Fast Image Diffusion for Feature Detection and Description

Lu Feng, Zhuangzhi Wu, and Xiang Long

***Abstract*—In this paper, we introduce a new multiscale 2D feature detection and description method based on optimal O(1) bilateral filter feature (OBFF). Existing methods detect and describe features by analyzing the scale space generated by linear and nonlinear diffusion kernel function, like Gaussian scale space and anisotropic diffusion scale space. By using the anisotropic diffusion scale space, KAZE features achieve significant progress on the 2D feature detection by using the anisotropic diffusion scale space. It makes the blurring locally adaptive and retains better feature localization accuracy and distinctiveness than the SIFT method. Our method OBFF also generates the nonlinear scale space of image to detect the local feature. The optimal bilateral filter is advantage in object boundary preserving and antinoise ability and dramatically speed up feature detection in nonlinear scale space. We use the benchmark datasets to compare our method with state-of-the-art approaches.**

***Index Terms*—Bilateral filter, nonlinear scale space, feature detection, SIFT, binary descriptor.**

I. Introduction

In last decade, scale-invariant feature transform (SIFT) [1] and speed up robust feature (SURF) [2] have been widely applied to many feature matching applications. Both methods make use of the Gaussian scale space (GSS). The SIFT construct the GSS in a pyramidal framework and the SURF approximating Gaussian derivatives by box filters. However, lots of details like object boundaries are blurred in the GSS.

To overcome the drawback of GSS, some approaches have been proposed recently. These approaches choose the nonlinear diffusion scale space to extract local feature of image. The nonlinear diffusion scale space makes blur process adaptive to the local distribute of the image data. KAZE features were introduced by Alcantarilla *et al.* [3] and use Additive Operator Splitting (AOS) schemes [4] to approximate the Perona and Malik diffusion equation [5]. Repeatability and distinctiveness was increased compare with SIFT and SURF in this method due to the use of nonlinear diffusion scale space. Since AOS schemes require solving a large system of linear equations to obtain a solution, the computational complexity of KAZE is very higher than those approaches using GSS. To obtain low-computationally, Alcantarilla *et al.* [6] replace AOS schemes with mathematical framework named Fast Explicit Diffusion (FED) [7] to approximate the nonlinear diffusion scale space. FED schemes are easy to implement and more accuray than the AOS schemes. In this approach, FED schemes were embedded in a pyramidal framework to detect and describe the image feature. They call this approach as Accelerated-KAZE (A-KAZE). Wang *et al*. [8] use the bilateral filter to approximate the Perona and Malik diffusion equation and the relation between PM equation and bilateral filter is analyzed [9].

Due to the limited computational resource of camera enabled mobile devices, new methods have been proposed to reduce computational complexity while keeping up the performance of methods such as SIFT and SURF. ORB [10] and BRISK [11] speed-up feature detection and description by combining modifications of the FAST corner detector [12] and binary descriptors based on BRIEF [13] with scale and rotation invariance. ORB and BRISK feature are much faster to compute than SIFT and SURF, while showing comparable performance.

In order to obtain low-computational feature extraction, we take advantage of optimal O(1) bilateral filter to approximate the nonlinear scale space. To preserve low storage demand, we also use the Modified-Local Difference Binary (MLDB) descriptor [6] introduced by Alcantarilla *et al.* This descriptor overcomes the drawback of Local Difference Binary (LDB) descriptor and obtains rotation and scale invariant while exploiting gradient information from the nonlinear scale space in a very efficient way.

II. Related Work

A. Nonlinear Scale Space

To overcome the edge blurred in the GSS, the anisotropic diffusion filter has been introduced to preserve edges while smoothing details. The anisotropic diffusion filter can be written as follows:

$$\frac{\partial I}{\partial t} div(c(x,y,t)\cdot\nabla I), \tag{1}$$

where *div* is the divergence operator, ∇ is the gradient operator, I is the image luminance. $c(x,y,t)$ is the conductivity function which is the key to control the diffusion adaptive to the local image structure. Time parameter t is the scale parameter, and larger values lead to simpler image representations. In anisotropic diffusion the image gradient magnitude controls the diffusion. The conductivity function is defined as follow:

$$c(x,y,t) = g(|\nabla I_\sigma(x,y,t)|), \tag{2}$$

Manuscript received August 6, 2014; revised November 18, 2014.

The authors are with the Beihang University, Beijing, China (e-mail: fenglupeter@126.com, zzwu@buaa.edu.cn, long@buaa.edu.cn).

DOI: 10.7763/IJCTE.2016.V8.1020

where the function ∇I is the gradient of Gaussian smoothed version of the original image L. The equation (1) is also called Perona and Malik diffusion equation [5]. Perona and Malik described two different formations for the conductivity function g:

$$g_1 = exp(-\frac{|\nabla I_\sigma|^2}{k^2}), g_2 = \frac{1}{1+\frac{|\nabla I_\sigma|^2}{k^2}}, \quad (3)$$

the parameter k is the constrast factor that determines witch edges have to be kept and which ones have to be smoothed. Since there are no analytical solution for the equation (1), we need to use numerical methods to approximate the differential equations. The explicit schemes are the simplest solution. But the computational complexity make them impractical to use in the feature detection stage.

Alcantarilla *et al*. introduce Additive Operator Splitting (AOS) [4] and Fast Explicit Diffusion (FED) [6] schemes to accelerate the speed of nonlinear scale space generation. The AOS schemes are semi-implicit and efficient to be solved. However, AOS schemes need to solve large system of linear equation at each scale level. The FED schemes overcomes AOS's drawback and achieve more efficient result. Both schemes need to discretize the equation (1) as follow:

$$\frac{I^{i+1}-I^i}{\tau} = \sum_{l=1}^{m} A_l(I^i)I^{i+1}, \quad (4)$$

where A_l is a matrix that encodes the image conductivities for each dimension, τ is a constant time step in order to respect stability conditions. The AOS schemes generate the nonlinear scale space iteratively as follow:

$$I^{i+1} = (E-(t_{i+1}-t_i)\cdot A(I^i))^{-1}I^i, \quad (5)$$

where E is the identity matrix. The FED schemes are motivated form a decomposition of box filters in term of explicit schemes. The main idea of FED schemes is to perform M cycles of n explicit diffusion steps with varying step sizes τ_j as follow:

$$\tau_j = \frac{\tau_{max}}{2cos^2(\pi\frac{2j+1}{4n+2})}, \quad (6)$$

where τ_{max} is the maximal step size that holds the stability condition. The corresponding end time τ_n of one FED cycle is:

$$\theta_n = \sum_{j=0}^{n-1}\tau_j = \tau_{\max}\frac{n^2+n}{3}, \quad (7)$$

let $I^{i+1,0} = I^i$, a FED cycle with n variable step sizes τ_j is as follow:

$$I^{i+1,j+1} = (E+\tau_j A(I^i))I^{i+1,j}, j = 0,...,n-1, \quad (8)$$

during the whole FED cycle, the matrix $A(I^i)$ are kept constant.

B. Bilateral Filter and Anisotropic Diffusion

Bilateral filter uses both intensity and spatial distance to calculate the weight of the neighbor pixels. This filter is defined as:

$$I_p = \frac{1}{W_p}\sum_{q\in S} G_{\sigma_s}(|p-q|)G_{\sigma_r}(|I_p-I_q|)I_q, \quad (9)$$

where p and q are pixel location, L_p and L_p are intensity value of pixels, G_{σ_s} and G_{σ_r} are spatial and intensity Gaussian kernels with standard deviations σ_s and σ_r , W_p is the normalize factor. The total weights are combined with spatial weights. and intensity weights.

The nonlinear scale space based on numerical approximation methods for PM equation is time consuming and unstable. Buades, Coll and Morel [14] have established the link existing between bilateral filtering and PM equation. They have proven that for small neighborhoods, bilateral filtering using a box function as spatial weight, asymptotically behaves as the Perona-Malik model. In a discrete setting, Durand and Dorsey [15] have shown that the bilateral filter, if constrained to the four neighbors of each pixel, corresponds to a discrete version of the Perona-Malik equation. Subsequently, Barash [9] used adaptive smoothing as a link between anisotropic diffusion and bilateral filtering, each of which can be viewed as a generalization of the former.

The direct implementation of the standard bilateral filter requires $O(\sigma^2{}_s)$ operations per pixel, where σ_s is the radius of the effective support of the spatial kernel. The computational complexity is too intensive for time-critical applications. Consequently, a plenty of studies on its simplification and acceleration have been proposed. To speed up the BF filter, we use the sparse approximation with fixed number of box method proposed by Pan *et al*. [16]. Let L be the radius of the spatial support of the given spatial kernel K_s applied in the bilateral filter. Then all the candidate boxes together form a series B_l , where l is the radius of the box B_l and $l = 0,1,2,...,L$. For arbitrary K_s , it can be approximated using the weighted sum of all the candidate boxes, which is formulated as follows:

$$K_s(x-y) \approx \sum_{l=0}^{L} k_l B_l(x-y). \quad (10)$$

For the symmetry and monotonicity of spatial kernel, it is possible to find a real positive series k_l that minimizes the following squared error:

$$||K_s - \sum_{l=0}^{L} k_l B_l||_2^2, \quad (11)$$

where the spatial dependency is omitted for simplicity.

When L is large equation, the computational cost will be unbearable for time-critical application. We limit the number of boxes used in the approximation should not be larger than a predefined number N. For any $l \in [0,L]$, we align the center of the corresponding box with that of K_s and pad it with zeroes up to the same size as K_s . Then the columns of each

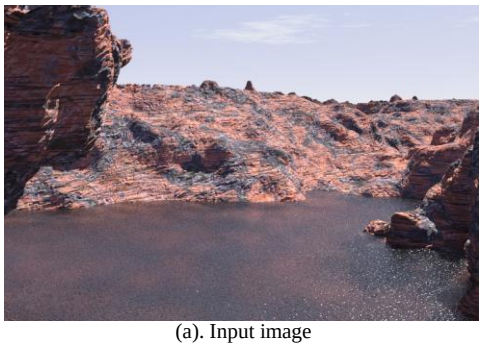
padded box are concatenated to form a column vector b_l. We then put these column vectors to form a matrix B of size $S \times (L+1)$, where $S = (2L+1)^2$ is the number of elements in K_s. We concatenate the columns of K_s to form a column vector q. k is defined as a column vector containing all the coefficients k_l, the optimization problem given in (11) can be reformulated as:

$$\hat{k} = arg\min_k \| q - Bk \|^2, s.t. \| k \|_0 \le N, \quad (12)$$

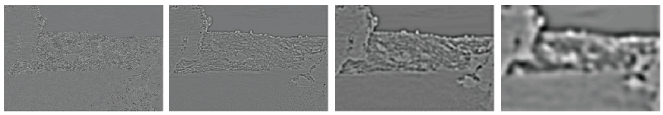
where $\| k \|_0$ is the L_0 norm of the vector k, which denotes the number of non-zero elements in k. We use Batch- Orthogonal Matching Pursuit (B-OMP) algorithm to solve equation (12). After we obtain the radiuses and coefficients of the boxes, the optimal O(1) bilateral filter can be given as equation (13):

$$u^{oBF} = \frac{\sum_y (\sum_{l|k_l \neq 0} k_l B_l(x-y)) K_r(I(x)-I(y)) I(y)}{\sum_y (\sum_{l|k_l \neq 0} k_l B_l(x-y)) K_r(I(x)-I(y))} \quad (13)$$

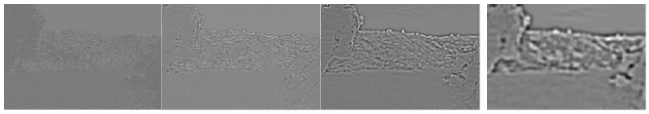
III. Optimal Bilateral Filter Feature



(a). Input image



(b). Difference of Linear scale space
$(\sigma_s = 0.8 - 32.0)$



(c). Difference of nonlinear scale space
$(\sigma_r = 0.6, \sigma_s = 0.8 - 32.0)$

Fig. 1. Comparison of DoSS and DoNSS.

A. Building Nonlinear Scale Space

Similar to SIFT method, the original image is repeatly convolved with OBF to produce each interval image for each octave. The spatial Gaussian kernel is increased per interval and intensity Gaussian kernel is fixed. The nonlinear scale space can be defined as follow:

$$L_i(x, y; k_i\sigma_s) = I_i(x, y) * OBF(x, y; k_i\sigma_s, \sigma_r) \quad (14)$$

where the $I_i(x, y)$ is the input image (original image is $I_0(x, y)$), $OBF(x, y; k_i\sigma_s, \sigma_r)$ is the OBF with two factor σ_s and σ_r, * is the convolution operation, $L_i(x, y; k_i\sigma_s)$ is the nonlinear scale space, k_i is the scale difference factor. Once a octave is built, the last image in the current octave is down sampled and used as the input image for the next octave.

From Fig. 1, we can find that the nonlinear scale space can preserve more edges than the linear scale space from the result of the difference of scale space.

B. Feature Detection

To increase the detection accuracy, we computer the determinant of the Hessian for each of the filtered images Li in the nonlinear scale space. The Hessian of filtered images are normalized by the scale factor, i.e. $k_{i,norm} = k_i / 2^{o^i}$ and

$$L_{i,Hessian} = k_{i,norm}^2 (L_{i,xx} L_{i,yy} - L_{i,xy} L_{i,xy}) \quad (15)$$

For computing the second order derivatives, we use concatenated Scharr filter with step size $k_{i,norm}$. Scharr filters, as shown in Fig. 2, have better rotation invariance than other filters or central differences differentiation. First, we search for maxima of the detector response in spatial location. For each step, the detector responses which higher than the threshold and maxima in a window of 3×3 pixels will be preserved. Then, for each of the potential response, we check that the response is a maxima with respect to other keypoints from level $i+1$ and $i-1$, respectively directly above and directly below in a window of size $\sigma_{i,s} \times \sigma_{i,s}$ pixels. Finally, the 2D position of the keypoint is estimated with sub-pixel accuracy by fitting a 2D quadratic function to the determinant of the Hessian response in a 3×3 pixels neighbourhood and finding its maximum.

-3	0	3
-10	0	10
-3	0	3

-3	-10	-3
0	0	0
3	10	3

Fig. 2. 3×3 Scharr filter template.

Due to the edge preserving property of bilateral filter, we can extract more feature points on the edge of the image region. Fig. 3 has shown that on the edge of the sky and the landscape many keypoints have been extracted. The feature points on the edge of the image region will benefit the 3D reconstruction of low-textured building.

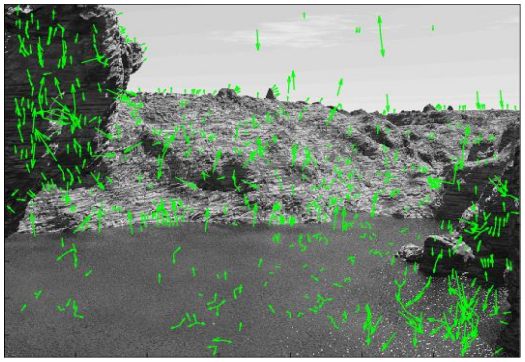
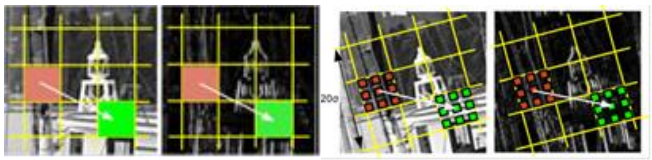


Fig. 3. Detected feature points by Scharr filter.

C. Feature Description

Binary descriptors which used in BRIEF, ORB and BRISK have widespread used since they make feature points be matched very efficiently. We use the Modified-Local Difference Binary (M-LDB) [6] descriptor which utilizes gradient and intensity information from the nonlinear scale space. Unlike LDB descriptor, M-LDB descriptor obtains rotation invariant by estimating the main orientation of the keypoint and the grid of LDB rotate accordingly. For the scale invariant, M-LDB subsamples the grids in steps by using the scale factor for the feature. M-LDB uses the derivatives computed in the feature detection step, reducing the number of operations required to construct the descriptor.

The LDB descriptor was introduced in [17] and follows the same principle as BRIEF, but using binary tests between the averages of areas instead of single pixels for additional robustness. In addition to the intensity values, the mean of the horizontal and vertical derivatives in the areas being compared is used, resulting in 3 bits per comparison. LDB proposes using various grids of finer steps, dividing the patch in 2×2, 3×3, 4×4, etc. grids, as shown in Fig. 4(a). The averages of those subdivisions are very fast to compute using integral images if the descriptor is upright (not rotation invariant).



(a). LDB binary test (b). M-LDB binary test

Fig. 4. LDB and M-LDB binary tests between grid divisions of a keypoint.

However, when considering the rotation of the keypoints integral images can not be used, and visiting all points in a rotated subdivision can be relatively expensive in computation time. Rotation invariance is obtained by estimating the main orientation of the keypoint as in KAZE, and the grid of LDB rotated accordingly. Instead of using the average of all pixels inside each subdivision of the grid, we subsample the grids in steps that are a function of the scale s of the feature. This approximation of the average performs well in our experiments. The scale-dependent sampling in turn makes the descriptor robust to changes in scale. This process is depicted in Fig. 4(b). M-LDB uses the derivatives computed in the feature detection step, reducing the number of operations required to construct the descriptor.

Given that M-LDB computes an approximation of the average of the same areas in the intensity and gradient images, the Boolean values that result from the comparisons are not independent of each other. Reducing the size of the descriptor by choosing a random subset of the bits [18] or with a more elaborated method such as that used in [10] is expected to improve the results, or at the very least reduce the computational load without decreasing performance.

IV. Experimental Results

In this section, we use the VLBenchmarks evaluation from [17] to evaluate the detector repeatability in the Oxford dataset and synthetic rotation and Gaussian noise experiments. For the latter case we use the Iguazu dataset introduced in [3]. The VLBenchmark reimplements the protocol introduced in [19] for local detectors evaluation. We compare the performance of OBFF with respect to BRISK, ORB, SURF, SIFT and A-KAZE. For BRISK, ORB, SURF and SIFT we used their OpenCV implementations, while for A-KAZE we use the original library provided by the authors. Table I shows average combined detection and description performance results considering the Matching Score (MS) and Recall (RC) as described in [20].

TABLE I: Evaluation Result of Descriptors

Featrues	Size	Bikes		Boat		Trees	
		MS	RC	MS	RC	MS	RC
SIFT	128Bytes	8	69	15	62	4	29
SURF	64Floats	27	68	11	52	7	23
ORB	256Bits	25	67	7	17	8	34
BRISK	512Bits	3	55	3	37	3	22
A-KAZE	486Bits	47	87	16	47	17	48
OBFF	486Bits	32	68	13	35	12	41

Fig. 5 shows a timing evaluation of the combined detection and description considering 1000 features extracted from the fist image of the Graffiti dataset. This image has a resolution of 800 ×640 pixels. All timing results were obtained with an Intel Core i7-3770 CPU and 8G RAM. From Table I and Fig. 1, we can see our method's score is lower than A-KAZE but the time consumption is less than it.

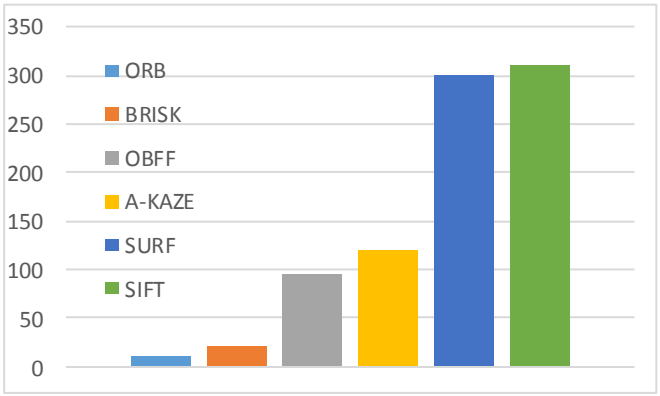


Fig. 5. Time evaluation of image feature descriptors.

V. Conclusion

In this paper, a novel method OBFF has been presented for image feature detection and description. By using the Optimal O(1) bilateral filter the speed of constructing nonlinear scale space and accuracy of feature detection have been significantly improved. We also use the M-LDB descriptor to speed up the feature points match process. The experiment results show that our method need optimization to achieve higher performance in accuracy and speed.

References

[1] D. Lowe, "Distinctive image features from scale in variant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[2] H. Bay, T. Tuytelaars, and L. Gool, "Surf: Speeded up robust features," *Computer Vision–ECCV*, pp. 404-417, Springer, 2006.

[3] P. Alcantarilla, A. Bartoli, and A. Davison, "Kaze features," *Computer Vision–ECCV*, pp. 214-227, Springer, 2012.
[4] J. Weickert, B. Romeny, and M. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 398-410, 1998.
[5] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629-639, 1990.
[6] P. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Proc. British Machine Vision Conf.*, 2013.
[7] S. Grewenig, J. Weickert, and A. Bruhn, "From box filtering to fast explicit diffusion," *Pattern Recognition*, pp. 533-542, Springer, 2010.
[8] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Sixth International Conference on Computer Vision*, 1998, pp. 839-846.
[9] D. Barash, "Fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 844-847, 2002.
[10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Proc. IEEE International Conference on Computer Vision*, 2011, pp. 2564-2571.
[11] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proc. IEEE International Conference on Computer Vision*, 2011, pp. 2548-2555.
[12] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *Computer Vision–ECCV*, Springer, pp. 430-443, 2006.
[13] M. Calonder, V. Lepetit, M. Ozuysal,T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281-1298, 2012.
[14] A. Buades, B. Coll, and J. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1499-1505, 2006.
[15] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics (TOG)*, vol. 21, pp. 257-266, 2002.
[16] X. Yang and K. Cheng, "LDB: An ultra-fast feature for scalable augmented reality," in *Proc. IEEE and ACM Intl. Sym. on Mixed and Augmented Reality (ISMAR)*, 2012.
[17] S.-D. Pan, X.-J. An, and H.-G. He, "Optimal O(1) bilateral filter with arbitrary spatial and range kernels using sparse approximation," *Mathematical Problems in Engineering*, vol. 2014, no. 1, pp. 1-11, 2014.
[18] J. Heinly, E. Dunn, and J. M. Frahm, "Comparative evaluation of binary features," in *Proc. Eur. Conf. on Computer Vision (ECCV)*, 2012, pp. 759-773.
[19] K. Lenc, V. Gulshan, and A. Vedaldi. (2012). VLBenchmarks. [Online]. Available: http://www.vlfeat.org/benchmarks/
[20] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. "A comparison of affine region detectors," *Intl. J. of Computer Vision*, vol. 65, no. 1-2, pp. 43-72, 2005.

Lu Feng graduated from Wuhan University of Technology in 2006 with a B.S. degree in software engineering. He is a PhD candidate at the Department of Computer Science and Technology, BeiHang University. His research interests include computer graphics, computational geometry, and image processing.

Zhuangzhi Wu is an associate professor at the Department of Computer Science and Technology, BeiHang University. He received his B.S. degree in mechanical engineering from Harbin Engineering University in 1991, an M.S. degree in industrial automation from BeiHang University in 1995, and a PhD degree in computer software and theory from BeiHang University in 2001. His research interests include computer graphics, computational geometry, digital geometry processing, 3-D anthropometry, and 3-D Vision Measurement.

Xiang Long received the bachelor degree in mathematics from Peking University, Beijing. He received the master and Ph.D. degrees in computer science and technology in Beihang University. Currently, he is a professor at Beihang University, Beijing. His research interests include high performance computing, computer architectures, reliability.