ISSN 1028-3358, Doklady Physics, 2011, Vol. 56, No. 9, pp. 479–482. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © V.F. Kravchenko, V.I. Ponomaryov, V.I. Pustovoit, 2011, published in Doklady Akademii Nauk, 2011, Vol. 440, No. 3, pp. 335–338.

TECHNICAL PHYSICS

Visualization of Three-Dimensional Video Sequence Images in Real-Time Mode on the Basis of Multilevel Wavelet Decomposition

V. F. Kravchenko[a], V. I. Ponomaryov[b], and Academician V. I. Pustovoit[c]

Received May 26, 2011

DOI: 10.1134/S1028335811090084

1. For the first time, on the basis of concepts described in [1–9], we propose and substantiate a method that has allowed us to visualize in real-time mode three-dimensional (3D) image video sequences. This novel approach is based on two-dimensional (2D) video sequences and employs the multilevel decomposition of wavelets, in particular, of those based on atomic functions. This makes it possible to realize their best approximation characteristics in the course of the 3D visualization. The procedure involving the multilevel decomposition of wavelets based on atomic functions (WAF) [7, 8] is compared to other well-known algorithms: L&K [3], SSD [4], GEEMSF [5], and classic wavelets (Daubechies, Symlets, Biorthogonal, and Coiflets).

The method developed employs reconstructions of images (depth maps) with subsequent 3D visualization. In this case, the anaglyph-formation technique using two adjacent images (frames) of a 2D video sequence is applied. The realization of the algorithms proposed for the 3D visualization in the real-time mode was performed on the basis of an EVM DM642 processor made by Texas Instruments for image registration and processing.

2. This novel method includes the following operations: the decomposition of 2D video sequences in image frames; the separation of a color image over RGB components; the calculation of depth maps (DM) on the basis of stereo pairs formed by adjacent frames with the use of multilevel WAF wavelet decomposition (M-WAF); the depth-map correction by means of the dynamic compression; the synthesis of the anaglyph by the interpolation using the method of the closest adjacent pixel (NNI); and, finally, the visualization of the 3D-image video sequence. The block diagram for the algorithm proposed is presented in Fig. 1.

The multilevel decomposition of the wavelet function is formed on the basis of the discrete wavelet transform (DWT) with respect to three components [6]; namely, horizontal, vertical, and diagonal components can be represented as

$$WT_s = W_s \angle \Theta_{W_s}, \quad W_s = \frac{\sqrt{|D_{h_s}|^2 + |D_{v_s}|^2 + |D_{d_s}|^2}}{|A_s|}, \quad (1)$$

where W_s is the normalized wavelet modulus at the decomposition level s; D_{h_s}, D_{v_s}, and D_{d_s} are the horizontal, vertical, and diagonal components describing

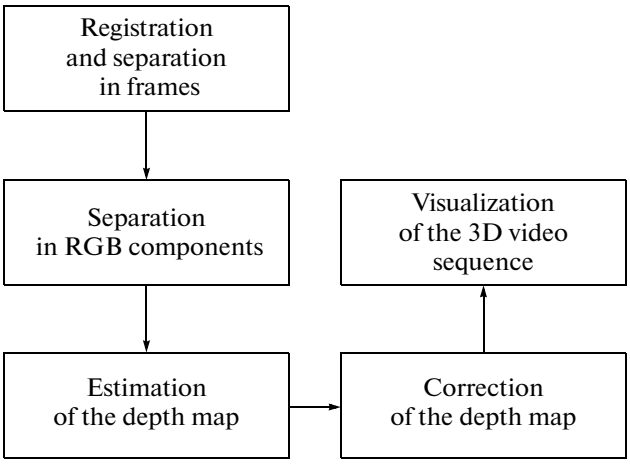

Fig. 1. Block-diagram for the algorithm proposed.

[a] *Kotel'nikov Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, Russia*
[b] *National Polytechnic Institute of Mexico, Mexico City, 04430 Mexico*
[c] *Scientific and Technological Center of Unique Instrumentation, Russian Academy of Sciences, Moscow, 117342 Russia*
e-mail: kvf@pochta.ru; vponomar@mail.ru; vponomar@ipn.mx; vlad_pst@yahoo.com

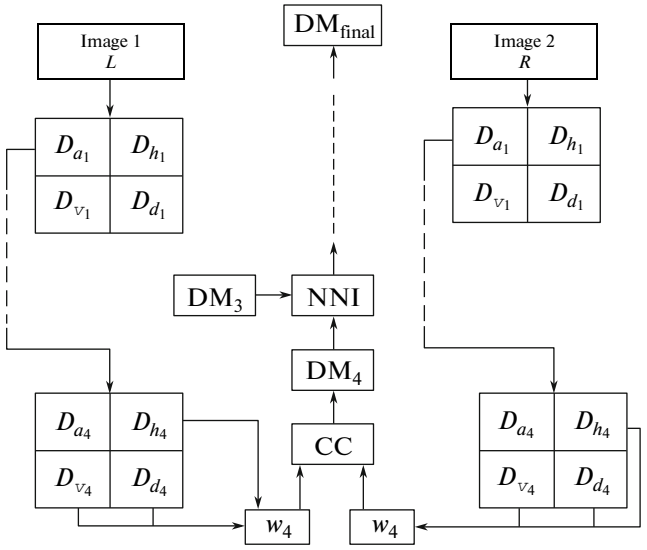

Fig. 2. Block diagram of the algorithm proposed to estimate the depth image.

image details at the same level s; A_s is the approximation component; and Θ_{w_s} is the phase calculated according to the expression

$$\Theta_{w_s} = \begin{cases} \alpha_s & \text{in the case of } D_{h_s} > 0, \\ \pi - \alpha_s & \text{in the opposite case of } D_{h_s} < 0, \end{cases} \qquad \alpha_s = \arctan\frac{D_{h_s}}{D_{v_s}}. \tag{2}$$

The estimate of the depth map (DM_s) at the decomposition level s for each (L_s and R_s) of the stereo pairs is based on the calculation of the normalized cross-correlation (CC_s) function in a sliding processing window (SO):

$$CC_s(x,y) = \sum_{i,j\in SO} \frac{W_{L_s}(i,j)W_{R_s}(x+i,y+j)}{\left(\sum_{i,j\in SO} W_{L_s}^2(i,j)\cdot \sum_{i,j\in SO} W_{R_s}^2(x+i,y+j)\right)^{1/2}}. \tag{3}$$

At each of the four decomposition levels, the algorithm calculates normalized wavelet values W_s for the current pair of adjacent frames in the sequence. Then, based on these values, we can calculate the cross correlations CC_s and the estimates for the depth map. The NNI method is applied to perform the interpolation between the current and more precise estimates for a depth map at each decomposition level in the course of the motion from a rough estimate to a more precise one. The final estimate for the depth map is formed with allowance for all its intermediate estimates as is seen in Fig. 2.

The classic methods for constructing anaglyphs, i.e., Photoshop, the least squares method, etc., are characterized by errors in the form of artifacts and phantasma effects, as well as by loss in the color quality. In order to improve the accuracy of depth-map estimates and minimize the errors indicated, we make use the method of dynamic compression in the anaglyph [9]. The method consists in the compression of the depth image in accordance with the formula $D_{new} = aD^P$, where D_{new} is the new depth map, with the quantities a and P satisfying the inequalities $0 < a < 1$ and $0 < P < 1$. At the completing stage of the anaglyph formation, the interpolation procedure is used in accordance with the NNI method.

Properties of the novel procedures and algorithms of 3D visualizations, which were known from the literature, were studied on the basis of standard numerical criteria: the quantity of "bad" pixels (QBP) [6] and the structural similarity index measure (SSIM) [10]. The visual subjective analysis of depth maps and generated anaglyphs was also used.

The values of the QBP criterion are calculated according to the following formula for each of the images under study:

$$\mathrm{QBD} = \frac{1}{N}\sum_{x,y} |d_E(x,y) - d_G(x,y)|^2, \tag{4}$$

where N is the number of pixels in either the image or the frame and d_E and d_G are, respectively, the estimated and true (ground truth) depth maps.

The value of the SSIM criterion is determined from the formula

$$\mathrm{SSIM}(x,y) = [l(x,y)]\cdot[c(x,y)]\cdot[s(x,y)], \tag{5}$$

where the functions of the luminescent $l(x,y)$, contrast $c(x,y)$, and structural $s(x,y)$ similarities are calculated in the following manner:

$$l(x,y) = \frac{2\mu_X(x,y)\mu_Y(x,y) + C_1}{\mu_X^2(x,y) + \mu_Y^2(x,y) + C_1},$$
$$c(x,y) = \frac{2\sigma_X(x,y)\sigma_Y(x,y) + C_2}{\sigma_X^2(x,y) + \sigma_Y^2(x,y) + C_2}, \tag{6}$$
$$s(x,y) = \frac{\sigma_{XY}(x,y) + C_3}{\sigma_X(x,y)\sigma_Y(x,y) + C_3}.$$

Here, X is the image to be estimated; Y is the ground-truth image; μ and σ determine the average value and the mean-square value for the images X or Y, respectively; and $C_1 = C_2 = C_3 = 1$.

The Aloe, Wood1, Lampshade1, and Venus (370 × 433 pixels in size; see http://vision.middlebury.edu/stereo/data) synthetic test images were studied. In addition, test color video sequences in the Avi format: Coastguard and Flowers (both of 300 frames,

Table 1. Values of the QBP and SSID criteria for synthetic stereo pairs of images

Image/Criterion	L&K	SSD	GEEMSF	WF Bio6.8	WF Coif1	WF Haar	WAF π_6	M-WF Coif1	M-WAF π_6
Aloe/SSIM	0.3983	0.6166	0.3017	***0.9267***	0.5826	0.5776	***0.9232***	0.5826	***0.9232***
Aloe/QBP	0.1121	0.4722	0.9190	0.0297	0.4517	0.4420	0.0130	0.4490	***0.0111***
Wood1/SSIM	0.1089	0.7142	0.7051	0.9367	0.7096	0.7072	0.9448	0.7096	***0.9448***
Wood1/QBP	0.1316	0.2376	0.2100	0.1258	0.2400	0.2402	0.1180	0.2400	***0.0919***
Lampshade1/SSIM	0.0861	0.6320	0.3124	0.7061	0.7061	0.7081	0.6897	0.7061	***0.7619***
Lampshade1/QBP	0.2430	0.2800	0.3410	0.2072	0.2071	0.2071	0.2017	0.2071	***0.1426***
Venus/SSIM	0.1990	0.4320	0.2145	0.5979	0.4530	0.4472	0.4604	0.4530	***0.6947***
Venus/QBP	0.3084	0.1428	0.2013	0.1694	0.5014	0.5010	0.1930	0.5011	***0.1091***

480 × 720 pixels in size, see http://trace.eas.asu.edu/yuv/index.html) and the photographed Video Test real-video sequence (200 frames, 480 × 720 pixels in size) were used to generate three-dimensional video sequences.

In the course of the depth-image formation, the cross correlation (*CC*) was calculated in a sliding window (SO) of 5 × 5 pixels, with the dynamic compression of all the algorithms under study with the parameters $a = P = 0.5$. The L&K method and the SSD algorithm were realized in accordance with [3] and [4], respectively. The GEEMSF procedure repeated that proposed in the original paper [5]. In the new algorithm based on the wavelet transformation, we employed both the classical types of wavelet functions (Coiflets, Daubechies, Biorthogonal, and Symlets) and wavelets based on atomic functions (up, fup, and π).

As is seen from Table 1, the criterion optimum values for the QBP and SSIM are attained when depth maps for images to be analyzed are estimated with the help of the algorithm based on the M-WAF π_6. The next algorithm (with respect to its efficiency) is based on the WAF π_6. Figure 3 demonstrates the best reconstruction of the depth map in the case of the algorithm using the WAF, which is compared to other well-known methods. The final stage of the visualization was performed on the basis of the anaglyph-formation method according to two adjacent images of the 2D video sequence [9].

A number of promising algorithms were realized with the help of the EVM DM642 digital processing system (Texas Instruments) [11], which made it possible to generate 3D video sequences virtually in real-time mode. In this case, we used the sequential connection of three DSP EVM DM642 systems when the first and the second processors realized the registration of current frames for both the 2D sequence and the formation of depth maps, whereas the third DSP allowed us to generate the anaglyph and its visualization on a display. Such a configuration of the digital processors significantly improved the processing speed so that it attained 20 (240 × 360 pixels, M-WAF) and 30 color frames (240 × 360 pixels, WAF) per second, which corresponds to the film speed (Table 2).

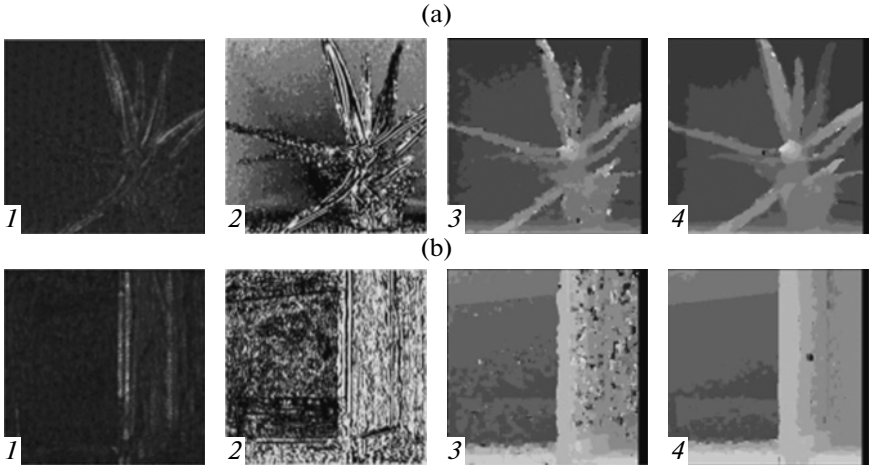

Fig. 3. Depth maps obtained for synthetic images (a) Aloe and (b) Wood1. Both are based on the (*1*) L&K, (*2*) SSD, (*3*) WF BIO 6.8, and (*4*) M-WAF π_6 algorithms.

Table 2. Processing time for various algorithms in processing video sequences

Algorithm	Matlab: time per frame, s (480 × 720 pixels)	DSP	
		3 sequential DSP: time per frame, s (240 × 360 pixels)	3 sequential DSP: time per frame, s (480 × 720 pixels)
Classic wavelets Coif1, Db6.8, Haar	6.16	0.031	0.071
WAF (up, fup_4, π_6)	6.19	0.031	0.071
M-Classic wavelets Coif1, Db6.8, Haar	6.76	0.048	0.08
M-WAF (up, fup_4, π_6)*	6.77	0.049	0.08

The conventional L&K and SSD algorithms that show a worse quality of reconstructing three dimensional video sequences require 22.59 and 16.26 s, respectively, for the 3D-frame visualization (with the use of MatLab). This testifies to the fact that these algorithms are not promising for image processing in the real-time mode.

The measured processing times included the time period from the registration of the 2D video sequence to the visualization of the generated anaglyph on a monitor. Thus, using lenses equipped with red and blue optic filters for the right and left eyes, respectively, we are able to observe 3D video sequences.

REFERENCES

1. I. Ideses and L. Yaroslavsky, Lect. Notes. Comp. Sci. **3212**, 273 (2004).
2. A. Smolic, P. Kauff, S. Knorr, et al., Proc. IEEE **11** (4), 607 (2011).
3. D. J. Fleet, *Measurement of Image Velocity,* Massachusetts: Kluwer Acad. Publ. (1992).
4. S. S. Beauchemin and J. L. Barron, ACM Comput. Surv. **27** (3), 433 (1995).
5. B. B. Alagoz, OncuBilim Algorithm and Syst. Labs. **8** (4), 1 (2008).
6. A. Bhatti and S. Nahavandi, Stereo Vision, Chapt. 2, pp. 27–48 (2008).
7. Yu. V. Gulyaev, V. F. Kravchenko, and V. I. Pustovoit, Dokl. Math. **75** (2), 325 (2007).
8. V. Kravchenko, H. Perez-Meana, and V. Ponomaryov, *Adaptive Digital Processing of Multidimensional Signals and Its Applications* (Fizmatlit, Moscow, 2009).
9. I. Ideses and L. Yaroslavsky, J. Opt. A: Pure Appl. Opt. **7**, 755 (2005).
10. W. S. Malpica and A. C. Bovik, Proc. IEEE Int. Conf. on Acoustics, Speech and Sign. Proc, 1149 (2009).
11. Texas Instruments. MS320DM642 Evaluation module with TVP Video Encoders. Technical Reference. 507345-0001. Texas Instruments. 2004.

Translated by G. Merzon