Jaccard Distance as Similarity Measure for Disparity Map Estimation

V.A. Gonzalez-Huitron A.E. Rodríguez-Mata L.E. Amabilis-Sosa R. Baray-Arana Isidro Robledo-Vega G. Valencia-Palomo

***Abstract*—High confidence in disparity map estimation is critical in several application fields. A novel framework that employs customized local binary patterns and Jaccard distance for stereo matching along stereo consistency checks is presented. The proposal contributes with a method that allows greater confidence in its estimates, without dependence on supervised learning, and capable of generating a dense map with low-cost filtering. The proposed framework has been implemented in CPU and GPU for parallel processing capability. First, Local binary patterns are obtained during the initial stage; then, the Jaccard distance is employed as a similarity measure in the stereo matching stage; subsequently, a matching consistency check is performed, and singular disparities are removed. A comparison among novel and state-of-the-art algorithms for sparse disparity map estimation is performed employing Middlebury and KITTI stereo Datasets where the quality criteria used were percentage of bad pixels (B), quantity of invalid pixels, processing time and running environments to put each framework into context, obtaining down to 2.07% bad matching pixels and performing better than state-of-the-art cost functions.**

***Index Terms*—Stereoscopic vision, disparity mapping, Jaccard, image processing**

I. Introduction

Computer vision within digital signal processing can be considered a set of techniques and models that allow visual information processing using digital images. From the beginning, developments in computer vision have been inspired by the study of the human visual system (HVS), which suggests the existence of different types of treatment of visual information depending on specific goals or objectives where one of them is an estimation of depths or distances through stereotypical techniques [1]. The main objective of computer vision is to model and automate the process of visual recognition and therefore, to distinguish between objects to follow the trajectory of objects in images, making decisions in diverse multidisciplinary fields, such as the estimation of distances, speed, acceleration, sizes, or shapes.

One of the critical areas of current interest in research is the disparity estimation from two or more views of a scene. The basic principle indicates that it is possible to determine the difference in the relative position from one image to another for each point or object from a pair of stereo images. This difference indicates the disparity for a given point or can be related to optical flow [2], [3]. The disparity is directly related to the depth of a scene, so a correct estimation of a disparity map is essential for the reconstruction of different scenarios [4], e.g., medical tissues, surfaces, distance estimation for control applications, as an important parameter for decision making in more complex computer vision systems [5]–[7], among others.

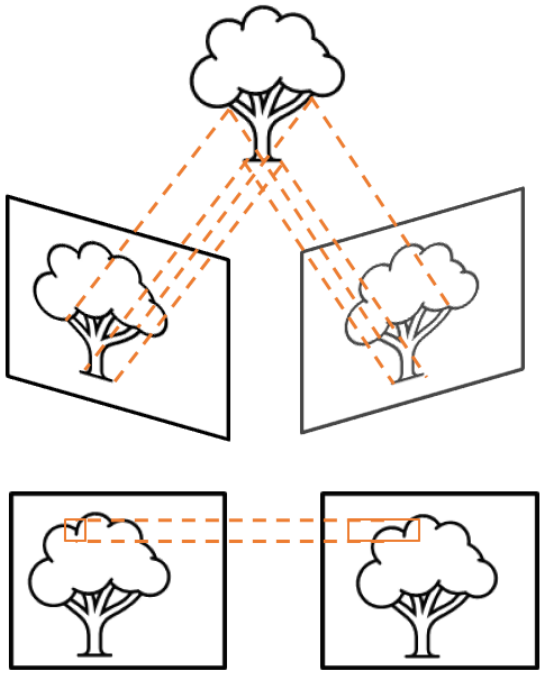


Fig. 1 Disparity estimation from a stereo pair.

The problem of estimation from stereo images consists of obtaining two images of the same scene commonly aligned horizontally, which allows locating a point of interest in a relative position with respect to the other image, this difference in relative position is known as disparity, in Fig.1 can be seen generically. When the disparity estimation is performed on points or regions of interest is known as a Sparse map, while performing the estimation for all points of the scene is known as a Dense map.

However, despite a plethora of algorithms in disparity map estimation, the complexity, difficulty, and applications of the

V.A. Gonzalez-Huitron, Instituto Politécnico Nacional, Santa Ana 1000, ESIME Culhuacan, Mexico-City 04440, Mexico.
Leonel Amabilis-Sosa. CONACYT-TecNM campus Culiacan, Juan de Dios Bátiz No. 310 Pte. , Col. Guadalupe, C.P. 80220 Culiacán Rosales, Sin.
Baray-Arana, Robledo-Vega and Rodriguez-Mata, Tecnológico Nacional de Mexico/IT Chihuahua, Av. Tec. 2909, 31200, Chihuahua, Mexico.
G. Valencia-Palomo, Tecnológico Nacional de México, IT de Hermosillo, Av. Tec. y Per. Poniente, S/N, 83170, Hermosillo, Mexico.
Corresponding author: abraham.rm@chihuahua.tecnm.mx

topic justify the design of new methods that can achieve an improvement over the most currently used methods. For example, in recent years, the usage of deep learning approaches has achieved outstanding results [8], [9]. However, their high computational cost for the training stages makes it difficult to apply them on a larger scale in other areas of knowledge. Other methods achieved results with low computational costs [10], as well as implementations in low energy consumption devices or embedded systems. However, the quality of disparity maps is seriously affected or limited to only a few scenarios. These drawbacks demand designing novel methods that can improve quality metrics while processing times are kept within the margin of better state-of-the-art algorithms. The novel method that has been designed in this study can resolve mentioned drawbacks where the principal contributions are as follows: 1) Proposed custom pre-processing stage based on Local Binary Patterns. 2) Introducing a stereo matching stage with a non-parametric similarity measure, new for disparity estimation, and 3) Including an additional disparity check with a stereo consistency check.

The rest of the paper is organized as follows. In Section II, some relevant and recent related works are described. In Section III, the novel proposal for sparse disparity estimation is presented, and the hardware implementation is detailed. In Section IV, the quality criteria, datasets, and experimental results are explained. Finally, in Section V, the conclusions are presented.

II. Related works

Nowadays, there is a wide range of methods for estimating disparity maps, some of which are briefly described in this section.

One of the most used algorithms is SGBM1 [11] that is thanks to their implementations in OpenCV libraries from version 2.4 to the most recent release to date, in popular programming languages like C++ and Python. In the first stage of this framework, the search for similar pixels is performed in a 3×3 window using the sum of absolute differences (SAD) criterion; during the next stage, a stereo consistency check is used. Finally, an interpolation and disparity refinement stage is performed employing the dynamic programming in five of eight possible directions, which is carried out through interpolations along the defined scanlines.

For the estimation of sparse disparity maps, MotionStereo algorithm [12] uses as the first stage a keyframe selection, and then, it performs a stereo rectification and shifting process to ensure the condition of horizontal epipolarity and to revise the absence of negative disparities. In the following stage, an extraction of binary characteristics should be performed as a tree binary decision process that allows the estimation of stereo correspondence with a lower number of operations and is independent of local window sizes. Every image value is converted to a binary vector within 32-bit integer values via Local Binary Patterns (LBP), then the calculation for the similarity between data sets is performed by means of the information gain, which is obtained from the entropy and covariance matrix of the data sets. To reduce the number of operations, only five possible disparities within the disparity range are evaluated, which are randomly selected, choosing the one with the lowest cost. Finally, a median filter with a 3×3 window is applied over the disparity map. MotionStereo is designed for application on mobile devices, so CPU or GPU usage is limited here, obtaining dense disparity maps through bilateral plane estimations and data interpolations.

Paper [13] presents the DCNN procedure, in which the disparity maps are obtained from previous training through the Middlebury dataset carried out in two stages. First, a non-parametric transform using Rank transform is used to achieve better results due to lightning changes in a scene, then a Companion Transform is performed, which aims to determine values for uniform or low texture regions for improving the performance of disparities estimation in such regions. The stereo matching stage is carried out through a convolutional neural network (CNN), which should select the disparity value with the lowest cost according to the hyperparameters that are determined during training. In the following stage, a second CNN to determine occlusions and invalid pixels in the disparity map is used, carrying out the corrections without the need for a filtering stage or stereo consistency check between left and right maps. MC-CNN-art proposed in [14] is based on previous work from DCNN, where the most notable difference is that the images must be converted to grayscale, and the refinement of disparities should be carried out by means of cross-based cost aggregation, semi-global matching, a left-right consistency check, subpixel enhancement; finally, a median and bilateral filtering should be applied.

The DAWA [15] method comprises two main steps. First, it uses adaptive support weights for stereo local matching. Aside from the geometric distance and color similarity, the adaptive weight distribution benefits pixels in the block matching with a smaller cost, where the window size is determined as 27 for the benchmark tests. Besides, a multiscale strategy with invalidation criteria to reduce match ambiguity and computational time is performed. In the second stage, a global interpolation using a variational formulation is performed. The energy functional penalizes deviations from the local disparity estimation at different scales.

Intel RealSense RGBD imaging systems denoted here as r200high framework, are described in [16]. The work presents a method implemented in FPGA for a fast estimation of disparity maps where image corrections and rectification employing hardware solutions are initially performed. Then, the Census transform is applied to the stereo pair with a 7×7 window; next, the search for similarities is done with Winner-Take-All criteria, and finally, a series of filters discard the disparities that are considered erroneous. The objective of this system is to deliver disparity maps with high certainty in the values that are presented for a correct estimation of depth. It should be noted that one of the important limitations is the need for specific hardware for the recreation of results, as well as the important limitation of maintaining the disparity range as a fixed value.

For Displets framework [17] firstly, using left and right images of a synchronized, calibrated and rectified stereo camera, a sparse disparity map is estimated at each pixel using

the semi-global matching framework followed by a simple left-right consistency check; next, matching costs features are obtained via CNN [18], following, the main image is decomposed into a set of planar superpixels via displets for long-range interactions, which are disparity maps from a given semantic class that are associated with an image mask and a score, for example, 3D car shapes that are provided for dense DM refinement.

SGM-Nets [19] aims to predict dense DM with semi-global matching, where a learning-based penalties estimation method is proposed. It consists of CNN where a grayscale image patch of 5×5 pixels and its position taken as input to the framework, then the proposed CNN (Nets) should predict the penalties for dense DM refinement. Real and synthetic stereo pairs are used to train the proposed method. The sparse DM is obtained via CNN matching cost and Zero Mean Normalized Cross-Correlation for comparison results.

CSPN [20] is a method designed for efficient propagation of values, which is carried out employing recurrent CNN. Here the affinity between pixels is determined through a CNN model, which can be applied to in-depth reconstruction, stereo estimation, and optical flow. As an initial stage, a map of sparse values and a reference image should be presented to realize a diffusion process derived from PDE. This diffusion process is iterative and is performed in a pyramidal way for propagation at different distances from the initial point. Because this process is based on CNN, it can be implemented on GPUs, reducing processing times. Like the other methods based on deep learning, the models and training are independent for each dataset.

Other methods can achieve outstanding results [13], [14], [17], [18] presenting the limitation of requiring a large amount of data and processing time for a previous training stage, limiting these methods only to scenarios to which the algorithm has been designed. Other techniques present processing times that are a barrier to their usage in various applications despite the results achieved [15]. It should be noted that a common feature of several methods is the traditional usage of SAD as a similarity measure. SAD is a measure that can be implemented easily but in complex scenarios, this criterion demonstrates limitations, such as textureless regions, uneven light conditions between the stereo pair, or differences in exposure. Also, in CNN-based solutions, the matching costs demand a significant time amount in training sessions, so there can be used other measures, favoring more straightforward implementations where no training session is required [21]. In this study, we present an alternative approach that can compromise quality, processing time, and easy implementation, where an adequate pre-processing in combination with a similarity measure such as Jaccard distance, and a novel disparity check stage appear to demonstrate results with a higher level of confidence.

III. Proposed method

The proposed method consists of the following stages: Estimating a modified Census transform for each pair of images in a stereo pair, followed by a stage of similarity estimation employing Jaccard's measurement used as a matching cost to obtain left and right disparity maps. Those maps during the next stage are used for a consistency check where only a left disparity map should be obtained, then a refinement of disparities is applied. As a result of described operations, a sparse disparity map is estimated. The block diagram of the proposed framework is presented in Fig.2.

A. Pre-processing Stage

As a first step, gradient magnitude from the stereo pair employing Sobel edge estimation kernel is obtained, thus reducing dissimilarities from exposure and illumination variances between scenes. Then, a modified Census transform [22] is performed to obtain an LBP binary vector for each pixel [23]. Where both operations, Census and LBP focus on obtaining a binary pattern around the processed pixel, both, Census is applied for every pixel in the local neighborhood while LBP contains two parameters, radius and sample points [24], where a minor binary string is generated for different window size. For census exist approaches for local sub-sampling [25], this is our proposal case. The modified Census transform is presented in the form as follows:

$$\Gamma(x,y) = \bigotimes_{k\in P} C(I(x), I(y)), \tag{1}$$

where I is the gradient magnitude from an image, x and y denote the position value corresponding to an image, P denotes the neighbourhood around the *x,y* position, and k is the binary value for each value related to the neighbourhood, Γ is the binary vector that should be calculated. The neighborhood P is defined only in eight directions around any given position; therefore, the binary vector length is equal to $8n$, where n is the window size for the modified Census transform. The position of the binary values is determined in a clockwise direction in a spiral pattern to arrange the vector values. The binary vector is suitable for Jaccard distance computation at each pixel.

In Fig. 3, an example of the preprocessing stage is shown, where a comparison of Census transform, and our modified transform with Sobel magnitude for an n-slice is presented. Our method shows a more consistent performance by maintaining more variations across the image.

B. Stereo Matching Stage

The Jaccard distance (2) can be explained as complementary to the Jaccard similarity coefficient, which is used frequently for medical image segmentation [26]. The Jaccard distance measures the dissimilarity between two sets, and it is computed for every binary vector, shifting one image horizontally for every disparity level, choosing the disparity related to the minimum Jaccard distance for every position. Two DM are obtained at this stage:

$$D_{Jaccard}\ (VL_{i,j}, VR_{i,j}) = 1 - \frac{|VL_{i,j} \cap VR_{i,j}|}{|VL_{i,j} \cup VR_{i,j}|}. \tag{2}$$

The Jaccard Index is defined as the proportion of the intersection size to the union size of the two data samples. For

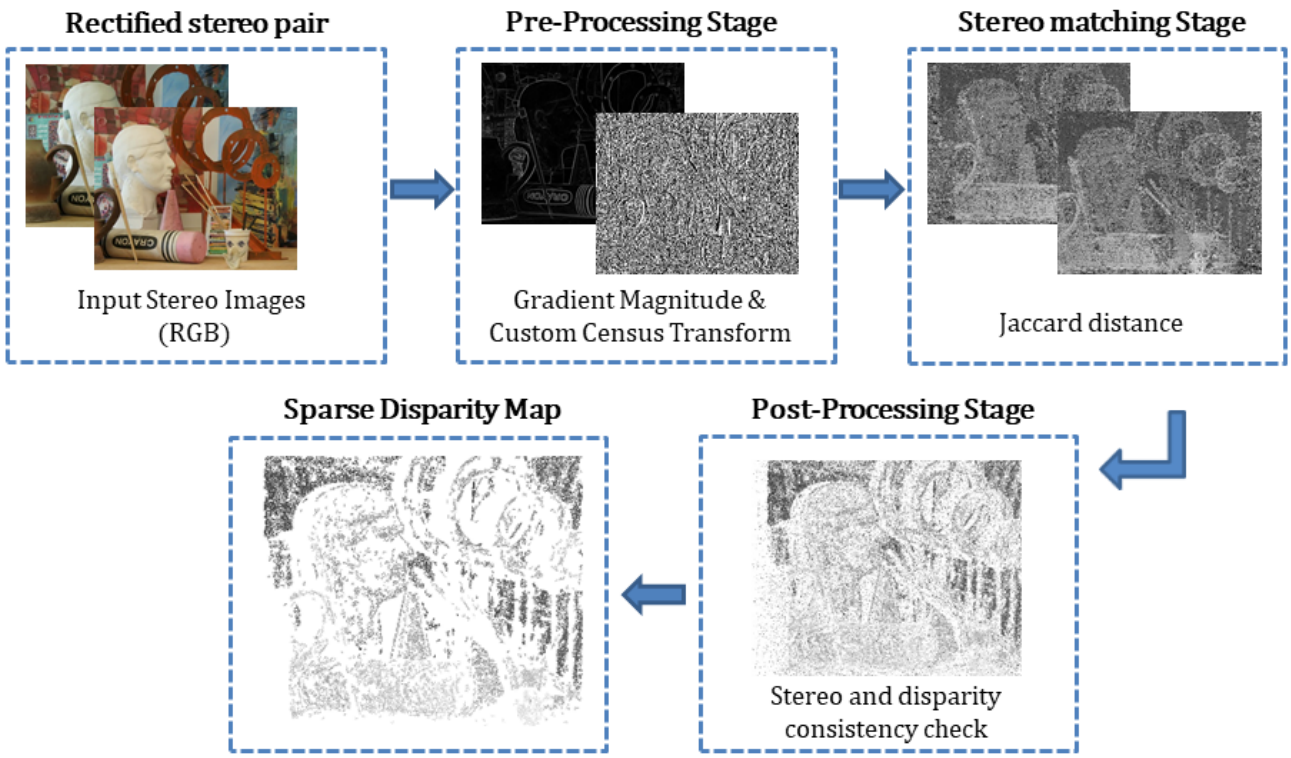
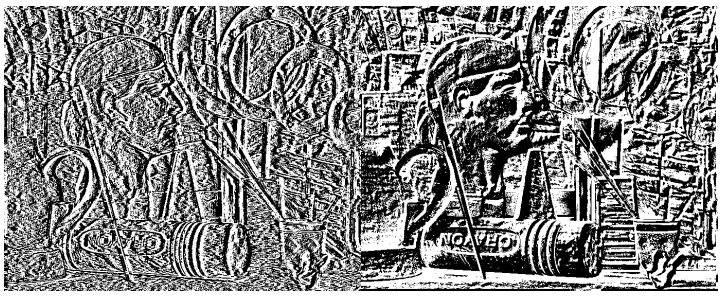


Fig. 2 Design for the proposed method.



(a) (b)

Fig. 3. Results for our proposed custom Census transform for ArtL image in Half Size from the Middlebury dataset where neighborhood size is 27; the 52nd slice is shown, image a) proposal, b) state-of-the-art approach.

our proposal, for each pixel a binary set is defined through a modified Census transform, therefore, comparing binary sets and obtaining a measure of its similarity at the core for the matching stage and disparity value estimation, when the Jaccard Index is best suited for obtaining a value that describes the intersection of two sets over the union of mentioned sets [27].

C. Post-Processing Stage

In this stage, the left one is only retained from both DM. First, a match consistency check is performed, [28]. Next, a disparity consistency check is performed. For every disparity value, a neighborhood area across the disparity value estimated is defined (an area of 12 disparities is used in this study), where only existing disparities among the defined range are preserved if the total number exceeds a threshold value T. Therefore, a disparity needs to be similar in location and intensity to be preserved, and outliers are removed. In this method, we use $T = 9$, the minimum disparities areas across a disparity range have to be preserved, and every other value is marked as a potential mismatch, finally, a NaN (Not A Number) label is assigned. Below, in (3) is shown for disparity map estimation, considering the sliding location process across an image to evaluate each pixel position for a given stereo-pair, where *n* and *m* is image width and height, *i* and *j* are columns, and rows and Γ is the custom Census transform for each stereo pair image, *w* is the disparity range and *k* disparity evaluated.

$$DM_{left} = \sum_{i=0}^{n} \sum_{j=0}^{m} \sum_{k=0}^{w} D_{Jaccard} \left(\Gamma_{left}(i,j), \Gamma_{right}(i,j-k) \right). \tag{3}$$

D. Hardware Implementations

The proposed framework has been implemented on a CPU Intel i3-7100 3.4GHz, 12Gb RAM and GPU NVIDIA Geforce GTX 1070 with 1920 CUDA cores and 8Gb GDDR5 in memory interface, using the IDE MATLAB R2019b. The implementation aims for an efficient and easy transferable routine, which can be easily deployed to other languages and IDEs. Both implementations, CPU, and GPU are suitable for parallel and multi-thread processing.

IV. Experimental Results

In this section, we present the data to be used, quality measures, qualitative and quantitative results as well as a brief discussion of them, where our main hypothesis is that to obtain a better MD, the confidence level of the values should be higher, even if this means that their number is lower.

A. Data

The Middlebury Stereo Vision dataset [29] was used for the performance evaluation of the novel framework and their comparison with better existing methods. The Half-size was selected due to being the most tested among frameworks in the Middlebury evaluation table. Also, the dataset KITTI Stereo 2015 [30] was used, which consists of 200 stereo pairs from real images.

B. Quality Criteria

The quantitative metric Percentage of Bad Matching Pixels (B) is used, which is shown in (4). Additionally, invalid pixels, which are the percentage of values marked as occlusion or mismatched disparities, labeled by any given method, are presented too.

$$B = \frac{1}{N} \sum_{(x,y)} (|DM_I(x,y) - DM_{GT}(x,y)| > \delta_d), \quad (4)$$

where DM_I is the estimated disparity, and DM_{GT} is the Ground Truth (GT), N is the total number of pixels in an image or frame, δ_d is the error threshold difference for each pixel evaluated. There are two results for estimated metric B: sparse and dense. The sparse result is presented by the DM estimated without any interpolation or reconstruction stages, also with occlusions marked. Dense DM is characterized as post-processed DM. A comparison of the developed framework and state-of-the-art methods has been performed between cost measures for sparse DM using the Middlebury and KITTI stereo datasets.

C. Evaluation Results

Experimental results for the Middlebury datasets were obtained for images in Half-Size format, and an 11x11 window size, where the criterion B results are shown in Fig. 4. The experimental results justify a reliable better performance of the novel framework that is superior in comparison with other frameworks (δ value was chosen at 1 for both datasets). Our proposed framework can maintain a high quality, and it demonstrates significantly fewer errors compared to related frameworks being surpassed only by DCNN and r200high (slightly). In Fig. 5, a visual comparison among different frameworks is presented in detail. There, the white values indicate a non-valid disparity or occlusion labeled by our framework, both must be estimated for a dense disparity map. Also, the error images are presented, where white values expose the correct estimated disparities, grey values denote invalid pixels, and black values are bad matched pixels (B). One can see that our framework presents a high number of invalid pixels, as well as r200high. It should be noted that the black pixels are minimal compared with the rest, denoting a minimal error in disparities marked as valid, this means a low B value is typically related to higher values in invalid pixels due to higher confidence in such disparities. Therefore, obtaining a more accurate point cloud is suitable for applications where high accuracy is needed, like medical or self-driving vehicles.

TABLE I
AVERAGE INVALID PIXELS AND PERCENTAGE OF BAD MATCHING PIXELS

ine Method	Invalid Pixels	Average B
DAWA	40.17	7.04
DCNN	45.55	0.01
MotionStereo	55.19	3.38
r200high	79.03	2.57
SGBM1	23.98	16.69
Jaccard-DM ine ine	84.62	2.70

To get a complete picture of the performance of disparity mapping methods, in addition to quality metrics, it is necessary to consider the number of pixels that are marked as invalid. This is important for defining parameters in subsequent stages of data interpolation and disparity refinement. For the percentage of invalid pixels, the following interpretation is presented: in the case of a low percentage of invalid pixels, the method tends to estimate and retain a high number of disparity values. Consequently, the occlusions or errors that must be corrected in later stages are at minimum level. For example, algorithms SGBM1 or DAWA demonstrate higher B values. In this scenario, the confidence in disparity values is low. On the other side, a high value of invalid pixels indicates a low quantity of estimated disparities but with higher confidence, as demonstrated for r200high and Jaccard-DM frameworks. Table I shows the percentage of invalid pixels for the methods compared in this study, along with the average Percentage of Bad Matching Pixels (B), where a lower B value is related to a higher percentage of invalid pixels value. The first one presents significant limitations in hardware implementation, which makes their transfer to other applications or programming languages limited. At the same time, our method has demonstrated competitive results in quality with a high percentage of invalid pixels to ensure a high level of confidence.

In Table II, the experimental results for KITTI 2015 Stereo dataset are presented for occluded and non-occluded labeled stereo pairs and for Sparse DM estimated by different match-

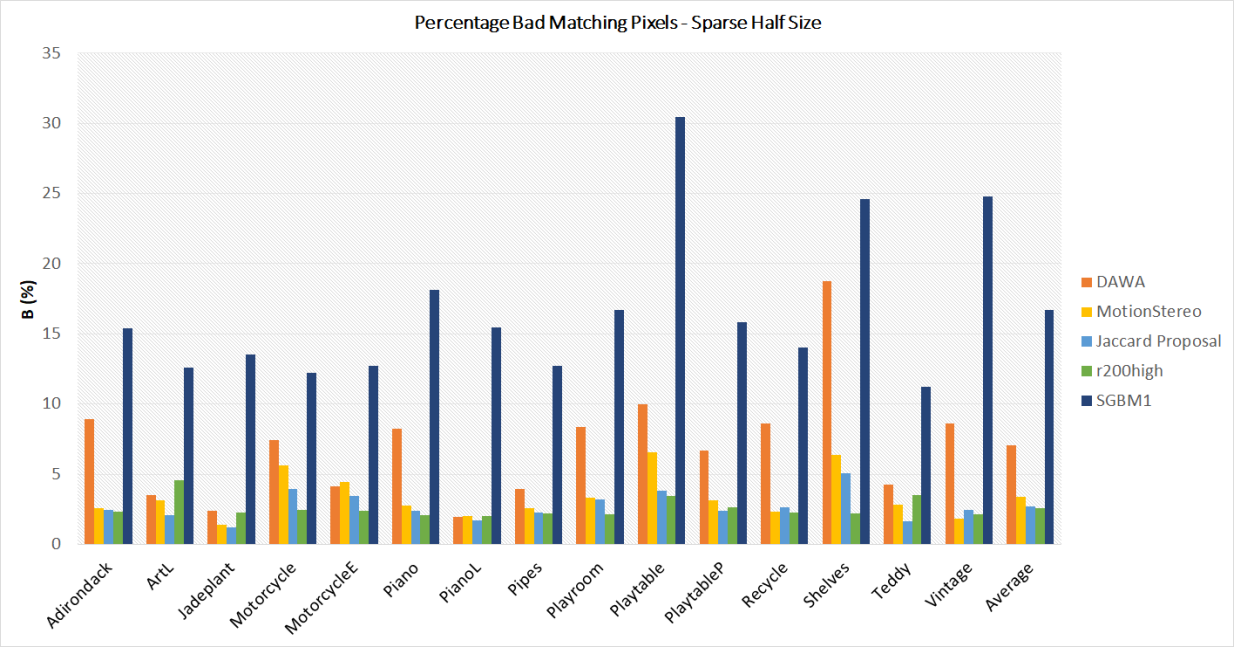


Fig. 4 B results for sparse disparity map estimation for the Middlebury dataset half-size, where our proposal obtains similar values compared to commercial solutions such as r200high, and performing better than deep learning approaches.

TABLE II
AVERAGE RESULTS FOR SPARSE DM ESTIMATION AMONG DIFFERENT MATCHING COST

ine Method	Average B		Invalid pixels		Time (secs)
	occ	noc	occ	noc	
SAD	6.34	6.33	86.13	86.14	**2.37**
NCC	7.01	7.01	**86.04**	**86.05**	3.29
Hamming	1.65	1.65	95.46	95.46	4.61
Jaccard	**0.81**	**0.98**	97.52	97.52	5.34
ine ine					

TABLE III
AVERAGE RESULTS FOR DENSE DM ESTIMATION AMONG DIFFERENT FRAMEWORKS AND RUNNING ENVIRONMENT

ine Method	Average B
SGBM	10.86
Displets	3.43
SGM-Net	3.66
MC-cnn-acrt	3.89
CSPN	**1.74**
Jaccard	16.85
ine ine	

ing costs. As one can see, our proposed framework exceeds state-of-the-art matching costs, where the minimum B value is obtained. Invalid pixels show larger values for our method, which means a major confidence in the DM obtained but with fewer values estimated. In Fig. 6, the experimental results for KITTI 2015 Stereo dataset are shown, where the sparse DM estimated, dense DM, and their respective error image are presented. In the sparse error images, our novel framework shows a larger gray area representing invalid pixels, and white values, which are the correct ones. A lesser quantity of black values (error) indicates higher confidence in the white ones. Consequently, density maps can be reconstructed with more straightforward and faster methods achieving acceptable results. This can be observed in dense DM results, where dense error images shown as blue values indicate correct estimations, and red ones are the incorrect ones. As one can observe, our method estimates dense DM with a minimum number of sparse values but with high confidence in those estimations.

Dense DM results for evaluation metrics are presented in Table III where our method employs two-step basic interpolations. Firstly, five values moving median along the vertical direction is performed, then, nearest-neighbor interpolation is applied. This basic interpolation and the results obtained show the importance of high confidence in sparse DM values, achieving competitive results in comparison with more robust and complex refinement methods, such as ones performed via deep learning with end-to-end results, where the training process only includes stereo images and dense DM from the training dataset, in consequence, deep learning models are generated for each dataset from state-of-the-art. It should be noted that our method is capable of generating dense DM with less than 3% of estimated values in sparse DM with even low-cost and complex interpolation methods. In Fig. 7, Dense DM is presented for some stereo pairs and the compared methods for KITTI Stereo 2015, where subjective results are similar to our approach even with a low cost and low complex refinement disparities stage. It should be noted that the method with the best score at the date of preparation of this study (CSPN) uses a comprehensive and complex refinement procedure based on sparse DM maps with a minimum number of values (500 values) estimated.

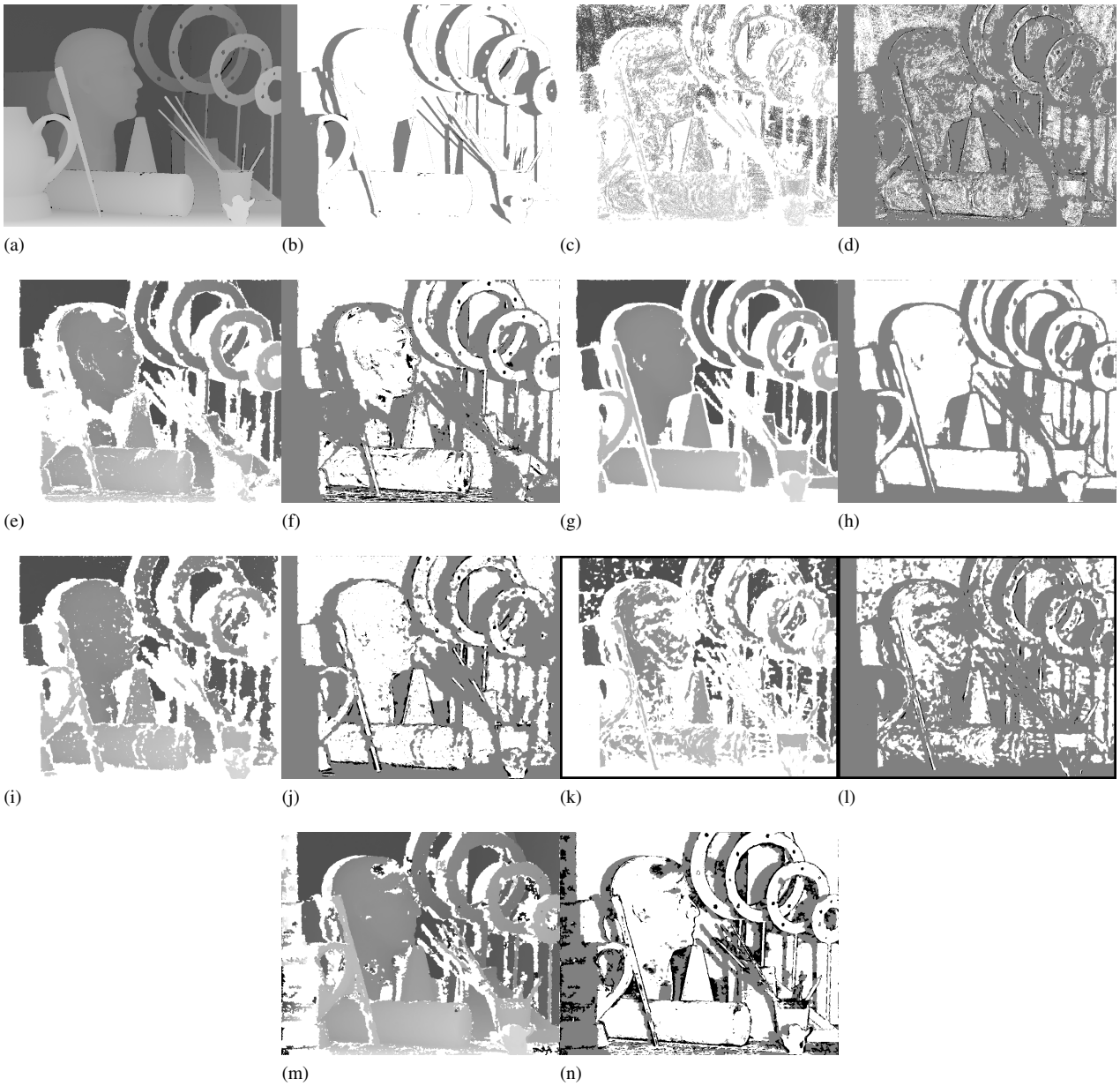


Fig. 5 Sparse disparity maps and error images, respectively from a) & b) Groundtruth and non-occluded mask, c) & d) Proposed framework, e) & f) DAWA, g) & h) DCNN, i) & j) MotionStereo, k) & l) r200high and m) & n) SGBM1.

V. Conclusions

In this paper, a comparison between different methodologies for the sparse disparity maps estimation was presented, where the main contributions of the novel method are the following: first, in the design of the novel framework, a customized Census transform was used with the capacity to be implemented on GPUs; second, the proposition to use Jaccard's distance as a similarity measure to resolve the stereo matching problem that is also implemented on GPUs, demonstrating an improvement in terms of accuracy and confidence in the estimated DM. Finally, quantitative and qualitative results show that higher confidence values used to obtain a more accurate sparse DM with fewer values can generate dense DM with even low-cost and state-of-the-art filtering techniques. Objective and subjective comparisons in terms of quality presented in this study have justified the competitiveness of the novel method in comparison with better existing frameworks, where the Jaccard Index can be employed as a similarity measure for disparity estimation in applications where high-depth accuracy is highly regarded above the continuity or pixel quantity estimations, also without the need of training datasets, which is a significant advantage when a vision solution must be replicated and implemented on specific hardware. As future research directions, some queries need to be addressed to improve the current proposal:

Pursue a real-time implementation in a lower-level programming language such as C/C++ or Rust

Define a strategy to estimate confidence values prior to the

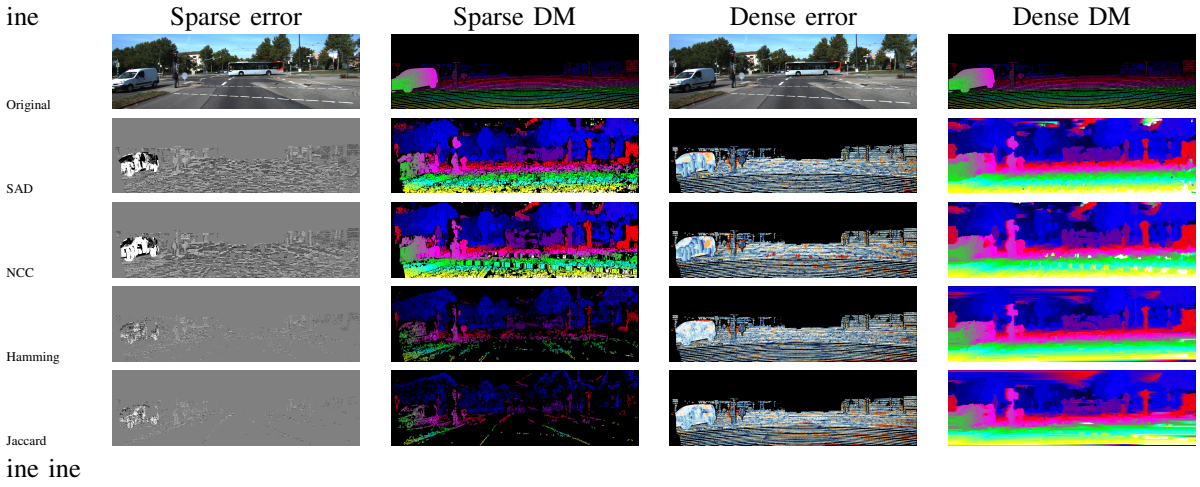


Fig. 6 Results for training image 17 from KITTI 2015 dataset with different matching costs.

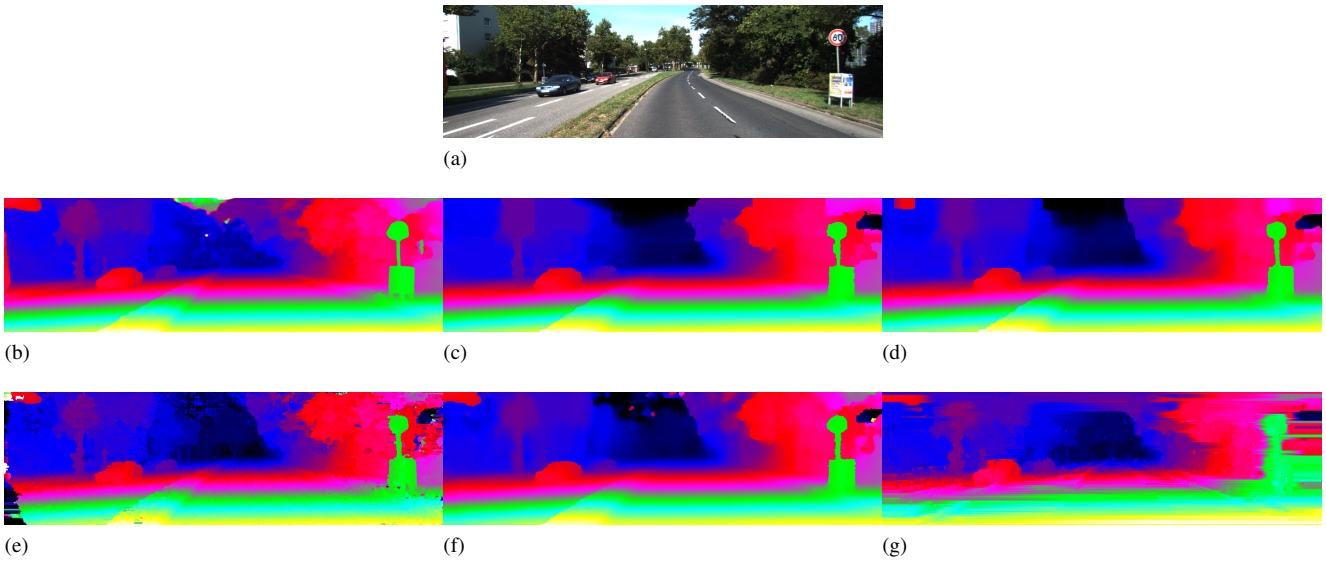


Fig. 7 Qualitative comparison with dense disparity maps for image 12 from testing KITTI dataset, a) Original b) CSPN c) Displets d) MC-cnn-acrt e) SGBM f) SGM-Net g) Jaccard.

cost function to reduce computational operations for each DM estimated

Evaluate with newer state-of-the-art datasets and test with more disparity accuracy demanding applications

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.

[2] S. Trejo, K. Martinez, and G. Flores, "Depth map estimation methodology for detecting free-obstacle navigation areas," in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 916–922, IEEE, 2019.

[3] J.-N. Zhang, Q.-X. Su, P.-Y. Liu, H.-Y. Ge, and Z.-F. Zhang, "Mudeepnet: Unsupervised learning of dense depth, optical flow and camera pose using multi-view consistency loss," *International Journal of Control, Automation and Systems*, vol. 17, no. 10, pp. 2586–2596, 2019.

[4] R. Fan, X. Ai, and N. Dahnoun, "Road surface 3d reconstruction based on dense subpixel disparity map estimation," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3025–3035, 2018.

[5] B.-S. Shin, X. Mou, W. Mou, and H. Wang, "Vision-based navigation of an unmanned surface vehicle with object detection and tracking abilities," *Machine Vision and Applications*, vol. 29, no. 1, pp. 95–112, 2018.

[6] L. Ting and D. Yuelin, "A novel method of human tracking based on stereo vision," in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, pp. 883–889, IEEE, 2018.

[7] J. S. Peixoto, A. R. Cukla, M. A. de Souza Leite Cuadros, D. Welfer, and D. F. Tello Gamarra, "Gesture Recognition using FastDTW and Deep Learning Methods in the MSRC-12 and the NTU RGB+D Databases," *IEEE Latin America Transactions*, vol. 20, pp. 2189–2195, aug 2022.

[8] K. Batsos and P. Mordohai, "Recresnet: A recurrent residual cnn architecture for disparity map enhancement," in *2018 International Conference on 3D Vision (3DV)*, pp. 238–247, IEEE, 2018.

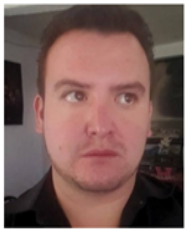
[9] S. J. Lee, H. Choi, and S. S. Hwang, "Real-time depth estimation using recurrent cnn with sparse depth cues for slam system," *International Journal of Control, Automation and Systems*, vol. 18, no. 1, pp. 206–216, 2020.

[10] C. Lin, Y. Li, G. Xu, and Y. Cao, "Optimizing ZNCC calculation in binocular stereo matching," *Signal Processing: Image Communication*, vol. 52, pp. 64–73, 2017.

[11] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, vol. 30, no. 2, pp. 328–341, 2007.
[12] J. Valentin, A. Kowdle, J. T. Barron, N. Wadhwa, M. Dzitsiuk, M. Schoenberg, V. Verma, A. Csaszar, E. Turner, I. Dryanovski, *et al.*, "Depth from motion for smartphone ar," *ACM Transactions on Graphics (ToG)*, vol. 37, no. 6, pp. 1–19, 2018.
[13] W. Mao, M. Wang, J. Zhou, and M. Gong, "Semi-dense stereo matching using dual cnns," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1588–1597, IEEE, 2019.
[14] J. Zbontar, Y. LeCun, *et al.*, "Stereo matching by training a convolutional neural network to compare image patches.," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2287–2318, 2016.
[15] J. Navarro and A. Buades, "Semi-dense and robust image registration by shift adapted weighted aggregation and variational completion," *Image and Vision Computing*, vol. 89, pp. 258–275, 2019.
[16] L. Keselman, J. Iselin Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, "Intel realsense stereoscopic depth cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–10, 2017.
[17] F. Guney and A. Geiger, "Displets: Resolving stereo ambiguities using object knowledge," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4165–4175, 2015.
[18] M. S. Hamid, N. Abd Manap, R. A. Hamzah, and A. F. Kadmin, "Stereo matching algorithm based on deep learning: A survey," *Journal of King Saud University-Computer and Information Sciences*, 2020.
[19] A. Seki and M. Pollefeys, "SGM-nets: Semi-global matching with neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 231–240, 2017.
[20] X. Cheng, P. Wang, and R. Yang, "Learning depth with convolutional spatial propagation network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2361–2379, 2019.
[21] V. Gonzalez-Huitron, V. Ponomaryov, E. Ramos-Diaz, and S. Sadovnychiy, "Parallel framework for dense disparity map estimation using hamming distance," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 231–238, 2018.
[22] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European conference on computer vision*, pp. 151–158, Springer, 1994.
[23] M. Rahman, S. Rahman, M. Shoyaib, *et al.*, "MCCT: a multi-channel complementary census transform for image classification," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 281–289, 2018.
[24] C. Singh, E. Walia, and K. P. Kaur, "Color texture description with novel local binary patterns for effective image retrieval," *Pattern recognition*, vol. 76, pp. 50–68, 2018.
[25] S. Cervantes Alvarez, A. Mexicano Santoyo, J. A. Cervantes, R. Rodríguez, and J. Fuentes Pacheco, "Binary Pattern Descriptors for Scene Classification," *IEEE Latin America Transactions*, vol. 18, pp. 83–91, mar 2020.
[26] S. Kosub, "A note on the triangle inequality for the jaccard distance," *Pattern Recognition Letters*, vol. 120, pp. 36–38, 2019.
[27] V. Verma and R. K. Aggarwal, "A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective," *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1–16, 2020.
[28] R. H. Bhalerao, S. S. Gedam, and K. M. Buddhiraju, "Modified Dual Winner Takes All Approach for Tri-Stereo Image Matching Using Disparity Space Images," *Journal of the Indian Society of Remote Sensing*, vol. 45, no. 1, pp. 45–54, 2017.
[29] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *German Conference on Pattern Recognition*, pp. 31–42, Springer, 2014.
[30] M. Menze, C. Heipke, and A. Geiger, "Joint 3D estimation of vehicles and scene flow," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, p. 427, 2015.

Victor Alejandro Gonzalez-Huitron attained his Ph.D. in communications and electronics in 2017 and M.S. degree in engineering in microelectronics in 2013, from Instituto Politecnico Nacional. He is currently working as a professor and researcher at TecNM. His research interests include computer vision, image processing and digital signal processing.



Abraham Efraim Rodriguez-Mata He obtained a degree in Chemical Engineering in 2009, Master and PhD in Automatic Control at Cinvestav of IPN Mexico. He is currently a researcher is TecNM campus Chihuahua and a member of the National System of Researchers (SNI) of Mexico. He is interesting in robust control, active disturbance rejection and nonlinear observer design, artificial intelligence applications to various engineering problems.

Guillermo Valencia-Palomo received his PhD in Automatic Control and Systems Engineering from the University of Sheffield, U.K., in 2010. Since 2010, he has been with Tecnológico Nacional de México, IT Hermosillo. He is Associate Editor of IEEE Access, IEEE Latin America Transactions, Int. J. of Aerospace Engineering

Rogelio E.Baray Arana received the BSc(1987) and the MSc.(1990) degrees in electrical engineering from the Chihuahua Institute of Technology, Chihuahua, Chih., Mexico. His professional experience includes development of projects and consultancy for several companies. He currently works as a Research Professor at the Chihuahua Institute of Technology, in Chihuahua, Chih.

Isidro Robledo-Vega Isidro Robledo-Vega received the BSc degree in industrial engineering in electronics in 1989 and the MSc degree in electronics engineering with a computer science option in 1996 from the Technological Institute of Chihuahua, Mexico, and the PhD degree in computer science.

Leonel Ernesto Amabilis-Sosa PhD in Environmental Engineering, graduated from Mexico's highest university with honors. He is also an expert in multiparametric statistics applied to various engineering problems. He is currently interested in applications of artificial intelligence and machine learning to various engineering phenomena related to environmental research.