computing@computingonline.net
www.computingonline.net

Print ISSN 1727-6209
On-line ISSN 2312-5381
International Journal of Computing

PRIMITIVE VISUAL RELATION FEATURE DESCRIPTOR APPLIED TO STEREO VISION

Dario Rosas, Volodymyr Ponomaryov, Rogelio Reyes-Reyes

SEPI Esime Culhuacan, Instituto Politécnico Nacional, Av. Santa Anna 1000, Mexico City, Mexico,
drosasm0701@alumno.ipn.mx, vponomar@ipn.mx, rreyesre@ipn.mx

Paper history:
Received 23 August 2018
Received in revised form 19 Sep. 2018
Accepted 20 September 2018
Available online 30 September 2018

Keywords:
Image Local Descriptor;
Dense Depth Map;
Visual Primitives;
Vision Stereo;
PCA;
GPU.

Abstract: In this study, we present a novel local image descriptor, which is very efficient to compute densely, with semantic information based on visual primitives and relations between them, namely, coplanarity, cocolority, distance and angle. The designed feature descriptor covers both geometric and appearance information. The proposed descriptor has demonstrated its ability to compute dense depth maps from image pairs with a good performance evaluated by the Bad Matched Pixel criterion. Since novel descriptor is very high dimensional, we show that a compact descriptor can be sustitable. An analysis of size reduction was performed in order to reduce the computational complexity with no lose of quality by using different algorithms like max-min or PCA. This novel descriptor has a better results than state-of-the-art methods in stereo vision task. Also, an implementation in GPU hardware is presented performing time reduction using a NVIDIA R GeForce R GT640 graphic card and Matlab over a PC with Windows 10.

1. INTRODUCTION

Computer vision is an interdisciplinary field that seeks to perform process as similar to human vision, employing methods that can understand digital images and video, such as acquisition, processing and analyzing. Some tasks in computer vision include segmentation, object detection and identification by extracting high-dimension data from the real word and transforming data using descriptor that can interface with other processes.

Stereo vision is one of the most active research areas in the computer vision. Therefore, a variety of solutions and variations of existing methods have been presented for specific needs or requirements. The goal of stereo vision is to estimate the depth of a scene by disparity maps, matching similarities from a pair of images. A taxonomy of existing stereo algorithms that allows the dissection and comparison of individual algorithm components is presented in [1]. This taxonomy is based on four steps that stereo algorithms typically perform:

1. Matching Cost
2. Cost aggregation
3. Disparity computation
4. Disparity refinement

The sequence of the steps depends on the type of an algorithm, where local algorithms typically follow the steps 1,2,3 but some others combine steps 1,2 and use matching costs based on the support region. On the other hand, global algorithms do not perform an aggregation step but rather seek a disparity assignment (step 3) that minimizes a global cost function (step 1).

Some authors focus their efforts in one or more steps, depending on particulars goals. Difference matching cost have been studied; the most common is based in pixel difference and includes squared intensity differences (SAD) and absolute intensity differences (AD); also, in the video processing field, the mean absolute difference (MAD) and mean-squared error are more frequently used. Other approaches use gradient-based measures and non-parametric measures, such as rank and census transform. It is also possible to perform a preprocessing step, using histogram equalization or Gaussian filters.

Local and windows-based methods aggregate the matching cost over a support region employing squared windows or Gaussian convolutions, shiftable windows or windows with adaptive sizes.

Algorithms can be classified by the disparity computation step, local methods, global methods and dynamic program methods.

Local methods emphasize on the matching cost computation and cost aggregation steps, computing the final disparity by a "*winner take all*" methods. While global methods often skip the aggregation, they formulate an energy minimization framework. The objective is to find a disparity function that minimizes a global energy. More recently, max-flow and graph-cut methods have been proposed to solve a special class of global optimization problem. Dynamic programming methods find the global minimum for independent scanlines as an optimization problem. These approaches work by computing the minimum-cost path through the matrix of all pairwise matching costs between two corresponding scanlines.

Most state-of-art methods rely on local measure to estimate the similarity of pixels across images and then on impose global shape constraints using some aggregation cost such as dynamic programming [2], level sets [3], graph-cuts [4], PDE [5], or EM [6].

Image descriptors can be classified as global and local descriptor. 2D local features such as SIFT are commonly used in object detection task, while global descriptors, such as visual contours have been proved to provide a semi-global overview of a scene and give more information than local features about the shape of an object, also, they are flexible enough for task such as classification and recognition.

2D visual contours and their relations have been used in computer vision and robotics in various contexts; for example, in contour relations [7], they are used as features for object recognition. Similarly, Henricsson [8] uses geometrical relations such as proximity, curvilinearity and symmetry between contours to describe objects based on combinations of these relations. Contours in computer vision are important because they provide a means to group the local features together as well as saving the spatial relations between these contours [9].

1.1 RELATED WORKS

Let present brief review of similar papers. Local image descriptor has already been used in dense matching, although in a more traditional way to match only sparse pixels that are feature points. More of the existing stereo vision algorithms are based on pixel difference and present matching cost and disparity refinement. For example, in [10] (SAD+Wavelet) techniques are performed with aggregation cost using a multilevel disparity map (DM) approach and matching cost, such as SAD combined with a wavelet; finally, an adaptive filter is used during the postprocessing step. A variation of this approach is presented in [11] (MDEC+SSIM) technique, where a pyramid DM estimation is used with SSIM measure as the matching cost. Methods based on a global approach usually present better performance at high computational cost, such as graph-cuts, belief propagation, or semi-global matching. Paper [12] presents an algorithm based on Randow Walk with Restar Algorithm (RWR) updating the matching cost aggregated into superpixels.

Paper [13] presents a hybrid method using transition pixel values in horizontal and vertical orientations and a polynomial curve fitting, showing robustness under radically different radiometric conditions. This approach uses a "*winner take all*" disparity computation. Yong in [14] presents a feature detector using SURF, SIFT, and HOG algorithms to find interesting points and to evaluate the quality of the points detected; then, a regression of the multimodal image is used to compute the disparity map.

Promising Daisy descriptor [15] advocates an approach based on SIFT and GLOH, it has been designed to obtain robustness to perspective and lighting changes and have been proved to be optimal for dense matching. Another local descriptor mainly used for image correlation is the Scale Invariant Descriptor (SID) [20]. This descriptor uses a combination of log-polar sampling with spatially varying filtering that converts image scaling and rotation into translations. Scale invariance is achieved by taking the Fourier Transform Modulus (FTM) of the transformed signals because the FTM is translation invariant.

In this paper, we propose novel local feature descriptor based on visual primitives (VP). Additionally, the semantic information is obtained using the relation between visual primitives (VPR). These relations are cocolority, coplanarity, normal distance and the angle between them. The principal difference between of novel descriptor and existing approaches is that it can extract structural and semantic information from an image; additionally, the designed descriptor has demonstrated robustness against radiometric distortions. Also, a feature descriptor size reduction is applied in order to save computational cost and memory. The designed descriptor is implemented in a GPU to accelerate processing speed, which is important for real-time applications. The designed descriptor is used with traditional depth map estimation algorithms, confirming their performance via traditional quality metrics.

The remainder of the paper is organized as follows. In Sect. 2, the novel local feature descriptor is explained and the framework for disparity estimation is presented with the dimension

reduction. In the Sect. 3 the experimental results are presented. Finally, Sect. 4 concludes this study by discussing the results of the proposed approach.

2. VISUAL PRIMITIVE RELATION DESCRIPTOR

The framework of designed descriptor is explained as follows; for a given input image in RGB space, we first should compute a space color conversion from the RGB color space to the CIELab space and then apply the monogenic filter in the L channel. This filter gives the information about visual primitives: magnitude, phase, and orientation. Next, this information should be used to obtain the relation between them and to form the feature vector. We take advantage of the two degrees of freedom when designing monogenic filters to extract information. To measure quality performance, the designed feature descriptor is used as a metric for stereo matching similarities across a pair of images. Then, this measure is used in the traditional block matching algorithm to estimate a depth map.

2.1 VISUAL PRIMITIVES

The visual primitives are a set of visual descriptors. These primitives [21] describe edge structures by means of several properties that are relevant for edges only. They have been used to formalize different contexts in visual scenes, as well as 6D motion and 3D spatial context. These descriptors have been employed in several applications such as learning of object representations, pose estimation, motion estimation, and vision-based grasping.

The primitives express explicitly important structural properties of the edges such as local orientation, phase, color, and motion; this information is encoded in a multi-dimensional feature, where geometric and appearance cues are separated. Information about these different properties can be extracted from images by applying a variety of linear and non-linear local filtering operations.

Current work makes use of the monogenic signal presented by Felsberg and Sommer [36]. It uses a bandpass filter that is radially symmetric around the origin ('even') in both the frequency domain and image domain. The Log-Gabor is used as even filter, as follows:

$$G_e(w) = \exp - \frac{\log\left(\frac{|w|}{w_0}\right)^2}{2(\log \sigma_0)^2}, \quad (1)$$

$$F_{o1} = i\frac{w_x}{|w|}F(w), \quad (2)$$

$$F_{o2} = i\frac{w_y}{|w|}F(w), \quad (3)$$

$$f_m(x) = [f_e(x); f_{o1}(x); f_{o2}(x)]. \quad (4)$$

Two odd parts of the filter, F_{O1} and F_{O2}, using the Riesz transform, are presented in eqs. 2 and 3. Each of the two resulting filters are odd-symmetric, with the axis of symmetry along the two image axes. After filtering, we can present the monogenic signal as a combination of the three parts (one even, two odd) as a vector shown in eq. 4:

$$p_i = \big(A(x_i), \theta(x_i), \varphi(x_i)\big). \quad (5)$$

These three components can be explained as spherical polar coordinate system, using the radius, elevation angle, and azimuthal angle. The local amplitude is the radial part of the representation A(x); the local phase is found from the angle between the even part and the combined odd part φ; and the local orientation θ is the orientation of the odd filter and represents the dominant direction in an image at point x. The visual primitive is a vector as shown in eq. 3. Fig. 1 presents an example of the visual primitives from an image to obtain visual primitives.

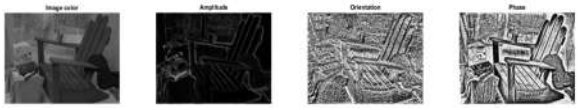


Figure 1- Visual primitives from the image Airdock.

2.2 RELATIONS BETWEEN VISUAL PRIMITIVES

Since primitives carry geometrical and appearance information, the primitives have attributes such as the mean color, position and orientation. The mean color is defined in the CIELab color space because of the statistically less correlated behavior of an image in the CIELab space. These attributes together with the geometrical and visual of primitives give relations between primitives that can be used within the context of various reasoning processes. Let describe certain primitive relations.

Angle: The angle between two primitives is defined by using the orientations of the primitives as:

$$RA(p_i, p_j) = arcos\left(\frac{p_i . p_j}{|p_i||p_j|}\right). \quad (6)$$

Normal Distance: The normal distance between two primitives is defined by the distance from one primitive's position to the line created by the others primitive orientation and position. Therefore, the distance between the ith and jth primitives in a scene is defined as:

$$RND\left(p_i, p_j\right) = |W_i - (W_i U_i) U_i|, \quad (7)$$

Coplanarity: The coplanarity of entities can be measured by their elongation with a common plane. We define the coplanarity between two primitives as the mean angle between a common plane and the best-fit lines of the primitive. Therefore, the coplanarity between the ith and jth primitive in the scene can be defined as follows:

$$RCP(p_i, p_j) = \frac{1}{2}\left(\pi - arcos\left(\frac{np_i}{|n||p_i|}\right) - arcos\left(\frac{np_j}{|n||p_j|}\right)\right). \quad (8)$$

Cocolority: The cocolority between two primitives is defined as the color difference between the colors on the primitive. The color difference is calculated in such a way:

$$RC = \sqrt{\left(L_i - L_j\right)^2 + \left(a_i - a_j\right)^2 + \left(b_i - b_j\right)}. \quad (9)$$

The relations between primitives are illustrated in Fig. 2.

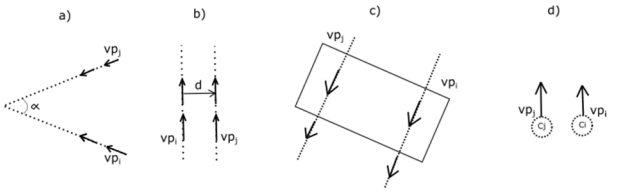


Figure 2 -Illustration of primitives relations. a) Angle; b) Normal distance; c) Coplanarity; d)Cocolority.

2.3 VISUAL PRIMITIVE RELATIONS DESCRIPTOR

We present a formal definition of the designed visual primitive relations descriptor (VPR). For a given input color I_{RGB}, we convert it from the RGB to the CIELab I_{Lab} space since the cocolority relation is defined in eq. 9. The monogenic signal is performed using S scales and σ Gaussian kernels, so S*σ filters are used. Each filter is then convolved with the L channel of image I_L obtaining 3*S*σ different components of the monogenic signal: $f_e\{s_i, \sigma_j\}$, $f_{O1}\{s_i, \sigma_j\}$, $f_{O2}\{s_i, \sigma_j\}$, calculated at s_i scale wavelengths and σ_j Gaussian kernel with i=1...,S and j=1...,Σ. We obtain the visual primitives with different responses.

At each pixel location, the designed descriptor consists of a vector made of values from the visual primitive relations located on a squared window centered on the location. Let to h(x, y) represent the vector formed of the values at location (x, y) in an image:

$$h_{RA}(x,y) = \left[RA_{x,y}\{s_i, \sigma_j\}(1,1), \ldots, RA_{x,y}\{s_i, \sigma_j\}(u,v)\right]^T, \quad (10)$$

where $RA_{x,y}$ is the angle relation described in eq. 6 between location (x, y) and location (u, v) in the neighborhood inside window W from the filter chosen response at scale s_i and Gaussian kernel σ.

We normalize this vector to unit and to denote the normalized vectors by $\tilde{h}(x, y)$. If σ is the value of Gaussian kernels used and S is the number of scales of the monogenic signal, the feature vector of the angle relation $D_A(x_0, y_0)$ for a location (x_0, y_0) is defined as the concatenation of h vectors:

$$D^{RA}(x_0, y_0) = \begin{bmatrix} \widetilde{h_{RA}}(x_0, y_0)_{s_1}^{\sigma_1 T}, \ldots, \widetilde{h_{RA}}(x_0, y_0)_{s_S}^{\sigma_1 T} \\ \vdots \\ \widetilde{h_{RA}}(x_0, y_0)_{s_1}^{\Sigma\ T}, \ldots, \widetilde{h_{RA}}(x_0, y_0)_{s_S}^{\Sigma\ T} \end{bmatrix}. \quad (12)$$

For the normal distance and coplanar relations, we perform the same structure as the angle relation. The relationship vector $h_{ND}(x, y)$ and $h_{CP}(x, y)$ at point (x, y) is shown as follows:

$$h_{ND}(x,y) = \left[RND_{x,y}\{s_i, \sigma_j\}(1,1), \ldots, RND_{x,y}\{s_i, \sigma_j\}(u,v)\right], \quad (13)$$

$$h_{CP}(x,y) = \left[RCP_{x,y}\{s_i, \sigma_j\}(1,1), \ldots, RCP_{x,y}\{s_i, \sigma_j\}(u,v)\right. \quad (14)$$

where $RND(x, y)$is the normal distance relation between the location (x, y) and location (u, v) described in eq. 7 and $RCP(x, y)$ is the coplanar relation shown in the eq. 8. The feature vector of the normal distance relation $D_{ND}(x_0, y_0)$ and D_{RCP} is concatenated as follows:

$$D_{ND}(x_0, y_0) = \begin{bmatrix} \tilde{h}(x_0, y_0)_{s_1}^{\sigma_1 T}, \ldots, \widetilde{h_{ND}}(x_0, y_0)_{s_S}^{\sigma_1 T} \\ \vdots \\ \widetilde{h_{ND}}(x_0, y_0)_{s_1}^{\Sigma\ T}, \ldots, \widetilde{h_{ND}}(x_0, y_0)_{s_S}^{\Sigma\ T} \end{bmatrix}, \quad (15)$$

$$D_{CP}(x_0, y_0) = \begin{bmatrix} \tilde{h}(x_0, y_0)_{s_1}^{\sigma_1 T}, \ldots, \widetilde{h_{CP}}(x_0, y_0)_{s_S}^{\sigma_1 T} \\ \vdots \\ \tilde{h}(x_0, y_0)_{s_1}^{\Sigma\ T}, \ldots, \widetilde{h_{CP}}(x_0, y_0)_{s_S}^{\Sigma\ T} \end{bmatrix}. \quad (16)$$

Because the cocolority relation does not depend on the filter parameters, the feature vector is defined as:

$$D_C(x_0, y_0) = \left[RC_{x,y}(1,1), \ldots, RC_{x,y}(u,v)\right]^T, \quad (17)$$

where $RC(x, y)$ is the cocolority relation between the location (x, y) and the neighborhood (u, v) inside the window W.

The final descriptor $D_{VP}(x_0, y_0)$ for location (x_0, y_0) is defined as the concatenation of the D vectors of primitives relations:

$$D_{VP}(x_0, y_0) = \begin{bmatrix} D_{RA}(x_0, y_0), D_{ND}(x_0, y_0), D_{CP}(x_0, y_0), \\ D_C(x_0, y_0) \end{bmatrix}^T, \quad (18)$$

The order of the elements in the vector D_{VP} is selected using PCA analysis. The vector is sorted using the eigenvalues from maximum to minimum, forming the elements of the descriptor as follows:

$$D_{VP}(x_0, y_0) = \begin{bmatrix} D_{RA}(x_0, y_0), D_C(x_0, y_0), D_{ND}(x_0, y_0), \\ D_{CP}(x_0, y_0) \end{bmatrix}^T, \quad (19)$$

2.4 Computational Complexity

VPR descriptor is parameterized by the number of scales S, the value of Gaussian kernels σ, and the size of the rectangular window W. Assuming that the image has P pixels, the filters in the frequency domain with size P are created. These filters are convolved with the image spectrum to produce different response versions of the visual primitive components.

Therefore, at each location of an image, the relations between primitives are computed inside a block window W, where it should be used $(2W + 1)^2$ for each primitive relation. Therefore, computing all the descriptors of an image requires: $4\Sigma S$ convolutions, and $[(2W + 1)^2 - 1] \times P \times \Sigma \times S$ multiplications.

Table 1. Computational complexity of VPR descriptor.

	Daisy	VPR Descriptor
Conv.	$2H \times Q + 1(1D)$	$4 \times \Sigma \times S$
Sampling	$P \times (Q \times T + 1)$	-
Operations	$2P \times H + P \times H$	$[(W \times x + 1)^2 - 1] \times \Sigma$

Table 1 shows a complexity comparison between the Daisy descriptor and VPR descriptor. The Daisy descriptor parameters are: H is the number of bins in a histogram, Q is the number of convolved orientations layers, and T is the number of histograms at a single layer. One can see that the VPR descriptor has a quadratic growth in contrast with DAISY.

Table 2. Number of operations required for the proposed VPR and Daisy descriptor.

	Daisy	VPR Descriptor
Convolutions	$49(1D)$	16
Sampling	1,562,500	-
Operations	1,000,000	20,000,000

Daisy feature descriptor proposes the use of the following parameters: $H = 8$, $Q = 3$, and $T = 8$; So, we can see that DAISY requires 49 convolutions by 1D direction, 25 sampling per pixel and 24 operations by pixel. If we insert parameters $W = 4$, $S = 2$, σ=2, it can be observed that the designed descriptor requires $192P$ operations and 16 convolutions. Novel descriptor have advantage in comparison with Daisy one, it can be easily parallelized because each relation at point x, y is calculated separately. A summary of this is shown in Table 2 considering an image with size 250×250.

The designed framework that appears to demonstrate the competitive quality performance was implemented using an Intel Core i7-3770 CPU at 3.40Ghz with 8 GB of RAM memory. Time values were computed using this CPU. Table 3 presents the processing time values for the Daisy and VPR descriptor.

The advantage of VPR descriptor is that can be implemented using parallel programing. The main process of our work is based on the visual primitives and the monogenic signal, so they should be computed in GPU. We use the Felsberg's monogenic filters described in eqs. 1 and 2, that allow to compute the visual primitive components by Fast Fourier Transform (FFT). The filters are computed in the frequency and in the image spectrum domains in order to obtain the visual primitives components. Once we obtained the visual primitives, the relations can be computed in parallel matter.

Let perform a kernel by each block window W, this means that we compute the relation between primitives for each location (x, y) in one kernel. Each a kernel should perform $[(2W + 1)^2 - 1] \times \Sigma \times S$. We applied this method for each primitive relation that should be computed.

The time for the primitive calculation is for only one scale and for one sigma value. The calculation time for a window size 3x3 is $0.084ms$ for each visual primitive relation. Finally, the total time value to compute designed descriptor result is $0.76ms$ and it is shown in Table 3, while Daisy descriptor time in GPU computes result during $9.96ms$ as it was presented in [24].

Table 3. Processing times required for the proposed VPR and for Daisy descriptors.

Technique	Time CPU(Seconds)	Time GPU (Miliseconds)
Daisy	0.25	9.96
VPR	47.04	0.76

We emphasize that the novel descriptor VPR based on visual primitives and their relations can obtain the structure and semantic information of an image. The novel descriptor is robust against radiometric distortions such as illumination and exposure changes. Additionally, VPR descriptor can be used together with state-of-the-art methods to improve quality. Even the computation cost could be

higher than that of Daisy descriptor; the quality results shows that it is worth it.

2.5 Feature Descriptor Reduction

The descriptor's size can be obtained from eq. 19, where is $[\Sigma \times (2W+1)^2, 3S+1]$. As is shown, the descriptor's size grows exponential, using the parameters $S = 2, \Sigma = 4$ and $W = 4$ the descriptor size will be $[324,7]$ for each pixel in an image. If the image size is $[490,720]$, we have $685{,}843{,}200$ feature values, and using double format, we will need up 4,677,684,480 bytes. In order to reduce this size, different reduction algorithms are applied, such as, statistic algorithms, direction approach and PCA.

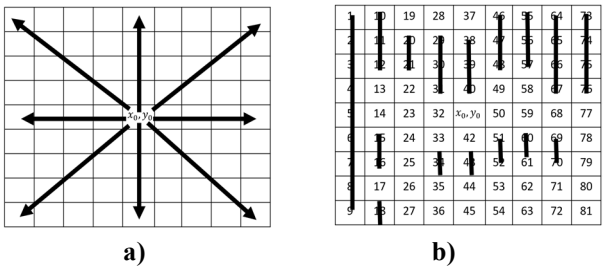
For statistic algorithms, at each vector relation descriptor D_R, we applied an operation O that can be max, min, or mean algorithm, so the final descriptor is formed as follows:

$$D_{VP} = [O(D_{RA}), O(D_C), O(D_{ND}), O(D_{CP})] \quad (20)$$

and the dimension is reduced to $[\Sigma, 3S+1]$.

Second approach to reduce the descriptor dimension consists in computing the relations around specific directions. Let consider the center of the window x_0, y_0, and calculating the relation between the point (x_0, y_0) along eight directions (North, South, East, West, NE, NW, SW, SE) as shown in the Fig. 3a. By this way, the final dimension is $[\Sigma \times 8 \times W, 3S+1]$.

At least, a PCA analysis should perform along the location of the descriptors. The eigenvalues are calculated for each location inside the window and select the half top values. The locations used to compute the relations are shown in the Fig. 3b. The dimensions after applying PCA reduction are $\left[\Sigma \times \frac{(2W+1)^2}{2}, 3S+1\right]$.



a) b)

Figure 3 -Location calculated for image reduction a)8 directions; b) PCA.

The computation complexity reduction for each algorithm is summered in the Table 4. One can see observing this table that the statistic methods get the best reduction and PCA method only can reduce the size to the half.

Table 4. Computational complexity reduction by different algorithms.

	Max/Min/ Meann	8 Dirct.	PCA
Operations	1,750,000	8,000,000	10,000,000

3. EXPIREMENTS AND RESULTS

In this section, we discuss the results of the experiments that we performed to justify the performance of the designed descriptor in the reconstruction of the disparity maps. First, to understand the influence of VPR parameters, we perform a parameter sweep experiment. Then, a comparison of the designed descriptor and other descriptors in depth map estimation is performed.

3.1 Data

To evaluate the performance of the proposed method, we used the dataset Middlebury [23]. The 2014 dataset was employed for testing and comparing with disparity maps. These datasets contain up to nine different pair images with their ground truth at full size (width: 1330-1390 pixels, height: 1110 pixels) half size, and one-third size. The 2014 dataset contains 33 image pairs divided into three sets, 10 for training, 10 for testing and 13 additional images without a ground truth provided. Additionally, this dataset presents two views of each image pair, taken under several different illuminations (L) and different exposures (E).

3.2 Quality criteria

In quality analysis, the quantitative metric Percentage of Bad Matching Pixels (B) is employed, justifying the performance of the proposed framework. To compute the selected metric, the ground truth GT for density maps (DM) obtained from the Middlebury Stereo Vision website for each a stereo pair and the DM estimates obtained by proposed descriptor are employed.

The B values are calculated as follows:

$$B = \frac{1}{N}\Sigma_{x,y}(|DM_I(x,y) - GT(x,y)| > \delta_d) \quad (20)$$

where N is the total number of pixels in an image or frame, DM_I is the estimated disparity, and GT is the ground truth. δ is the error threshold difference for each a pixel valuated, commonly used value is 2.0.

3.3 Comparison with other descriptors

To compare the novel feature descriptor with other descriptors, we used the database Middlebury

2014 at a quarter side of the original image. We employed commonly used SID and Daisy descriptors for comparison because Daisy is one of the most cited descriptors among state-of-the-art-methods, and SID is based on a monogenic signal.

Table 5. Bad Percentage Pixel for the dataset Middlebury.

Image	Daisy	SID	VPR	
Adirondack	25.76	17.50	**24.35**	
Jadeplant	**27.58**	52.41	45.36	
Motorcycle	17.06	10.71	**16.32**	
Piano	**22.57**	13.93	23.37	
Pipes	27.10	30.60	**25.26**	
Playroom	24.31	17.05	**22.00**	
Recycle	**18.16**	15.87	19.26	
Image	**Daisy E**	**VPR E**	**Daisy L**	**VPR L**
Adirondack	25.62	**24.69**	40.27	**34.6**
Jadeplant	24.81	**23.20**	28.26	**20.5**
Motorcycle	**16.16**	16.34	25.44	**21.4**
Piano	**22.68**	24.02	33.05	**28.7**
Pipes	26.48	**25.45**	41.96	**39.2**
Playroom	23.78	**21.96**	32.75	**29.7**
Recycle	**22.83**	23.67	29.01	**38.9**

We apply the parameters $S = 2, \sigma = 4$ and $W = 4$, the parameters for SID and Daisy have been chosen according to their respective works. Table 5 shows the results in the tested dataset for these three feature descriptors using traditional block matching to compute the depth map. The first column presents results for the Daisy descriptor. This descriptor demonstrates sufficiently good performance, with B criterion value less than 20%; the worst performance can be seen for the *Jadeplant* image, and the best performance appears in the *Motorcycle* image. The second column is the B for SID; it appears that the descriptor demonstrates better performance, but since the SID descriptor performs image reduction, the disparity map is also reduced, and it cannot be compared directly. Additionally, the SID cannot be performed for large images. The third column presents results for the novel descriptor, where one can see that they are very close to those of Daisy. Next, the columns show the experiment results for testing the Daisy and VPR descriptor in the case of image exposure (E) and lighting (L) changes. Finally, we can conclude that the novel descriptor shows better performance in almost all tested images, even with exposure and lighting changes.

Observing the results for the image Piano, one can see more differences between the Daisy descriptor and the designed one, but these differences cannot be easily seen in the depth map. The Daisy descriptor exhibits lower performance when lighting differences are present, and the VPR descriptor appears to demonstrate robustness against these changes.

As we used traditional block matching for cost aggregation, we cannot resolve the occlusion problems, so for the areas where the depth could not be calculated correctly, most of the differences appear because of occlusion.

Table 6. Bad Percentage Pixel for descriptor dimensions reduction.

Image	Max	Min	Mean	8 Dirs.
Adirondack	48.66	50.90	40.02	28.88
Jadeplant	47.49	47.53	49.35	49.36
Motorcycle	35.01	36.40	30.46	19.65
Piano	45.95	48.29	37.72	27.03
Pipes	46.95	47.59	41.42	28.38
Playroom	41.66	44.07	34.33	24.77
Playtable	41.03	41.56	36.87	30.24
Recycle	46.45	48.57	37.14	26.98
Shelves	40.09	41.19	37.69	30.51
Vintage	56.53	56.04	55.73	61.90
Image	**PCA**	**VPR**		
Adirondack	26.14	24.35		
Jadeplant	48.59	45.36		
Motorcycle	17.73	16.32		
Piano	24.94	23.37		
Pipes	26.62	25.26		
Playroom	23.06	22.00		
Playtable	27.58	19.26		
Recycle	23.06	24.35		
Shelves	28.31	27.55		
Vintage	62.11	63.48		

The Table 6 shows the bad percentage pixel when the dimension descriptor reduction was applied. As it can been seen, for eight directions in case of using PCA method VPR algorithm presents better results in comparison when the statistic approach is applied, even statistic algorithm performs sufficiently bigger reduction. Calculating the relations only at the locations obtained by the eigenvalues give us the best results with a mean quality lost less than 2%. Appling the calculations for eight directions can be obtained good performance with a lost quality around 5%. So, we can conclude that using PCA analysis is the best way to reduce to the half the descriptor dimension but saving quality.

4. CONCLUSIONS

Novel descriptor VPR based on visual primitives and the relations between them has been designed as a preprocessing step in stereo vision algorithms. The performance of the novel descriptor has been tested in the disparity maps computing. The VPR descriptor appears to demonstrate better performance in comparison with the Daisy and SID descriptors; also, it can be used for images of large sizes. Additionally, experiments with different exposure and illumination changes have been performed, demonstrating that VPR descriptor provides better robustness in comparison with other descriptors, in the case of lighting changes, which are more challenging. The improvement in stereo

vision algorithm was performed in the preprocessing step; the novel descriptor demonstrates the ability to improve the quality results when implemented with state-of-the-art methods. Computing our descriptor takes more computational complex using single core, so is important to seek for a faster computation of the descriptor for all image pixels. This could have implications beyond stereo reconstruction because dense computation of ijmage descriptors is fast becoming an important technique in other task, such as object recognition, object detection or facial aging analysis.

5. REFERENCES

[1] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, Vol. 47, Issue 1, pp. 7-42, 2002.

[2] S. Zhu, R. Gao, Z. Li, "Stereo matching algorithm with guided filter and modified dynamic programming," *Multimedia Tools and Applications*, Vol, 76, Issue 1, pp. 199-216, 2017.

[3] O.D. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," *Proceedings of the European Conf. on Computer Vision*, June 1998.

[4] V. Kolmogorov, R. Zabih, "Multi-camera scene reconstruction via graph cuts," *Proceedings of the European Conf. Computer vision*, Springer, Berlin, Heidelberg, pp. 82-96, 2002.

[5] L. Alvarez et al., "Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach," *Journal of Visual Communication and Image Representation*, Vol. 13, Issue1, pp. 3-21, 2002.

[6] C. Strecha, R. Fransens, L. Van Gool, "Combined depth and outlier estimation in multi-view stereo," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, Vol. 2.

[7] X. Wang, J. Keller, P. Gader, "Using spatial relationships as features in object recognition," *Annual Meeting of the North American*, pp. 160-165, 1997.

[8] O. Henricson, "Interfering homogeneous regions from rich image attributes," *Automatic Extraction of Man-Made objects from Aerial and Space Images*, Centro Stefano Franscini Ascona, pp. 13-22, 1991.

[9] J. D. Winter, J. Wagemans, "Contour-based object identification and segmentation: stimuli, norms and data, and software tools," *Behav. Res. Methods Instrum. Comput.*, Vol 36, Issue 4, pp.604-624, 2004.

[10] V. Gonzalez-Huiltron, V. Ponomaryov, "Robust approach for disparity map estimation based on multilevel decomposition," *IEEE Latin America Transactions*, Vol. 14, Issue 6, pp. 2968-2973, 2016.

[11] V. Gonzalez-Huiltron, V. I. Ponomaryov, E. Ramos-Diaz, S. Sadovnychiy, "Parallel Framework for Dense Disparity Map Estimation Using Hamming Distance," *Signal, Image Video Process*., vol. 12, no 2, pp. 231-238, 2016.

[12] S. Lee et al., "Robust stereo matching using adaptive random walk with restart algorithm," *Image and Vision Computing*, Vol. 37, pp. 1-11, 2015.

[13] S. Birchfield, C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 20, Issue 4, pp401-406, 1998.

[14] X. Yong et al., "Descriptor evaluation and feature regression for multimodal image analysis," *Machine Vision and Applications*, Vol. 26, Issue 7, pp. 975-990, 2015.

[15] E. Tola, V. Lepetit, P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol 32, Issue 5, pp. 815-830, 2010.

[20] I. Kokkinos, A. Yuille, "Scale invariance without scale selection," UCLA: Department of Statistics. [Online]. https://escholarship.org/uc/item/9m811940, 2008.

[21] N. Pugeault, F. Wörgötter, N. Krüger, "Visual primitives: local, condensed, semantically rich visual descriptors and their applications in robotics," *International Journal of Humanoid Robotics*, Vol. 7, No. pp. 379-405, 2010.

[22] M. Felsberg, G. Sommer, "The monogenic signal," *IEEE Trans. Signal Processing*, Vol. 49, Issue 12, pp. 3136-3144, 2001.

[23] D. Scharstein, R. Szeliski, and H. Hirschmller, *Middlebury Stereo Vision Page*. [Online]. Available: vision.middlebury.edu/stereo/, 2016.

[24] E. Iranmehr, S. Kasaei, "An efficient FPGA implementation of DAISY descriptor based on pipeline and multicycle architectures," *Int. Journal of Mechatronics*, Vol. 8, Issue 27, pp. 3745-3752, 2018.

Dario Ivan Rosas Miranda *received the communications and electronics engineer degree in 2013 and Master of Science in microelectronics degree in 2015. Currently, he is doctor student*

researcher at Instituto Politécnico Nacional. His research interest includes computer vision, image processing.

Volodymyr Ponomaryov *received the Ph.D. degree in 1974 and D.Sci. in 1981. His research interests include signal/image/video processing, pattern recognition, and real-time filtering. He has also been a promoter of 42 Ph.D.s. He has published of about 500 international scientific and conference papers, and 23 patents of ex USSR, Russia and Mexico, and five scientific books.*

Rogelio Reyes-Reyes *received the B.S. degree in Communications and Electronics Engineering from the Mechanical–Electrical Engineering School of National Polytechnic Institute (IPN) of Mexico, in 1999, the M.Sc. degree in Microelectronics and the Ph.D. degree in Communication and Electronic Engineering from the Graduate Section of the IPN, in 2003 and 2009, respectively. In 2002, he joined the Computer Department of the Culhuacan Campus of the IPN where he is now a professor. His main research fields are video and image processing, embedded security systems and related fields.*