Optical Engineering

SPIEDigitalLibrary.org/oe

Efficient active depth sensing by laser speckle projection system

Xuanwu Yin
Guijin Wang
Chenbo Shi
Qingmin Liao

Efficient active depth sensing by laser speckle projection system

Xuanwu Yin, Guijin Wang,* Chenbo Shi, and Qingmin Liao
Tsinghua University, Department of Electronic Engineering, Rohm Building 6-107, Beijing, China

Abstract. An active depth sensing approach by laser speckle projection system is proposed. After capturing the speckle pattern with an infrared digital camera, we extract the pure speckle pattern using a direct-global separation method. Then the pure speckles are represented by Census binary features. By evaluating the matching cost and uniqueness between the real-time image and the reference image, robust correspondences are selected as support points. After that, we build a disparity grid and propose a generative graphical model to compute disparities. An iterative approach is designed to propagate the messages between blocks and update the model. Finally, a dense depth map can be obtained by subpixel interpolation and transformation. The experimental evaluations demonstrate the effectiveness and efficiency of our approach. [DOI: 10.1117/1.OE.53.1.013105]

Keywords: depth sensing; range finding; depth recovery; three-dimensional image acquisition.

Paper 131371 received Sep. 4, 2013; revised manuscript received Nov. 28, 2013; accepted for publication Dec. 4, 2013; published online Jan. 20, 2014.

1 Introduction

Depth estimation is a basic problem in computer vision, widely applied in fields as diverse as automatic driving, shape measurement in industry, biomedical imaging, and human-computer interaction. Refer to Ref. 1 for an overview.

In recent years, estimating depth by speckle projection has attracted considerable attention. Siebert and Marshall[2] introduced three structures of speckle systems to capture three-dimensional (3-D) human body. Garcia et al.[3] used ground glass to generate speckles and computed depth by correlating the captured image with reference images at different depth layers. Chen and Chen[4] projected speckles with a common projector and built up a binocular system. Schaffer et al.[5] used a laser and a diffuser to generate speckles and computed the depth using a temporal correlation method. However, these systems are restricted to near range and cannot be applied in bright environment due to the use of visual light. Recently, researchers tended to use infrared projectors and cameras[6,7] to expand the sensing range for indoor applications. In these systems, the depth was estimated by stereo matching[8] between the real-time captured image and the reference image. However, the high contrast of speckles could not be well handled by conventional methods. Wang et al.[9] utilized adaptive binarization to handle the high contrast and proposed a progressive approach to estimate the depth map. Nevertheless, none of the approaches mentioned has considered the impact of ambient lighting.

To tackle these problems, in this paper, we propose a novel approach to sense the indoor depth information. In our system, infrared laser speckle pattern is projected onto the scene and a single camera is used to capture images. First, we propose a novel direct-global separation approach to remove the ambient lighting and extract the pure speckle pattern. After that, a set of sparse support points are selected. Then a disparity grid is built upon these support points and a generative graphical model is introduced to search disparities. We propagate messages between blocks and update the disparities and the model iteratively. Finally, the depth map can be obtained by transforming the disparities into depth values. The experiments demonstrate the efficiency and effectiveness of our approach.

The rest of this paper is organized as follows. In Sec. 2, we present the system structure and the framework of the algorithm. The preprocessing, including speckle pattern extraction and feature computing, is introduced in Sec. 3. Section 4 presents how to build the graphical model and search for the disparities with this model. Experimental results are presented in Sec. 5. In the last section, we make a conclusion and a discussion of future work.

2 System Structure and Algorithm Framework

As shown in Fig. 1, the system is made up of one pattern projector and one infrared digital camera. We use the Kinect's projector, which consists of an infrared laser generator and a diffuser, to project the speckle pattern onto the scene. Aligned with the projector in Y and Z direction, the infrared camera captures the real-time speckle pattern reflected by the scene.

Figure 2 gives an example of the inputs of our scheme, a real-time image and a fixed reference image. The reference image is a flat plane perpendicular to the optical axis, captured at a certain distance. Thus, the setup is equivalent to a binocular system. The depth can be computed by estimating disparities between the real-time image and the reference image.

Figure 3 presents the framework of our proposed approach. First, we separate the direct component from the image to extract the pure speckle pattern. With this speckle pattern, we compute the matching costs of both left-to-right and right-to-left. Sparse support points are selected by evaluating the matching cost and left-right consistency. Then a disparity grid is built upon these support points. For each block in the disparity grid, we introduce a generative

*Address all correspondence to: Guijin Wang, E-mail: wangguijin@tsinghua.edu.cn

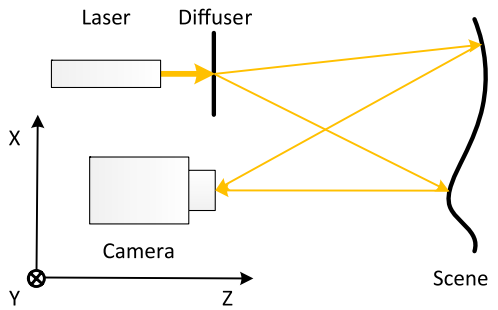

Fig. 1 System setup.

graphical model to represent the relationship between the support points and the unmatched points. Thus, disparities are computed by minimizing an energy function derived from the graphical model. The new reliably matched points are added as support points, and the graphical model is updated. Such process is repeated until convergence. Finally, the depth map is obtained by subpixel interpolation and transformation.

3 Speckle Pattern Extraction and Feature Computing

As the illumination of real-time captured image and the reference image can be significantly different, it is not reliable to simply utilize local intensities[10] to describe the speckles. To better describe the speckles, we first extract the pure speckle pattern, which is represented by binary features. Note that the fixed reference image is processed only once.

In the image, the intensity at (u, v), $i(u, v)$ can be viewed as a combination of two components, namely direct and global,[11]

$$i(u, v) = i_d(u, v) + i_g(u, v), \tag{1}$$

where $i_d(u, v)$ is the direct component due to direct illumination for the speckle pattern, and $i_g(u, v)$ is the global component due to ambient lighting. Nayar et al.[11] used high-frequency illumination to separate the direct and global components with multiple images.

In this work, we try to extract the direct component with a single image and obtain the pure speckle pattern. By analyzing tens of images captured under various illuminations, we find that the intensities of speckles are spatially nonuniform, while the smallest values in the neighborhood of a certain radius δ_s around the speckles are stationary. According to Ref. 12, the radius δ_s can be fixed as constant. Figure 4 presents a representative image patch. The red inset contains a single speckle of relatively low intensity and the green inset contains four visibly brighter speckles. It can be observed that the smallest value in the red inset is almost the same

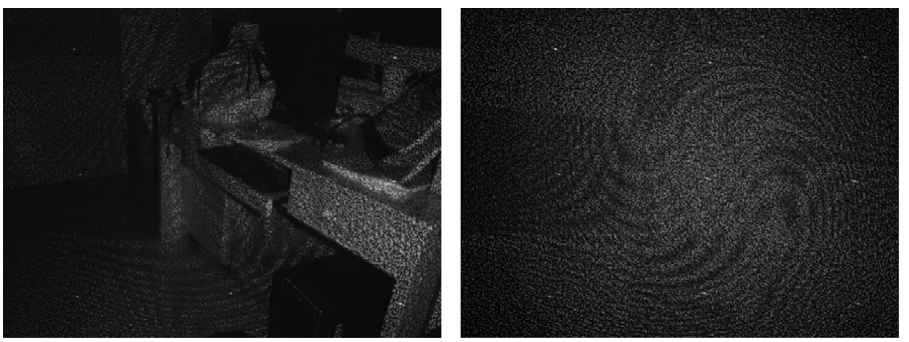

Fig. 2 Input image pair. Left: real-time image. Right: reference image.

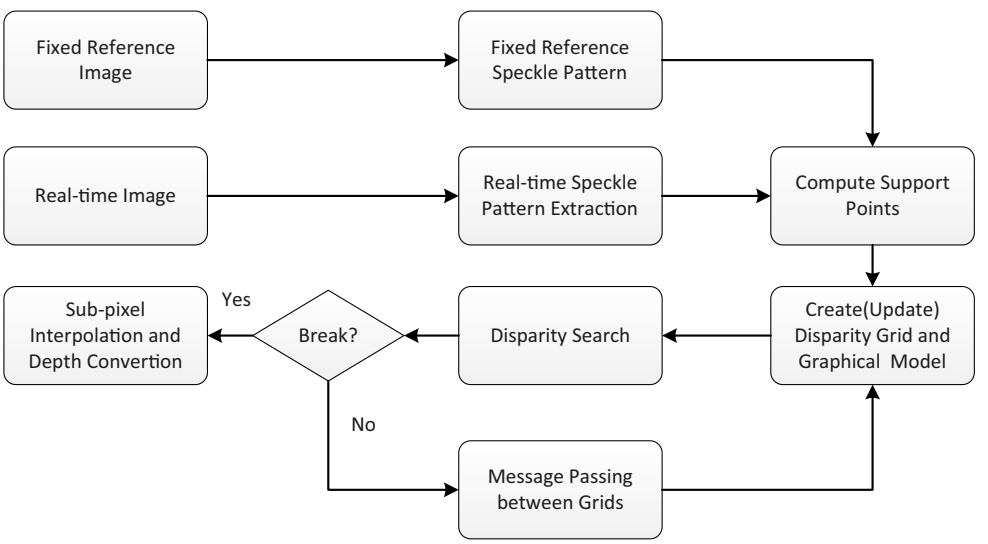

Fig. 3 Processing flow of the proposed approach.

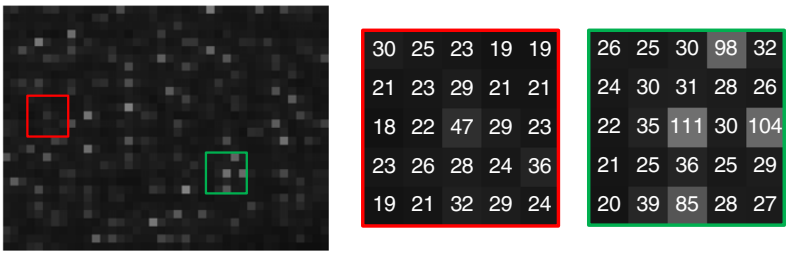

Fig. 4 A sample of real-time image and local details at different locations. (a) Sample patch in real-time image. (b) Detail for red inset in (a). (c) Detail for green inset in (a).

as that in the green inset. That is, in a window larger than $(2\delta_s)\times(2\delta_s)$, the lowest intensities reflect the global component.

Let x_n, $n=1,\ldots,N,\neq$ be the observed intensity values in a window of size $W_s\times W_s(W_s>2\delta_s)$, and X_i, $i=1,\ldots,N$ be its order statistics. The global component is estimated with the weighted average of these observations,

$$\hat{i}_g(u,v)=\frac{\sum_{k=1}^{N}w_kX_k}{\sum_{k=1}^{N}w_k}, \tag{2}$$

with the weight defined as

$$w_k=\frac{2}{1+\exp[\lambda(X_k-X_1)^2]}. \tag{3}$$

The parameter λ is a constant that controls the weight of each point, i.e., the contribution of a point to the estimated global component. Thus, the direct component at (u,v) can be estimated by

$$\hat{i}_d(u,v)=i(u,v)-\hat{i}_g(u,v). \tag{4}$$

The direct component, as the pure speckle pattern, is then utilized to compute the feature vector for each point.

To describe a certain point, we apply the CENSUS transform[13] in a $W_f\times W_f$ window around it to form a binary feature vector, denoted as $\mathbf{f}$. Let x_n^d, $n=1,\ldots,N$, be the estimated direct component with Eq. (4) in this window. The feature vector $\mathbf{f}$ can be treated as a binary string formed by a sequence of one-bit digits,

$$\mathbf{f}=\sum_{i\neq N/2}2^iU(x_i-x_{N/2}), \tag{5}$$

where $x_{N/2}$ is the center of the window, and $U(\bullet)$ is a step function,

$$U(x)=\begin{cases}1 & \text{if } x\geq 0\\ 0 & \text{otherwise}\end{cases}. \tag{6}$$

4 Depth Estimation with Iteratively Updated Graphical Model

Once the features have been computed, we can compute correspondences between the real-time image and the reference image to estimate the disparities. Geiger et al.[14] built a prior on the disparities by forming a triangulation on support points, which has been proved to be very effective. However, as there are out-of-range regions and non-Lambertian surfaces in the infrared images, the support points will not distribute as uniformly as that in their visual-light images. To cope with this problem, we built a disparity grid and introduced an iterative updating scheme inspired by Shi et al.[15] and Wang et al.[9]

4.1 *Support Points*

Support points are the points that can be robustly matched between the input image pair.[16,17] In this work, we select support points by evaluating the matching cost and uniqueness.

Let $\mathbf{o}=(u,v,\mathbf{f})^T$ be an observation in the real-time image, defined as the concatenation of its image coordinates, $(u,v)\in\mathbb{N}^2$, and a feature vector, $\mathbf{f}\in\mathbb{R}^Q$. Let $\mathbf{O}^r=\{\mathbf{o}_1^r,\ldots\mathbf{o}_N^r\}$ be the set of observations in the reference image that lie on the epipolar line associated with $\mathbf{o}$, with each observation formed as $\mathbf{o}_n^r=(u-d_n,v,\mathbf{f}_n^r)^T$, in which d_n is the disparity between $\mathbf{o}$ and $\mathbf{o}_n^r$. The matching cost between $\mathbf{o}$ and $\mathbf{o}_n^r$ is measured by the distance of the feature vectors,

$$\text{Cost}(\mathbf{o},\mathbf{o}_n^r)=D_f(\mathbf{f},\mathbf{f}_n^r), \tag{7}$$

where $D_f(\bullet,\bullet)$ is a distance function. In this paper, Hamming distance is used. To get rid of ambiguous matches, only the points whose absolute difference between the best and second best matches is large enough are retained. The uniqueness is measured by the left-right consistency, i.e., correspondences are retained only if they can be matched from left-to-right and from right-to-left.[9,14]

4.2 *Disparity Grid and Generative Graphical Model*

To merge the local information of the support points, we build a disparity grid on the real-time image by dividing the image into $M_g\times N_g$ blocks, with each block of size $W_g\times W_g$. The grid structure is illustrated in Fig. 5(a). For each block, the disparities of support points inside it and those in its four-neighborhood blocks are taken as the candidate disparities for it. These disparities form a candidate disparity set $\mathbf{D}=\{dc_1,\ldots,dc_M\}$, $dc_m\in\mathbb{N}$. This cross-block supporting scheme allows for the effective propagation of neighboring information between blocks.

We propose a generative graphical model to model the relationship between the candidate disparities and the observations, and the relationship between observations, as shown in Fig. 5(b). Denote by $\mathbf{o}$ the point we want to compute. These two relationships are described by two conditional independence assumptions:

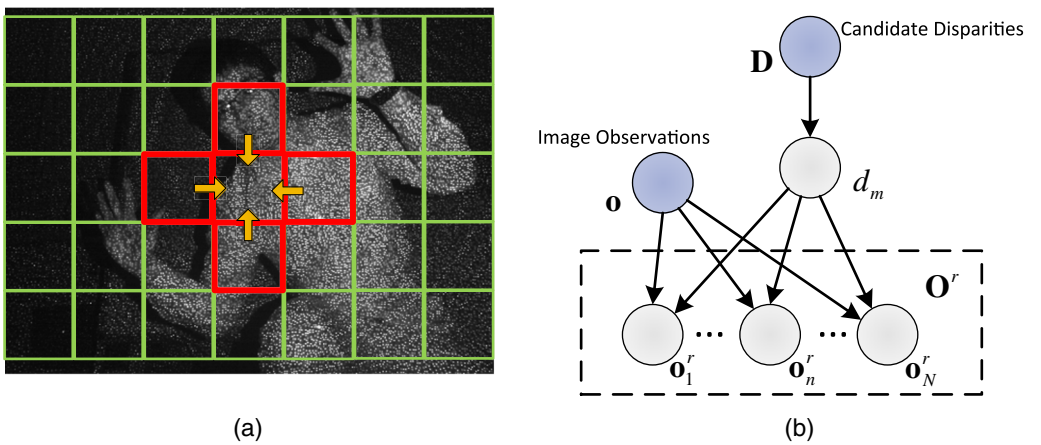

Fig. 5 Disparity grid and graphical model. (a) Grid setup and cross-block supporting. (b) Graphical model. Observations and candidate disparities are conditionally independent given d_m, the disparity between $\mathbf{o}$ and $\mathbf{o}_m^r$. And observations in $\mathbf{O}^r$ are conditionally independent given d_m and $\mathbf{o}$.

1. the observations $\{\mathbf{o}, \mathbf{O}^r\}$ and the candidate disparities $\mathbf{D}$ are conditionally independent given disparity d_m;
2. the observations in $\mathbf{O}^r$ are conditionally independent with each other given d_m and $\mathbf{o}$.

By the first assumption, the joint probability is factorized as

$$p(d_m, \mathbf{o}, \mathbf{o}_n^r, \mathbf{D}) = p(\mathbf{o}_n^r|\mathbf{o}, d_m, \mathbf{D})p(d_m|\mathbf{o}, \mathbf{D})p(\mathbf{o}, \mathbf{D}) \propto p(\mathbf{o}_n^r|\mathbf{o}, d_m)p(d_m|\mathbf{D})/Z, \tag{8}$$

with $p(\mathbf{o}_n^r|\mathbf{o}, d_m)$ being the likelihood, $p(d_m|\mathbf{D})$ being the prior, and Z being the normalization factor. The prior can be modeled as an equally weighted Gaussian mixture model

$$p(d_m|\mathbf{D}) \propto \sum_{dc \in \mathbf{D}} \exp\left[-\frac{(d_m - dc)^2}{2\sigma^2}\right]. \tag{9}$$

A constrained Laplace distribution is applied to represent the likelihood:

$$p(\mathbf{o}_n^r|\mathbf{o}, d_m) \propto \begin{cases} \exp[-\beta \cdot D_f(\mathbf{f}, \mathbf{f}_n^r)] & \text{if } d_m = d_n \\ \epsilon & \text{otherwise} \end{cases}, \tag{10}$$

where ϵ is a positive number, with a typical value $1.0e^{-6}$. The condition $d_m = d_n$ is derived from the fact that given a disparity, there is a deterministic mapping of $\mathbf{o}$ in the reference image.

Once the prior and likelihood are formulated, the disparity can be estimated by the maximum *a posteriori* estimation,

$$\hat{d} = \underset{d_m}{\operatorname{argmax}}\, p(d_m|\mathbf{o}, \mathbf{O}^r, \mathbf{D}), \tag{11}$$

where $\mathbf{O}^r$ is the set of observations in the reference image that lie on the epipolar line associated with $\mathbf{o}$. Similar to Eq. (8), the posterior is factorized as

$$p(d_m|\mathbf{o}, \mathbf{O}^r, \mathbf{D}) \propto p(d_m, \mathbf{o}, \mathbf{O}^r, \mathbf{D}) \propto p(\mathbf{O}^r|\mathbf{o}, d_m)p(d_m|\mathbf{D}). \tag{12}$$

By the second independence assumption, the conditional distribution over $\mathbf{O}^r$ can be further factorized as

$$p(\mathbf{O}^r|\mathbf{o}, d_m) = \prod_{n=1}^{N} p(\mathbf{o}_n^r|\mathbf{o}, d_m), \tag{13}$$

i.e., the product of all likelihoods. Note that the factorization in Eq. (13) holds and makes sense only if none of the likelihood takes on the zero value, which is the reason that ϵ is set to nonzero in Eq. (10). By plugging Eqs. (9), (10) and (13) into Eq. (12) and taking the negative logarithm, an energy function can be derived as follows:

$$E(d) = \beta \cdot D_f[\mathbf{f}, \mathbf{f}^r(d)] - \log \left\{\sum_{dc \in \mathbf{D}} \exp\left[-\frac{(d - dc)^2}{2\sigma^2}\right]\right\} + \log Z + C, \tag{14}$$

with $\mathbf{f}^r(d)$ being the feature vector at $(u - d, v)$ in the reference image, and C being a constant associated with ϵ only. Now, the disparity for each point can be estimated by minimizing the energy function

$$\hat{d} = \underset{d}{\operatorname{argmin}}\, E(d). \tag{15}$$

4.3 *Iterative Model Updating*

For out-of-range regions and non-Lambertian surfaces, the support points may not be sufficient. The disparity grid may not provide enough disparity information in these regions. To efficiently propagate the information between blocks, we propose an iterative scheme to update the support point set $\mathbf{D}$ and the graphical model.

The key of this iterative updating scheme is to select new reliably matched points during iterations. The energy in Eq. (14) itself is powerful to measure the reliability. Suppose for $\mathbf{o}$, the estimated disparity is $\hat{d}$. According to Eq. (14), the minimum matching energy is $E(\hat{d})$. The lower $E(\hat{d})$ is, the more reliably $\mathbf{o}$ is matched.

Algorithm 1 Iterative model updating and optimal disparity searching.

1: **repeat**
2: **for** each **o** that has not been reliably matched **do**
3: compute $E(\hat{d})$ and $\mathrm{Conf}(\hat{d})$ with the current model
4: **if** $E(\hat{d}) < E(d^*)$ and $\mathrm{Conf}(\hat{d}) > \mathrm{TH}_{\mathrm{Conf}}$ **then**
5: set $d^* = \hat{d}$
6: **if** $E(\hat{d}) < \mathrm{TH}_E$ **then**
7: add $\hat{d}$ into **D**
8: **end if**
9: **end if**
10: **end for**
11: recomputed the graphical model with the updated **D**
12: **until** after N iterations

We also use the confidence[10,18] to describe the goodness of matching, which is defined as

$$\mathrm{Conf}(\hat{d}) = \min_{\tilde{d}\neq\hat{d}}[E(\tilde{d}) - E(\hat{d})]. \tag{16}$$

This definition indicates the difference between the minimum energy and the second-minimum one.

Using both the matching energy and the confidence, we propose the iterative scheme presented in Algorithm 1 to update the model and search for d^*, the optimal disparity for **o**. At the beginning of each iteration, the matching energy and confidence for the points is computed with the current model. If the matching energy is lower than the current lowest matching energy $E(d^*)$, and the confidence is above some threshold $\mathrm{TH}_{\mathrm{Conf}}$, d^* will be updated to $\hat{d}$. Furthermore, if the matching energy is lower than some threshold TH_E, this point is treated as reliably matched and $\hat{d}$ will be added into **D**. At the end of each iteration, the model is recomputed with the updated support point set **D**.

We note that because the blocks can support neighboring blocks, the newly added support points can be used to update not only the block they belong to but also the neighboring blocks. Thus, the messages can propagate between blocks during iterations. Meanwhile, the thresholds, $\mathrm{TH}_{\mathrm{Conf}}$ and TH_E, can prevent the spurious messages from spreading out. Figure 6 gives an example of disparity messages propagating between blocks.

4.4 *Subpixel Interpolation and Depth Transformation*

Interpolation has been adopted to achieve subpixel accuracy.[19] We choose the linear interpolation method for its efficiency and effectiveness.

Suppose the integer disparity for a point is d, and the corresponding matching energy is $E(d)$. Let $\Delta L = |E(d) - E(d-1)|$ and $\Delta R = |E(d) - E(d+1)|$, the subpixel disparity d_{sub} is computed by

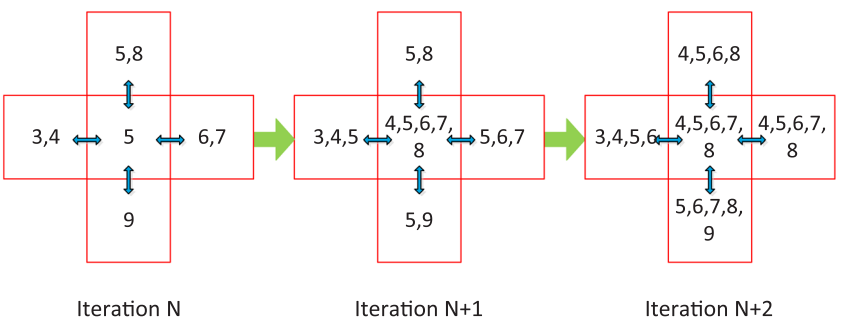

Fig. 6 Iterative message propagation. Numbers in the block are the estimated disparities of points in this block.

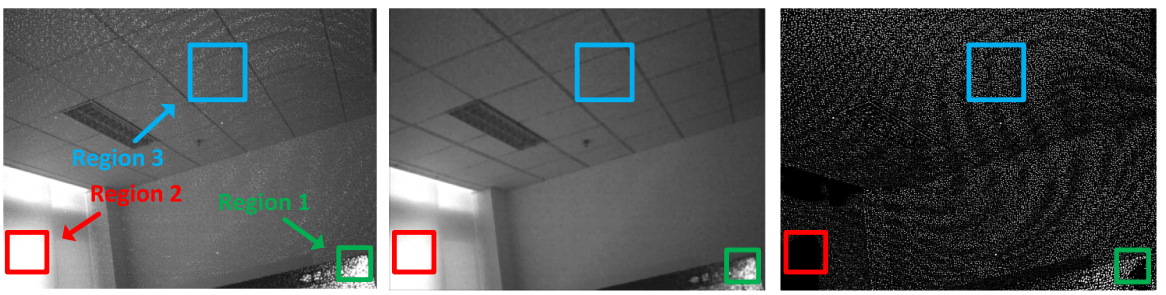

Fig. 7 Speckle pattern extraction result. (a) Original real-time image. (b) Global component (ambient lighting). (c) Direct component (speckle pattern). The green inset indicates a defocus blurred region, the red inset indicates a saturated region, and the blue inset indicates a normal region.

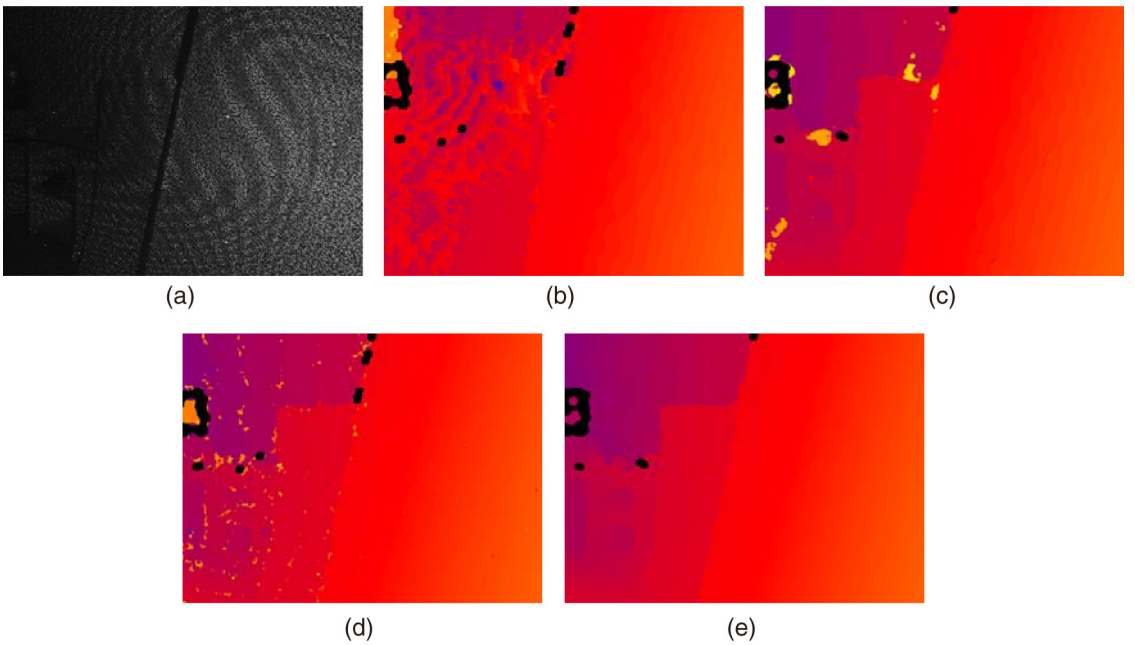

Fig. 8 Effectiveness of speckle pattern extraction and binary feature. (a) Real-time image. (b) Using intensity feature without pattern extraction. (c) Using intensity feature with pattern extraction. (d) Using binary feature without pattern extraction. (e) Using binary feature with pattern extraction.

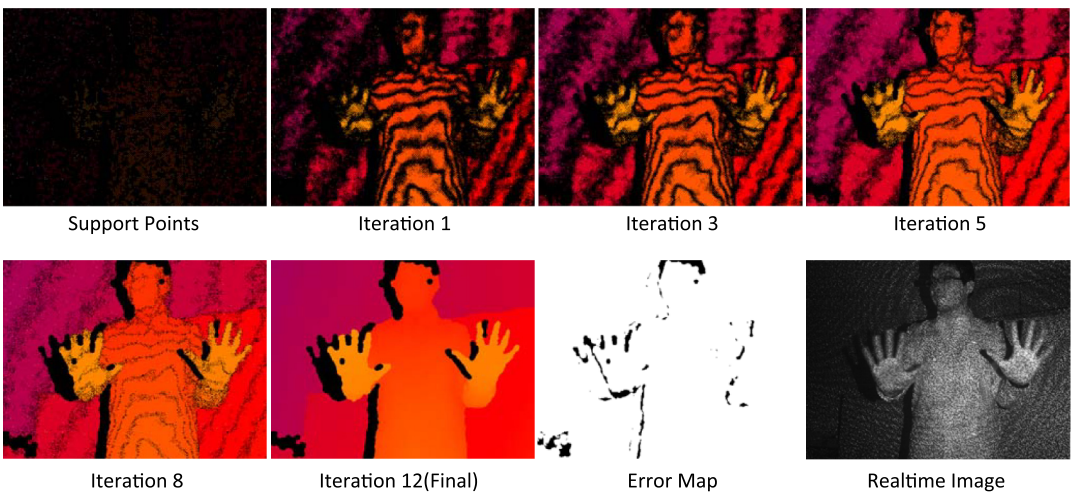

Fig. 9 Intermediate results during iterations.

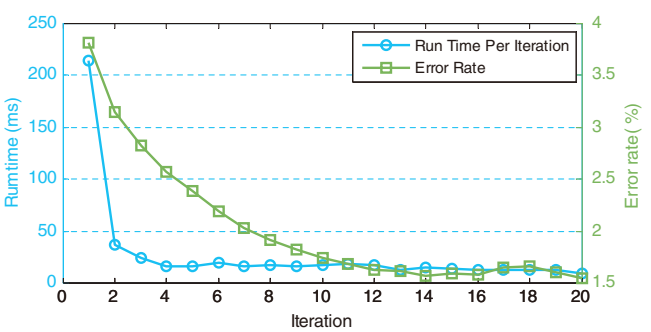

Fig. 10 Convergence of running time and error rate.

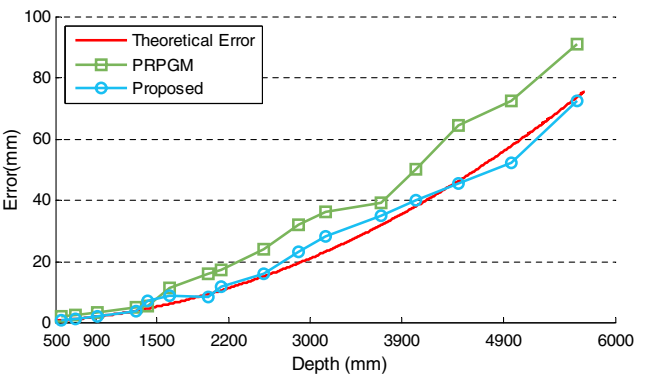

Fig. 11 Plane test for error analysis. Theoretical error is computed assuming that the disparity estimation error is 0.2 pixel. Depth is showed in logarithmic coordinate.

$$d_{\text{sub}} = \begin{cases} d + (\Delta L/\Delta R - 1)/2, & \text{if } \Delta L \le \Delta R \\ d - (\Delta R/\Delta L - 1)/2, & \text{if } \Delta L > \Delta R \end{cases}. \quad (17)$$

According to the triangular geometry, we can derive depth Z from disparity d_{sub} as

$$Z = \frac{s}{d_{\text{sub}} + s/Z_0}, \quad (18)$$

where s is a constant determined by the focal length of the camera and the equivalent baseline length, and Z_0 is the depth of the reference plane.

5 Experimental Evaluation

To evaluate the performance, we implement our approach in C++ on a desktop computer with 2.50 GHz Q9300 CPU and 4GB RAM. First, the effectiveness of pattern extraction and binary feature is presented. Then we show the efficiency and accuracy of the algorithm. Last, we give more results of several different scenes and compare our approach with other methods. Throughout all experiments, the parameters are fixed as $W_s = 5$, $\lambda = 0.05$, $W_f = 15$, $\sigma = 0.5$, $\beta = 0.05$, $\text{TH}_E = 100$, $\text{TH}_{\text{Conf}} = 24$.

5.1 *Speckle Pattern Extraction and Binary Feature*

As shown in Fig. 7, except for defocus blurred and saturated regions, most of the direct components can be extracted clearly. To show the effectiveness of our speckle pattern extraction method and binary feature, we test the algorithm on a typical scene that contains surfaces of different depths. We run the algorithm with or without the speckle pattern extraction using the binary or intensity feature, four combinations in total. The comparison is presented in Fig. 8.

From Fig. 8, it can be observed that the result is the worst for the scheme without pattern extraction and with intensity feature, while the result is the best for our scheme with the pattern extraction and with binary feature. Especially, the result is greatly improved by our scheme in faraway regions. In these regions, the low contrast between the speckle pattern and the ambient lighting make the matches ambiguous.

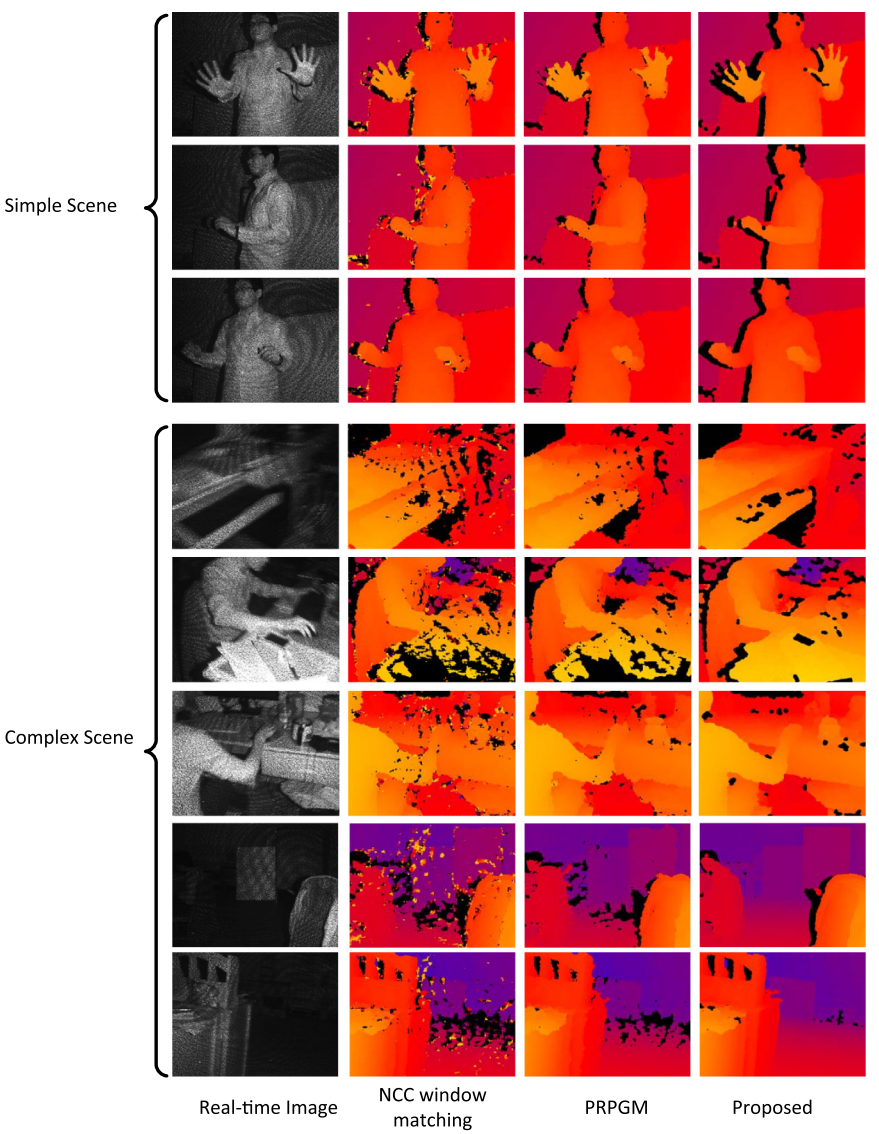

Fig. 12 Depth estimation under different scenes. The proposed method is compared with the NCC window matching and PRPGM method.

As the speckle pattern extraction can get rid of the ambient lighting's influence and the binary feature is insensitive to contrast, the combination of them can significantly reduce the ambiguities. It can also be seen from the results that depth stripes are more obvious when using the intensity feature, though subpixel interpolation has been applied to all setups. This indicates that the binary feature can achieve a higher accuracy than the intensity feature.

5.2 *Efficiency Analysis*

To show the efficiency of the proposed approach, we test the algorithm on a scene with one person inside. Figure 9 shows some results during iterations and the error map of the final result. It can be seen that our iterative message-passing scheme make the information propagate efficiently between blocks. The experimental results show that more reliably matched points are added earlier, which is in agreement with our analysis earlier. From the error map, we can see that the proposed approach is rather accurate except for some complex boundaries, e.g., the hands of the human.

Figure 10 presents the convergence of the error rate and running time. The error rate is the percentage of bad pixels whose disparity error is over some tolerance (1 pixel in this paper). We convert depth into disparity with the inverse transform of Eq. (18) to compute error rates. It can be seen from Fig. 10 that after about 12 iterations, the error rate has converged to no more than 1.7%. The running time per iteration decreases so sharply that it almost converged to a constant after four iterations. Hence, we choose to terminate the process after 12 iterations, for the balance between accuracy and efficiency. On average, the proposed approach can process more than 3 frames/s with the image resolution of 640×480.

5.3 *Plane Test and Accuracy Analysis*

We test the algorithm on a series of planes to analyze the estimation accuracy. These images are captured with planes perpendicular to the optical axis of the camera at different distances. By assuming that the mean value of the depth map is an unbiased estimator for the plane, the standard deviation is used to measure the estimation error. According to the analysis in Ref. 20, the theoretical random error of depth measurement is proportional to the square distance from the sensor to the object.

We compare the error of proposed method with the progressive reliable points growing matching (PRPGM) method introduced by Wang et al.,[9] and the results are plotted in Fig. 11. The theoretical error is computed by assuming that the disparity estimation error is 0.2 pixel. The error curve shows that the error of the proposed algorithm coincides with the theoretical error. For long-range planes, the proposed method performs better than the PRPGM method and can achieve a gain about 3 dB in the absolute error. Besides, the largest relative error of our method is still <1.5%, which verifies that the proposed approach can achieve an acceptable accuracy for different depth layers.

5.4 *Scene Test*

To demonstrate the wide applicability of our approach, we test our method on two videos. The first video is acquired with a fixed background and a human inside; the second is acquired with complex scenes, including different kinds of objects and non-Lambertian surfaces, such as LCD.

We compare our approach with the normalized cross correlation (NCC)-based local window matching and the PRPGM method, both introduced in Ref. 9. To make the

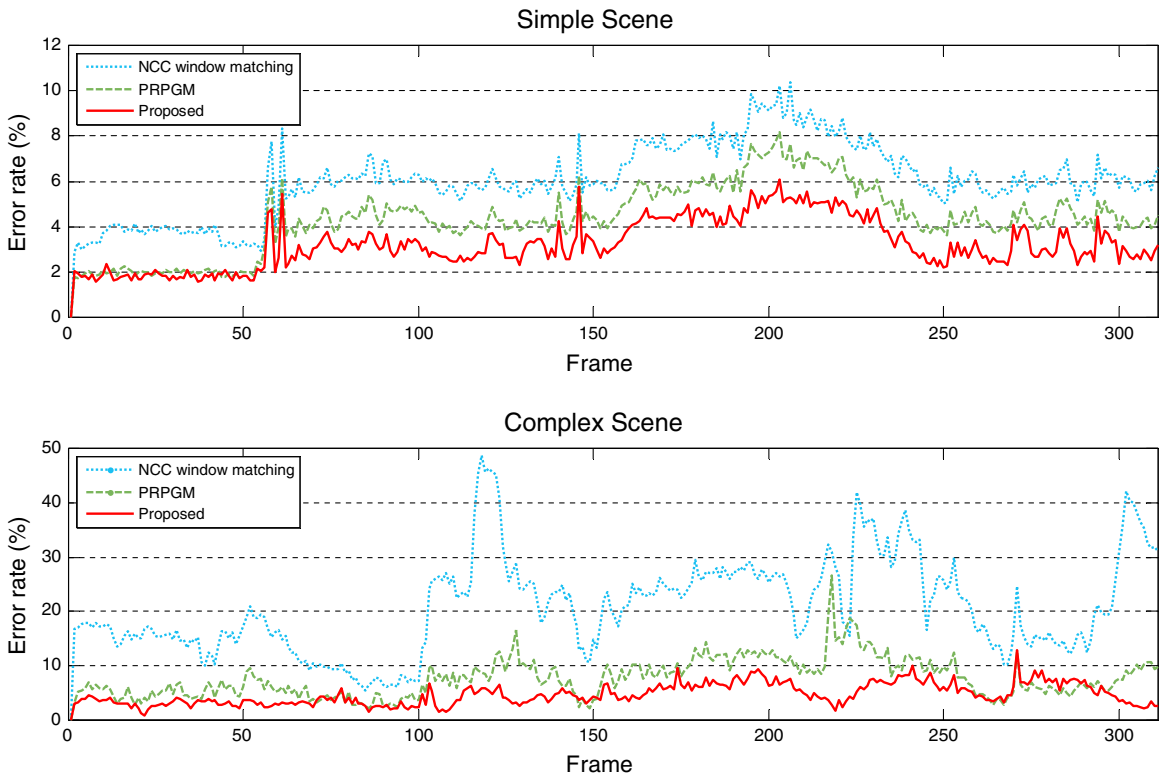

Fig. 13 Error rate under different scenes, compared with NCC window matching and PRPGM.

comparison fair, we apply the same speckle pattern extraction and subpixel interpolation to all three methods. Some representative frames are presented in Fig. 12, from which we can observe the following.

First, even though the speckle pattern extraction and the standard postprocessing are applied, there are still many mismatches left in the outcome of the NCC window matching. This indicates that the NCC method is not suitable to handle the speckle pattern, which is of high contrast.

Second, the PRPGM method behaves better than the NCC method; it has fewer holes left. This is because the reliably matched points can act as seeds to pass their disparity information to their neighbors. However, as the seeds only send out messages to their immediate neighbors, the information spreads very slowly. Points in regions of fewer seeds, such as boundaries, may receive messages from incorrect seeds. Even worse, if they are too far away from the seeds, they can hardly receive useful messages. The low speed of message passing may lead to inaccurate boundaries and unmatched patches, in particular, for complex scenes.

Third, the proposed method can cope with these problems. Through the cross-block support and the iterative model update, messages can pass on a scale of the block size. Therefore, regions with fewer support points can receive information from their neighboring blocks. Even the points in blocks of few support points can still gain substantial information after some iterations. Hence, more accurate boundaries can be obtained, and there are no unmatched patches left except for the regions where the speckle pattern cannot be extracted. The right-hand column of Fig. 12 demonstrates that our approach is widely applicable for different kinds of scenes.

Figure 13 presents the error rate of the three methods over the two videos. The proposed method can achieve the best performance for most of the cases. In the simple scenes, the proposed method can reduce about 2% of the error points compared with the PRPGM method; in complex scenes, an even greater gain can be achieved.

6 Conclusion and Future Work

In this paper, we have presented an active system to achieve effective depth sensing. We have shown that the speckle pattern extraction can effectively remove the ambient lighting, and the binary feature used is robust to contrast variation. The experiments have demonstrated that our iterative message-passing and model updating scheme can refine the disparity map efficiently, and our system is widely applicable. More importantly, the conditional independence in our graphical model makes it possible to process the points in parallel; thus the algorithm can be accelerated with graphics processing unit and field programmable gata array. To increase the robustness of our method, we intend to search for more robust preprocessing method, and a further plan is to refine the depth map by using the texture information of the scene.

Acknowledgments

This work is partially supported by NSFC 61271390.

References

1. F. Chen, G. M. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," *Opt. Eng.* **39**(1), 10–22 (2000).
2. J. P. Siebert and S. J. Marshall, "Human body 3D imaging by speckle texture projection photogrammetry," *Sensor Rev.* **20**(3), 218–226 (2000).
3. J. Garcia et al., "Three-dimensional mapping and range measurement by means of projected speckle patterns," *Appl. Opt.* **47**(16), 3032–3040 (2008).
4. Y.-S. Chen and B.-T. Chen, "Measuring of a three-dimensional surface by use of a spatial distance computation," *Appl. Opt.* **42**(11), 1958–1972 (2003).
5. M. Schaffer et al., "High-speed three-dimensional shape measurements of objects with laser speckles and acousto-optical deflection," *Opt. Lett.* **36**(16), 3097–3099 (2011).
6. B. Freedman et al., "Depth Mapping Using Projected Patterns," US Patent 20,100,118,123 (2010).
7. A. Shpunt and Z. Zalevsky, "Depth-Varying Light Fields for Three Dimensional Sensing," US Patent 8,374,397 (2013).
8. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision* **47**(1–3), 7–42 (2002).
9. G. Wang et al., "Depth estimation for speckle projection system using progressive reliable points growing matching," *Appl. Opt.* **52**(9), 516–524 (2013).
10. C. Shi et al., "Stereo matching using local plane fitting in confidence-based support window," *IEICE Trans. Inf. Syst.* **E95-D**(2), 699–702 (2012).
11. S. K. Nayar et al., "Fast separation of direct and global components of a scene using high frequency illumination," *ACM Trans. Graph.* **25**(3), 935–944 (2006).
12. A. Shpunt and Z. Zalevsky, "Three-Dimensional Sensing Using Speckle Patterns," US Patent 20,090,096,783 (2009).
13. R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European Conf. Computer Vision*, Stockholm, Sweden, pp. 151–158 (1994).
14. A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Computer Vision-ACCV*, Queenstown, New Zealand, pp. 25–38 (2010).
15. C. Shi et al., "An interleaving updating framework of disparity and confidence map for stereo matching," *IEICE Trans. Inf. Syst.* **E95-D**(5), 1552–1555 (2012).
16. M. Gong and Y.-H. Yang, "Fast unambiguous stereo matching using reliability-based dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(6), 998–1003 (2005).
17. N. Sabater, A. Almansa, and J. M. Morel, "Meaningful matches in stereovision," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(5), 930–942 (2012).
18. X. Hu and P. Mordohai, "A quantitative evaluation of confidence measures for stereo vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2121–2133 (2012).
19. I. Haller and S. Nedevschi, "Design of interpolation functions for subpixel-accuracy stereo-vision systems," *IEEE Trans. Image Process.* **21**(2), 889–898 (2012).
20. K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors* **12**(20), 1437–1454 (2012

Xuanwu Yin received his BS degree from Tsinghua University, Beijing, China, in 2011. He is currently working toward his PhD degree in electronic engineering with Tsinghua University, Beijing, China. His research interests include passive/active depth sensing technique, 3-D reconstruction, and computational imaging.

Guijin Wang received his BS and PhD degrees (with honor) from Tsinghua University, China, in 1998 and 2003, respectively, all in electronic engineering. From 2003 to 2006, he was a researcher at Sony Information Technologies Laboratories. Since October 2006, he has been with the Department of Electronic Engineering, Tsinghua University, China, as an associate professor. His research interests focus on wireless multimedia, depth sensing, pose recognition, intelligent human-machine UI, intelligent surveillance, industry inspection, and online learning.

Chenbo Shi received his BS and PhD degrees from the Department of Electronic Engineering, Tsinghua University, China, in 2005 and 2012, respectively. From 2008 to 2012, he published over 10 international journal and conference papers. He is the reviewer for several international journals and conferences. Now, he is a postdoctoral researcher in Tsinghua University. His research interests are focused on image stitching, stereo matching, matting, object detection, and tracking.

Qingmin Liao received his PhD degree in signal processing and telecommunications from the University of Rennes, France, in 1994. Currently, he is a professor and head of the Laboratory of Visual Information Processing at Graduate School at Shenzhen, Tsinghua University, Shenzhen, China, where he became interested in image/video analysis, computer vision and its applications.