COMPUTER SCIENCE

Depth Map Reconstruction Based on Features Formed by Descriptor of Stereo Color Pairs

V. F. Kravchenko[a], **V. I. Ponomaryov**[b,*],
RAS Academician V. I. Pustovoit[c], **and D. Rosas-Miranda**[b]

Received March 25, 2019

Abstract—An approach to depth map reconstruction from stereo pairs of color images in visualization of three-dimensional objects is proposed and substantiated for the first time. The approach makes use of a novel local image descriptor based on visual primitives and relations between them, namely, cocolority, coplanarity, distance, and angle. The new approach is compared with other well-known descriptors, such as DAISY and SID. Numerical experiments and an analysis and physical interpretation of their results obtained in the case of actual radiometric differences in the exposition or illumination of stereo image pairs have shown that the new approach is superior to other existing descriptors.

Stereo visualization is an actively developing area in three-dimensional (3D) computer visualization, where the main problem is to estimate the positions of objects in three-dimensional space on the basis of constructed depth maps (DM).

Computer visualization problems include segmentation and detection of objects in 3D by extracting information from multidimensional data in actual systems and transforming these data on the basis of efficient descriptors, which substantially reduce the dimension of the problem. Image contours and relations between them are the basic elements in computer visualization and robotics and represent generalized information on analyzed scenes in the form of a series of features important for recognition. Among descriptors of this type, Daisy and SID are the ones most efficient and widely used in practice.

Based on the ideas described in [1−12], an original method for reconstructing depth maps from stereo color pairs in 3D visualization is proposed and justified. The new image descriptor (VPR) relies on visual primitives and relations between them and characterizes distinctions in color, plane positions, distances, and angles between the primitives. A fundamental difference of VPR from other well-known descriptors is

[a] *Kotelnikov Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 125009 Russia*
[b] *Instituto Politecnico Nacional, Mexico, 04430 Mexico*
[c] *Scientific and Technological Center of Unique Instrumentation, Russian Academy of Sciences, Moscow, 117342 Russia*
*e-mail: vponomar@ipn.mx, vponomar@mail.ru

that it uses structural and semantic information. This improves its robust properties, specifically, in the case of nonideal recoding, illumination, and reflection in stereo color pairs. Numerical experiments and an analysis and physical interpretation of their results under actual radiometric differences in the exposition and illumination of stereo image pairs have shown that the new approach is superior to other available descriptors.

FORMULATION OF THE PROBLEM AND SOLUTION METHOD

Our theoretical description of the proposed image descriptor is based on the Riesz transform [9]. For a given two-dimensional signal $f(x, y)$, two odd signal components are formed:

$$F_{01,02}(w) = \begin{cases} i\dfrac{w_{x,y}}{\|w\|} F(w), & w \neq 0, \\ 0, & w = 0. \end{cases}$$

The Log-Gabor filter $G_e(w) = \exp\left(-\dfrac{\log\left(\dfrac{\|w\|}{w_0}\right)}{2(\log \sigma_0)^2}\right)$ is used, for which two odd components $G_{01}(w)$ and $G_{02}(w)$ are found, and the isotropic filter $G(w) = \sqrt{G_{01}^2(w) + G_{02}^2(w)}$ is applied.

After filtering, we form a vector of three components (two odd and one even) (so-called monogenic signal), namely, $f_m(x) = [f_e(x), f_{01}(x), f_{02}(x)]$. In spherical coor-
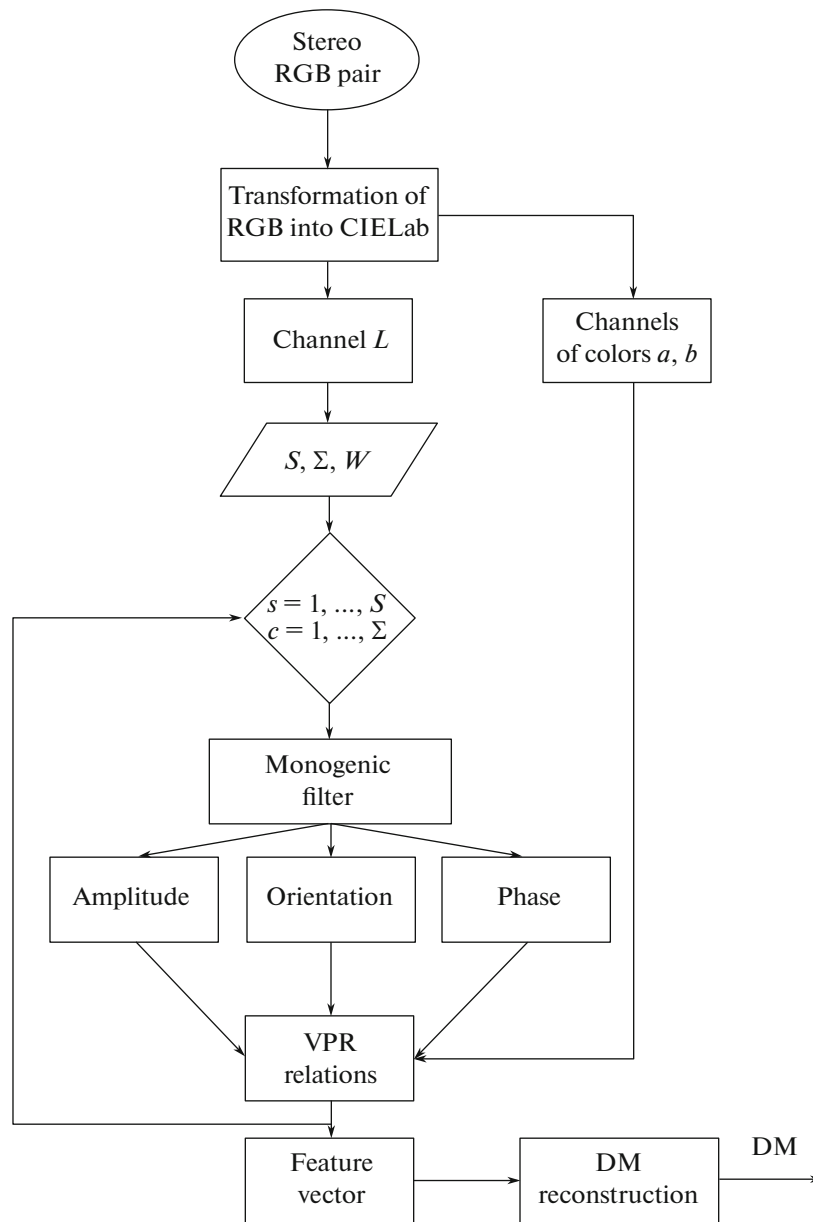
Fig. 1. Block diagram of the proposed method.

dinates, this vector is characterized by the length $A(x)$ and two angular coordinates:

$$\phi(x) = \arctan\left(\frac{f_0(x)}{f_e(x)}\right) \quad \text{and}$$

$$\theta(x) = \arctan\left(\frac{f_{02}(x)}{f_{01}(x)}\right).$$

The 2D visual primitive $\Pi_{2D}(x)$ is determined by these three parameters $A(x)$, $\phi(x)$, and $\theta(x)$, while the 3D visual primitive is characterized by the vector $\Pi_{3D}(x) = (A(x), \theta(x), \phi(x), d(x))$, where d determines the position of the primitive in three-dimensional space.

The dependences between two primitives are determined by a number of features. Specifically, cocolority is computed in the color space CIELab as follows:

$$R^C(\Pi_i, \Pi_j) = \sqrt{(L_i - L_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2}.$$

Other features are characterized by color, position, orientation, and relations between them, namely, by the angle between the primitives,

$$R^A(\Pi_i, \Pi_j) = \arccos\left(\frac{\Pi_i \Pi_j}{|\Pi_i||\Pi_j|}\right);$$
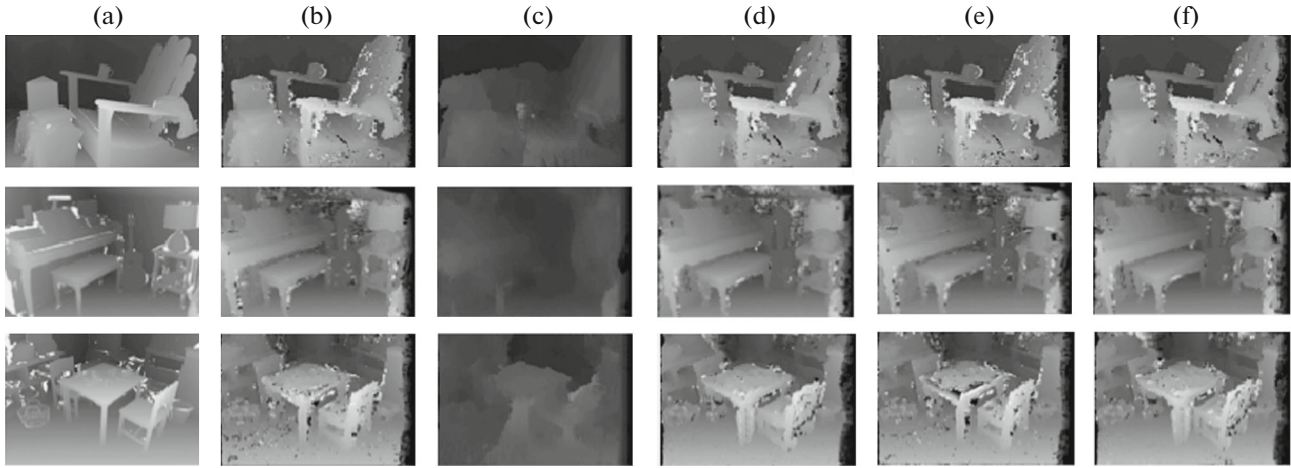
by the normalized distance between them,

Fig. 2. Results of depth map reconstruction: (a) true (GT) depth map and the maps reconstructed using the following descriptors: (b) Daisy, (c) SID, (d) VPR, (e) Daisy E (nonideal illumination), and (f) VPR E (nonideal illumination) in the cases of Adirondack, Piano, and Playtable images (from top to bottom).

$$R^{ND}(\Pi_i, \Pi_j) = \frac{1}{2}\left(\frac{|W_i \times \Pi_j|}{|\Pi_j|} + \frac{|W_j \times \Pi_i|}{|\Pi_i|}\right);$$

and by the coplanarity

$$R^P(\Pi_i, \Pi_j)$$
$$= \frac{1}{2}\left[\pi - \arccos\left(\frac{nA^O(\Pi_i)}{|n||A^O(\Pi_i)|}\right) - \arccos\left(\frac{nA^O(\Pi_j)}{|n||A^O(\Pi_j)|}\right)\right]$$

(see [13]). The novel descriptor VPR is based on relations between the primitives and is implemented as described in the block diagram of the method presented in Fig. 1, i.e., for a given color image I_{RGB}, its colors are transformed from RGB to the CIELab space. Cocolority is determined by the monogenic signal formed for S scales and Σ kernels ($i = 1, 2, ..., S$; $j = 1, 2, ..., \Sigma$) with the use of $S \times \Sigma$ Gaussian filters. Each filter performs the convolution of the image I_L of the channel L in the CIELab space and forms $3 \times S \times \Sigma$ components $f_m(x) = [f_e(x), f_{01}(x), f_{02}(x)]$ of the monogenic signal for different $S \times \Sigma$ filters with parameters s_i and σ_j in the Gaussian filters.

For each pixel, the new descriptor consists of a vector determined by the relations between the primitives (x, y) and (u, v) in a square window W centered at (x, y): $H_A(x) = [R_A\{s_i, \sigma_j\}(1, 1), ..., R_A\{s_i, \sigma_j\}(u, v)]$, where R_A is the angle determined by the relations between (x, y) and (u, v). The feature vector for angles between the primitives for $S \times \Sigma$ Gaussian filters is given by the matrix expression

$$D_A(x, y) = \begin{bmatrix} H_A(x,y)_{s_1}^{\sigma_1} & H_A(x,y)_{s_S}^{\sigma_1} \\ .... & .... \\ H_A(x,y)_{s_1}^{\sigma_\Sigma} ...... & H_A(x,y)_{s_S}^{\sigma_\Sigma} \end{bmatrix}. \quad (1)$$

Expressions for the features characterized by the normalized distances $D_{ND}(x, y)$ and the coplanarity parameters $D_{CP}(x, y)$ are found in a similar manner:

$$D_{ND}(x, y) = \begin{bmatrix} H_{ND}(x,y)_{s_1}^{\sigma_1} & H_{ND}(x,y)_{s_S}^{\sigma_1} \\ .... & .... \\ H_{ND}(x,y)_{s_1}^{\sigma_\Sigma} ...... & H_{ND}(x,y)_{s_S}^{\sigma_\Sigma} \end{bmatrix},$$

$$D_{CP}(x, y) = \begin{bmatrix} H_{CP}(x,y)_{s_1}^{\sigma_1} & H_{CP}(x,y)_{s_S}^{\sigma_1} \\ .... & .... \\ H_{CP}(x,y)_{s_1}^{\sigma_\Sigma} ...... & H_{CP}(x,y)_{s_S}^{\sigma_\Sigma} \end{bmatrix}. \quad (2)$$

The final descriptor in (x, y) is determined by the above-indicated features: $D_{VPR}(x, y) = [D_A(x, y), D_C(x, y), D_{CP}(x, y), D_{ND}(x, y)]$. Next, depth maps DM are reconstructed using a block matching procedure based on comparing the features of two stereo pairs.

NUMERICAL RESULTS

We studied synthetic benchmark images taken from the Middlebury Stereo Vision website [14, 15]. The 2005 dataset contains a series of different stereo pairs, together with ground-truth (GT) images (true depth maps) in full format (1390 × 1110 pixels), as well as in 1/2 and 1/3 formats. Additionally, the robustness of the new descriptor was examined using the 2014 dataset, which contains 33 stereo pairs divided into three groups (10 for learning, 10 for testing, and 13 additional pairs that are not illustrated by GT). Moreover, for each stereo pair, this dataset provides two images obtained in different illumination (L) or exposition (E) conditions.

All data in the experiments were processed together in order to confirm the efficiency and robustness of

Table 1. QBP values for VPR and other descriptors as applied to depth map reconstruction in the case of nonideal exposition (E) and illumination (L) of stereo pairs

Stereo pair/Descriptor	Daisy	VPR	Daisy E	VPR E	Daisy L	VPR L
Adirondack	25.76	**24.35**	25.32	**24.69**	40.27	**34.69**
Jadeplant	47.58	**45.36**	24.81	**23.20**	28.26	**20.59**
Piano	**22.57**	23.37	**22.68**	24.02	33.05	**28.76**
Pipes	27.10	**25.26**	26.48	**25.45**	41.96	**39.28**
Playroom	24.31	**22.00**	23.78	**21.96**	32.75	**29.74**
Colors	**43.24**	43.95	**43.42**	44.23	**50.87**	54.16
Classroom	**18.67**	20.83	**22.22**	24.86	**44.23**	46.98
Sticks	24.64	**23.38**	24.20	**23.49**	**43.49**	46.74
Backpack	10.89	**10.28**	10.92	**10.55**	**19.44**	19.62
Motorcycle	17.06	**16.32**	**16.16**	16.34	25.44	**21.41**
Pairwise averaged	26.18	<u>**25.51**</u>	24.00	<u>**23.87**</u>	39.57	<u>**37.15**</u>

the VPR descriptor as applied to the reconstruction of depth maps for stereo pairs obtained in different conditions.

The quality of depth map reconstruction was analyzed using the QBP criterion (the number of bad matching pixels). For each of the studied images DM, the QBP value was computed using the formula

$$QBP = \frac{1}{N} \sum_{(x,y)} (|DM_I(x,y) - DM_{GT}(x,y)| > \delta_d), \quad (3)$$

where N is the number of pixels in the image or frame; DM_I and DM_{GT} are the estimated and true (GT) maps, respectively; and $\delta_d = 1$.

In Fig. 2, the depth maps reconstructed by the VPR descriptor are visually compared with those produced by the DAISY and SID descriptors, which are best available in the literature; here, the traditional block matching technique is used to compute DM in all three methods.

Under different illumination (L) and exposition (E) conditions, the presented images show that the SID descriptor strongly smooths the objects and fails to estimate the depth maps correctly. The depth maps reconstructed by VPR and Daisy demonstrate a similar quality both in terms of a quantitative metric and in a subjective analysis. An analysis of the various images (Fig. 2, Table 1) suggests that the VPR descriptor has the best robustness in processing the inaccuracies of stereo image pairs.

Since the traditional block matching technique is used in the depth map reconstruction, the problem of occlusions cannot be resolved directly. The new VPR descriptor can be used in conjunction with other algorithms, such as semiglobal matching or graph cuts.

CONCLUSIONS

An analysis of the numerical results produced by the new method for depth map reconstruction suggests the following important conclusions:

(i) The proposed and substantiated method is based on visual primitives and relations between them, namely, cocolority, coplanarity, distance, and angle between the primitives. The method makes it possible to improve the quality of reconstructed depth maps.

(ii) The performance characteristics of the new VPR descriptor are superior to other descriptors, namely, DAISY and SID, which are widely used in the literature.

(iii) It has been confirmed experimentally that the VPR descriptor demonstrates the best robustness in depth map formation in the case of radiometric differences in the exposition and illumination of stereo image pairs.

REFERENCES

1. E. Ramos-Diaz, V. Kravchenko, and V. Ponomaryov, EURASIP J. Adv. Signal Process. **41** (1), 1−10 (2011).
2. V. F. Kravchenko, V. I. Ponomaryov, and V. I. Pustovoit, Dokl. Phys. **59** (11), 507−511 (2014).
3. V. F. Kravchenko, V. I. Ponomaryov, and V. I. Pustovoit, Dokl. Phys. **60** (11), 495−499 (2015).
4. V. F. Kravchenko, V. I. Ponomaryov, V. I. Pustovoit, and S. N. Sadovnichiy, Dokl. Phys. **62** (8), 374−378 (2017).
5. V. Huitron and V. Ponomaryov, IEEE Lat. Am. Trans. **14** (6), 2968−2973 (2016).
6. V. Gonzalez-Huilton, V. Ponomaryov, E. Ramos-Diaz, and S. Sadovnychiy, Signal Image Video Process. **12** (2), 231−238 (2018).

7. D. Rosas, V. Ponomaryov, and R. Reyes, Int. J. Comput. **17** (3), 171−179 (2018).

8. V. F. Kravchenko, H. M. Perez-Meana, and V. I. Ponomaryov, *Adaptive Digital Processing of Multidimensional Signals with Applications* (Fizmatlit, Moscow, 2009).

9. M. Felsberg and G. Sommer, IEEE Trans. Signal Process. **49**, 3136−3144 (2001).

10. Y. Xuanzi, Machine Vision Appl. **26** (7−8), 975−990 (2015).

11. Y. Wan, Z. Miao, Z. Tang, L. Wan, and Z. Wang, IEICE Trans. Inf. Sys. **95** (7), 2021−2024 (2012).

12. E. Tola, V. Lepetit, and P. Fua, IEEE Trans. Pat. Anal. Mach. Intell. **32** (5), 815−830 (2010).

13. N. Pugeault, F. Wörgötter, and N. Krüger, Int. J. Human. Robot. **7**, 379−405 (2010).

14. http://vision.middlebury.edu/stereo/data (November 2013).

15. http://vision.middlebury.edu/stereo/ (2016).

Translated by I. Ruzanova