in figure 12, uses the left image and the noisy disparity map. The second, in figure 13, adds the ambiguity information.

5.2.1 Architecture The purpose is to use ambiguity information to easily locate noisy areas and preserve details in non ambiguous areas (cf. figure 13). The network takes as input noisy disparity map, left image and normalized ambiguity (in $[0, 1]$). The residual disparity and ambiguity map are used at the end of the network to guide the denoising. In ambiguous areas, output disparity will be modified by the network. In unambiguous areas, disparity will remain unchanged.

5.2.2 Dataset description We use two sets of input images:

- Selection of image pairs originated from Data Fusion Contest (*DFC*) dataset (Bosch et al., 2019) (World-View 3 images)
- Image pair on Montpellier (Pléiades images).

Noisy disparity maps are generated with the Pandora's pipeline shown in figure 14. Ground truth (GT) are generated with the methodology described in (Cournet et al., 2020). One difficulty is the temporal inconsistencies between lidar and images, especially with regard to trees. This time gap can impact the quality of the neural network training.

5.2.3 Training For training, images are divided into patches of size 256×256. The loss function used is L1 loss modified to constraint the network to avoid adding errors:

$$Loss = |denoised - GT| * W$$

$$W = \begin{cases} 1.25 & \text{if } |denoised - GT| > |noisy - GT| \\ 1 & \text{otherwise} \end{cases}$$

First results have shown the advantage of combining both image sets. With the same network architecture, results were different for both datasets. With the DFC dataset the denoising is more aggressive. The disparity maps tend to be very smooth but coarse errors are well fixed. With Montpellier dataset the denoising is lighter. Disparity maps are less smoothed and building edges are preserved. But large errors are not corrected. So, each database has its own advantages and disadvantages. In order to combine advantages, training is performed by merging the datasets.

Moreover, as noisy disparity maps contain a lot of errors, neural network will favor a smoothing on the whole image in order to correct disparity changes. To avoid this effect, ground truth patches are integrated during training: for 30% of batch epoch the disparity map in input is replaced by the ground truth. Thanks to this method, we can observe that the disparity map is less smoothed: the buildings are less impacted by the denoising. The training is stopped after 50 epochs.

	% error points	Average error	Median error	% error point corrected
Noisy map	13.00	2.58	1.81	
Denoised map with classical approach	12.19	2.38	1.75	9.51
Denoised map with U-net	15.24	5.29	1.65	29.15
Denoised map with U-net Ambiguity	13.93	5.62	1.66	26.04

Table 3. Comparison of denoising methods.

5.2.4 Results The figure 15 shows results on Buenos Aires. This image does not belong to the training dataset. Consequently, it demonstrates the ability of the neural network to generalize. The addition of ambiguity gives suitable results, better than original U-Net architecture. It allows to preserve building edges, whereas without ambiguity building edges are trimed. Compared to classical methods, disparity maps denoised with deep learning methods are smoother but details are preserved. On figure 15, we can notice that errors introduced with denoising using deep learning methods are mainly located on trees. It derives from the training database where there exists a time shift between images and lidar reference.

Table 3 presents the comparison between ground truth and denoised disparity map. Methods based on neural networks fill more errors than classical approaches but they also produce
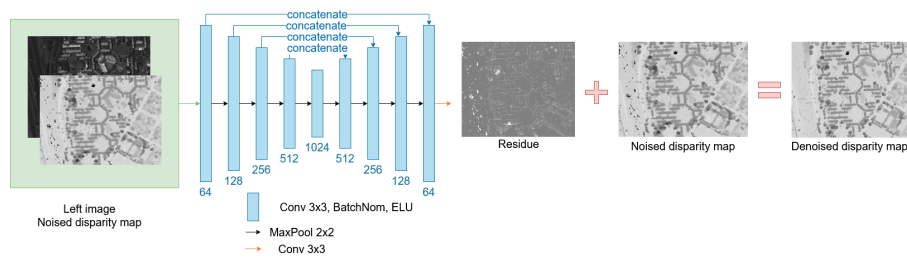
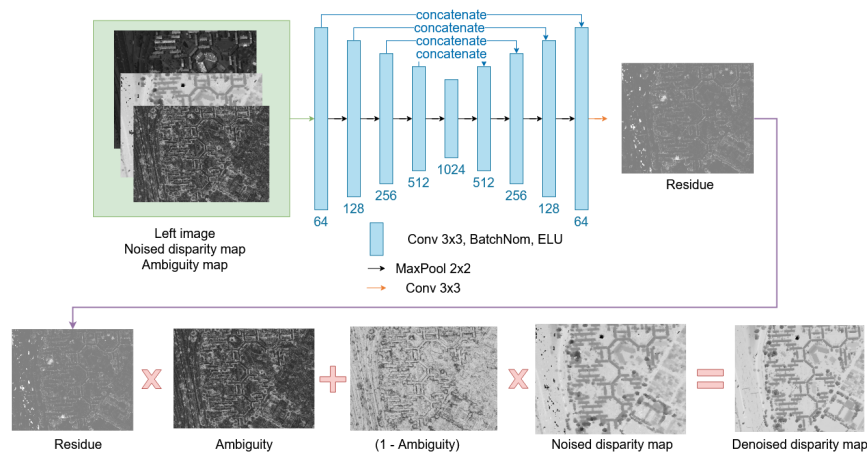Figure 12. Encoder-decoder architecture for denoising.

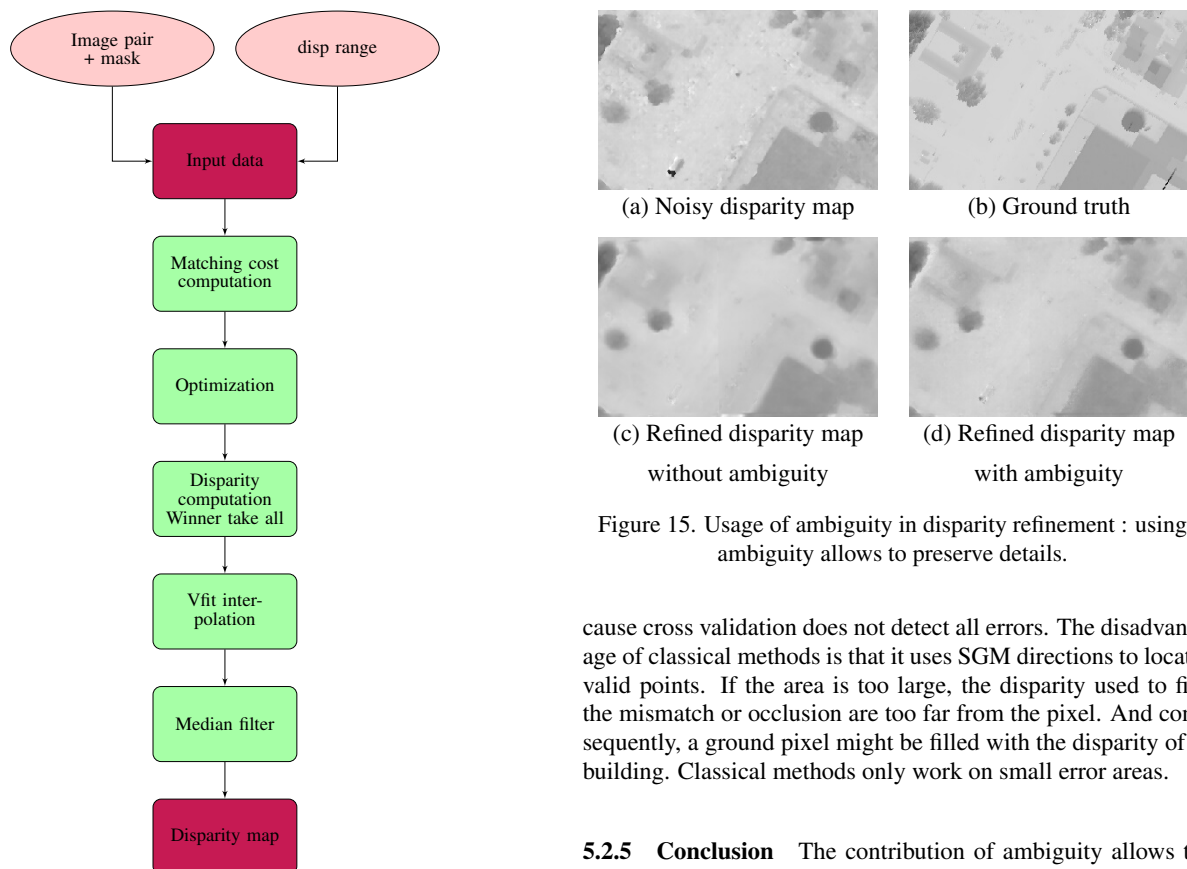Figure 13. Encoder-decoder architecture for denoising with ambiguity.

Figure 14. Pandora's pipeline.

(a) Noisy disparity map   (b) Ground truth

(c) Refined disparity map   (d) Refined disparity map

without ambiguity   with ambiguity

Figure 15. Usage of ambiguity in disparity refinement : using ambiguity allows to preserve details.

cause cross validation does not detect all errors. The disadvantage of classical methods is that it uses SGM directions to locate valid points. If the area is too large, the disparity used to fill the mismatch or occlusion are too far from the pixel. And consequently, a ground pixel might be filled with the disparity of a building. Classical methods only work on small error areas.

5.2.5 Conclusion The contribution of ambiguity allows to improve denoising results compared to original U-net denoising architecture. This architecture outperforms classical denoising methods. A semantic segmentation may be useful to produce statistics by classes. It also allows to perform a specific denoising by classes.

more errors. Classical denoising methods fill less points be-

6. CONCLUSIONS AND PERSPECTIVES

In this article, we have proposed a new confidence metric, called *Ambiguity integral metric*. It is based on the ambiguity concept of the cost curve profile, that characterizes the difficulty in identifying the correct disparity to select. This metric offers good performances to evaluate the quality of the disparity map, while keeping reasonable computation time in order to be included in space-industry ground processing platform. The implementation of this metric is available in Pandora.

Moreover, we demonstrate that this metric can be directly exploited in the stereo-matching pipeline to improve disparity estimation performance and therefore enhance 3D pipeline results. First, it can be introduced in the SGM optimization step. This can be done by weighting, by this confidence measure, the influence of points in the SGM regularization in order to reduce impact of ambiguous points. Results are promising but have to be improved especially at object borders, where there are disparity gaps that correspond to ambiguous areas. Secondly, this metric can be included in the disparity refinement step. We use ambiguity as an input of a disparity refinement deep learning architecture in order to easily locate noisy areas and preserve details in non-ambiguous areas. The contribution of ambiguity allows to improve the disparity refinement step performance.

Our work in progress include taking into account the neighborhood of a pixel when computing its ambiguity. Also, as the ambiguity only reflects on the probability of a given match to be wrong, we plan on estimating the disparity error for ambiguous points. This new metric would represent the risk associated on a disparity choice for ambiguous points. The further the minima are, the higher the risk is, because if the wrong point has been selected then the error will be significant. By including both ambiguity and risk concepts, we would be able to take into account the whole information of the cost curve.

REFERENCES

Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G. D., Brown, M., 2019. Semantic stereo for incidental satellite images. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1524–1532.

Bosch, M., Kurtz, Z., Hagstrom, S., Brown, M., 2016. A multiple view stereo benchmark for satellite imagery. *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 1–9.

Cournet, M., Sarrazin, E., Dumas, L., Michel, J., Guinet, J., Youssefi, D., Defonte, V., Fardet, Q., 2020. Ground truth generation and disparity estimation for optical satellite imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 127–134.

Gidaris, S., Komodakis, N., 2017. Detect, replace, refine: Deep structured prediction for pixel wise labeling. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7187–7196.

Gouveia, R., Spyropoulos, A., Mordohai, P., 2015. Confidence Estimation for Superpixel-Based Stereo Matching. *2015 International Conference on 3D Vision*, 180–188.

Haeusler, R., Nair, R., Kondermann, D., 2013. Ensemble Learning for Confidence Measures in Stereo Vision. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 305–312.

Hirschmuller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328-341.

Hu, X., Mordohai, P., 2012. A Quantitative Evaluation of Confidence Measures for Stereo Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2121–2133.

Höllmann, M., Mehltretter, M., Heipke, C., 2020. Geometry-Based Regularisation for Dense Image Matching via Uncertainty-Driven Depth Propagation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020, 151–159.

Kim, S., Kim, S., Min, D., Sohn, K., 2019. LAF-Net: Locally Adaptive Fusion Networks for Stereo Confidence Estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 205–214.

Lebègue, L., Cazala-Hourcade, E., Languille, F., Artigues, S., Melet, O., 2020. CO3D, A worldwide one one-meter accuracy DEM for 2025. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2020, 299–304.

Mehltretter, M., Heipke, C., 2019. CNN-Based Cost Volume Analysis as Confidence Measure for Dense Matching. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2070–2079.

Michel, J., Sarrazin, E., Youssefi, D., Cournet, M., Buffe, F., Delvit, J. M., Emilien, A., Bosman, J., Melet, O., L'Helguen, C., 2020. A new satellite imagery stereo pipeline designed for scalability, robustness and performance. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020, 171–178.

Min-Gyu Park, Yoon, K., 2015. Leveraging stereo matching with learning-based confidence measures. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 101–109.

Poggi, M., Tosi, F., Mattoccia, S., 2017. Quantitative Evaluation of Confidence Measures in a Machine Learning World. *2017 IEEE International Conference on Computer Vision (ICCV)*, 5238–5247.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv e-prints*.

Scharstein, D., Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1), 7–42.

Spyropoulos, A., Komodakis, N., Mordohai, P., 2014. Learning to Detect Ground Control Points for Improving the Accuracy of Stereo Matching. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1621–1628.

Stucker, C., Schindler, K., 2020. Resdepth: Learned residual stereo reconstruction.

Žbontar, J., LeCun, Y., 2015. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. *arXiv e-prints*, arXiv:1510.05970.