Vis Comput (2015) 31:1587–1600
DOI 10.1007/s00371-014-1035-1

ORIGINAL ARTICLE

Compressive environment matting

Qi Duan · Jianfei Cai · Jianmin Zheng

Published online: 30 September 2014

Abstract The existing high-quality environment matting methods usually require the capturing of a few thousand sample images and spend a few hours in data acquisition. In this paper, a novel environment matting algorithm is proposed to capture and extract the environment matte data effectively and efficiently. First, the recently developed compressive sensing theory is incorporated to reformulate the environment matting problem and simplify the data acquisition process. Next, taking into account the special properties of light refraction and reflection effects of transparent objects, two advanced priors, group clustering and Gaussian priors, as well as other basic constraints are introduced during the matte data recovery process to combat with the limited image samples, suppress the effects of the measurement noise resulted from data acquisition, and faithfully recover the sparse environment matte data. Compared with most of the existing environment matting methods, our algorithm significantly simplifies and accelerates the environment matting extraction process while still achieving high-accurate composition results.

Keywords Picture/image generation · Modeling and recovery of physical attributes

Q. Duan · J. Cai (✉) · J. Zheng
School of Computer Engineering,
Nanyang Technological University, Singapore, Singapore
e-mail: asjfcai@ntu.edu.sg

Q. Duan
e-mail: duanqi1983@gmail.com

J. Zheng
e-mail: asjmzheng@ntu.edu.sg

1 Introduction

Image matting and composition are two fundamental techniques in digital image editing, which can extract a foreground object with arbitrary shape and property from a source image and composite it into a new environment seamlessly. Conventional image matting and composition techniques handle opaque objects well but fail in transparent objects, which have complex light refraction and reflection effects. Environment matting and composition algorithms were therefore developed, which generalize the conventional matting and composition by incorporating the information of how a foreground object refracts and reflects light from the environment.

The concept of environment matting and composition was first introduced by Zongker et al. [32]. They demonstrated the ability of environment matting and composition in simulating light reflection, refraction and scatter effects, which greatly improves the composition results for transparent objects. Since then, various methods [4,20,21,30,33] have been proposed to either improve the accuracy further by introducing more complex processing or simplify the process to achieve real-time environment matting. Although the real-time method [4] has very low complexity in data acquisition and matte extraction, it is only applicable to colorless and purely specular transparent objects and cannot achieve high-accuracy environment matting and composition. On the other hand, the existing high-accuracy environment matting methods usually require the capturing of a few thousand sample images and spend a few hours to complete the data acquisition (see Table 1), which is very undesired for practical applications.

To reduce the complexity of data acquisition while still achieving high-quality environment matting, in this paper, we incorporate the recently developed compressive sensing

Table 1 Comparisons among different environment matting algorithms with a background image size of $n \times n$ and k denoting the sparsity of the light transport vector

Methods	# of captured images when $n = 512$	Data acquisition method	Feature Suitable transparent object
Zongker's method [32]	$\mathcal{O}(\log n)$		Support single-region mapping
	18 images	Image capture	General transparent object
Real-time method [4]			Support one-point mapping
	2 images	Image capture	Colorless and pure specular object
Chuang's method [4]	$\mathcal{O}(n)$		Support multiple-region mapping
	900 images	Image capture	General transparent object
Wavelet method [21]	$\mathcal{O}(n)$		Decompose background into patterns
	2400 images	Image capture	General transparent object
Wavelet noise method [22]			Decompose background into patterns
	256 HDR images	Image capture	General transparent object
Frequency method [33]	$\mathcal{O}(n)$		Support multiple-region mapping
	At least 2048 images	Image capture	General transparent object
Our method w/o Gaussian prior	$\mathcal{O}(k \log k)$		Support multiple-region mapping
	400 images	Video capture	General transparent object
Our method with Gaussian prior	$\mathcal{O}(k \log k)$		Support multiple-region mapping
	250–300 images	Video capture	General transparent object

technology into the environment matting problem. In particular, the environment matting extraction problem is converted into a sparse light transport vector recovery problem, which can be well solved by compressive sensing techniques with a small number of measurements. Moreover, to have an efficient and robust solution, our compressive environment matting design takes into account the important properties of the light transport vector such as group clustering and Gaussian prior. We also propose a hierarchical method to reduce the computational cost. The data acquisition in our design becomes a simple process of capturing a 3–4 min video, which is much easier than the existing methods that capture a few thousand images.

We would like to point out that the environment matting problem can be considered as a special case of image-based relighting and the compressive sensing theory has been applied in the general image relight problem in [24,26]. However, we argue that when applying compressive sensing in the specific problem of environment matting, there are some unique features. Our major contribution lies in the proposed solution to the compressive environment matting problem, which has some neat designs including hierarchical sampling, group-clustering-based recovery, Gaussian-prior-based denoising, and video-based data capturing that make the proposed compressive environment matting framework more practical, efficient and effective. Note that a preliminary version of this work has been published in [5]. Compared with the previous conference version, this paper reports more technical details and incorporates one additional prior, i.e. the Gaussian prior, which significantly improves the performance.

The rest of the paper is organized as follows. Section 2 reviews various related works and techniques including environment matting, light transportation measurement and compressive sensing theory. Section 3 shows how the classical environment matting problem can be converted and formulated into a problem of recovering sparse vectors and describes its related properties. In Sect. 4, we propose a hierarchical method to solve the compressive environment matting problem. Experimental results are shown in Sect. 5. Finally, we conclude the paper in Sect. 6.

2 Related work

2.1 Environment matting

Zongker et al. [32] first gave a mathematical formulation about the environment matting problem and used a sequence of structured backgrounds to estimate the environment matte data. The structured backgrounds consist of a hierarchy of finer and finer horizontal and vertical square-wave stripes. Later on, Chuang et al. [4] proposed two extensions to enhance the usability of the environment matting algorithm [32]. The first extension aims to improve the accuracy of the matting results, where 2D oriented Gaussian pattern instead of axis-aligned rectangle is used to recover light spatial variation and dispersion effects as well as multiple mappings of texture, which can better approximate the BRDF

(bidirectional reflectance distribution function) model. The second extension is targeted to achieve fast/real-time environment matting by making certain assumptions such as colorless and specularly refractive transparent objects and simplifying the matting process to the one with only one picture captured against a special backdrop.

There exist other solutions for environment matting and composition. Wexler et al. [30] used a probabilistic model to extract the matte by assuming that each background pixel has a probability of contributing to the final color of certain foreground pixel. Zhu and Yang [33] introduced a method to find an accurate matching between foreground pixels and the time-sequence backdrop in frequency domain based on Fourier analysis, which is more precise than spatial-domain matching [4,32]. Peers and Dutré [21,22] used wavelet background patterns and wavelet processes to calculate environment matting data. Most of these methods, especially the wavelet method [21] and frequency-based method [33], require a large number of image samples, which makes the capture process very complex and time-consuming (see Table 1).

2.2 Light transportation measurement

As pointed out in [32], environment matting is highly related to the image-based relighting problem, which aims to simulate the direct and indirect environment light illumination. Ideally to fully describe the light transportation property, an 8D function should be constructed because both incident and outgoing light fields are 4D functions including positions (2D) and all possible directions (2D) [17]. Considering that such 8D function is very difficult and time-consuming to measure, usually simplifications are made for practical applications. For example, only 4D light functions were measured in [7,18,32] by fixing the camera viewpoint and assuming 2D incident light from monitor or projector. Later, 6D light functions were measured in [19,25]. An 8D light function was approximately recovered by exploiting the symmetry property of light transport process in [12]. Considering thousands of sample images are needed for recovering the light field, many techniques have been introduced to accelerate the data acquisition process or reduce the number of sample images, including hierarchical methods [18,21,25], adaptive method [9], kernel-based method [29] and compressive sensing methods [24,26]. A detailed comparison between compressive environment matting and compressive image relighting will be discussed in Sect. 5.4.

2.3 Compressive sensing

The recently developed compressive sensing theory (CS) demonstrates an efficient way to reconstruct sparse signal with a small number of measurements. Suppose there is an unknown vector $x \in \Re^n$ and we have its measurement result $y = \Phi x$ where Φ is a $m \times n$ measurement matrix. If $m < n$, it is an ill-conditioned problem to recover a general x from y. But if x is a k-sparse signal with at most k ($k \ll n$) non-zero elements, we can recover the signal x by solving the following optimization problem:

$$\text{argmin } \|x\|_0, \text{ subject to } y = \Phi x \tag{1}$$

which is an NP-hard problem and can only be solved by the greedy algorithm. The main drawbacks of the greedy methods are that the reconstructed result is often suboptimal and it may be trapped into local optimal if the initial guess is not good. Furthermore, obviously it is very sensitive to measurement noise.

According to the compressive sensing theory [2,3,8], if the measurement matrix Φ satisfies the restricted isometry property (RIP) [3], the above NP-hard problem is equivalent to its convex relaxation

$$\text{argmin } \|x\|_1 \text{ subject to } y = \Phi x \tag{2}$$

which can guarantee a global optimal solution and can be reliably solved in polynomial time with only $\mathcal{O}(k\log k)$ non-adaptive measurements.

Note that measurement matrix Φ with RIP means that for any k-sparse vector x and arbitrary constant $\varepsilon \in (0, 1)$, the following inequality holds:

$$(1-\varepsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\varepsilon)\|x\|_2. \tag{3}$$

This inequality indicates that the eigenvalues of any submatrix of Φ are very close to 1. No matter where the non-zero elements are in the sparse vector, they will be equally sampled, which guarantees the recovery of the sparse signal with insufficient number of measurements. The popular matrices satisfying the RIP include Gaussian matrices, Bernoulli matrices and partial Fourier matrices [1].

Due to the excellent performance of compressive sensing in recovering sparse signals, it has been widely used in many applications, such as signal processing [10,23], image reconstruction [6,31], medical image analysis [13,16], shape prior modeling [34], so on. Also in many computer graphics research fields, compressive sensing helps to reduce the time complexity and accelerate the existing algorithms, for example, simulate light transport in volumetric media [11], accelerate existing ray tracing and rendering methods [27,28] and reduce the complexity for image-based relighting [24] and dual photography [26]. In the next part of this paper we will show how compressive sensing benefits the environment matting problem.

3 Environment matting model and properties

We begin by examining the environment matting equation introduced in [32], i.e.

$$C = F + (1-\alpha)B + \sum_{i=1}^{m} \mathcal{R}_i \mathcal{M}(T_i, A_i) \tag{4}$$

where C, F and B are the colors of image, foreground, background respectively, α is the opacity value of the foreground object, $\mathcal{R}_i$ is the reflectance that describes the contribution of light emanating from environment to the object, m is the number of texture maps, and $\mathcal{M}$ is the "texture-mapping operator" calculating the average color of region A_i in background texture T_i. In general, this model of Eq. (4) is comprehensive because it takes into account not only the foreground and background information but also the light refraction and reflection effects.

Considering the practical scenario of capturing sample images of transparent objects with one background monitor or projector and no other environment light source, F can be ignored since the camera only captures the background light that is not absorbed by the transparent object. Moreover, the second item and the third item in Eq. (4) can be combined because they both describe the contributions from the background light source, although one is for direct transmission and the other is for refraction and reflection.

Based on the above analysis, similar to [21,22], we model the environment matting and composition problem as

$$C = S + \rho \mathcal{T} \mathcal{L} \tag{5}$$

where ρ, called absorption index, is a scalar for one color channel describing the light absorption effects, $\mathcal{L}$ is a $N \times 1$ vector representing the N-dimensional background image and $\mathcal{T}$ is the light transport vector of $1 \times N$ illustrating the contribution of light emitted from each background region to an object pixel. Note that in Eq. (5), we also introduce an offset term S, which is to compensate the deviations caused by uncalibrated monitor and camera.

In this way, the environment matting problem becomes: *given a set of background images $\mathcal{L}$ and the corresponding captured image colors C, how to recover the light transport vector $\mathcal{T}$ as well as ρ and S*? After obtaining all these parameters, we can easily simulate the light refraction and reflection effects under any new background image.

3.1 Property of light transport vector

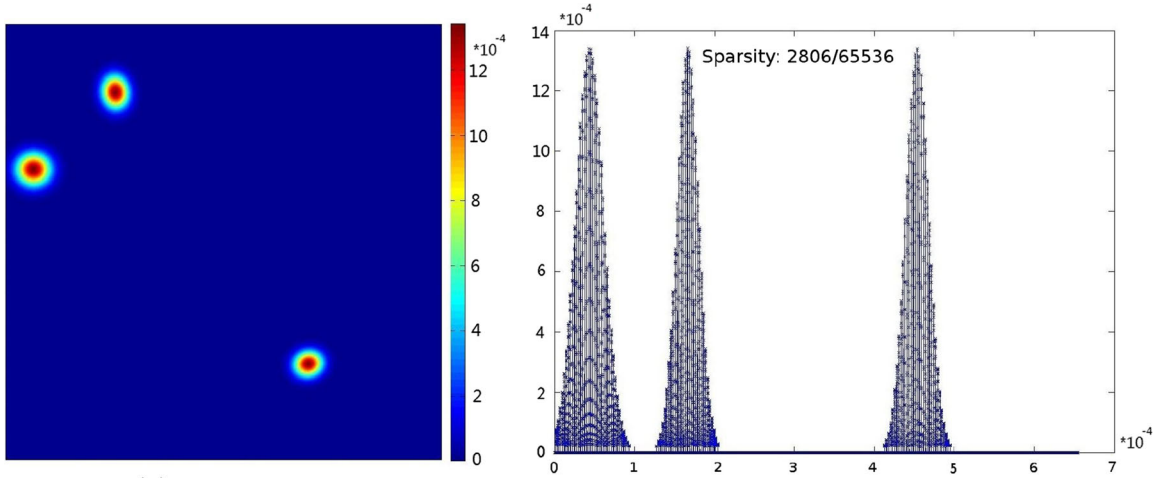
Apparently, the essence of our environment matting model is to recover the light transport vector $\mathcal{T}$. Before introducing the recovery process, we summarize the important properties of the light transport vector as follows.

- *Value range*: Each element of $\mathcal{T}$ represents the percentage of contribution from a background pixel. Thus, elements of $\mathcal{T}$ should be non-negative and the sum of all the elements is equal to 1, i.e. $\|\mathcal{T}\|_1 = 1$.
- *Sparsity*: Typically, for a transparent object, there are only a small number of environment regions making contribution to a foreground pixel. That means the light transport vector is a sparse vector, where only elements for those limited background regions are non-zero while the others are all zeros.
- *Group clustering*: In addition to sparsity, another interesting phenomenon is that the regions that contribute to certain foreground pixel are not randomly or uniformly distributed over the whole background plane. Actually most of them are neighboring and can be divided into several main parts (e.g. refraction part, reflection part and transmission part), which has been proven in [4,30,32]. This suggests that the non-zero elements $\mathcal{T}$ can be clustered into several local groups, which is a very useful prior for accelerating the CS-based recovery [14].
- *Gaussian prior*: More precisely, the light refraction and reflection effect of a transparent object should satisfy the BRDF property [4,15], which means that the non-zero elements of $\mathcal{T}$ should have 2D Gaussian distribution. This is another very important constraint which is used later to verify the accuracy of our matte result.

To give a clear illustration of the light transport vector and its properties, we generate a composition result artificially and construct its corresponding light transport vector shown in Fig. 1. Particularly, the 256×256 contribution map in Fig. 1 visualizes the contribution coefficient of each subregion in the background plane, where it is assumed that there are three separate regions (light refraction, reflection and transmission parts) contributing to the composition color result of one foreground pixel and the weights of each region are distributed according to the oriented Gaussian distributions [4]. Figure 1 shows the corresponding light transport vector $\mathcal{T}$ obtained by scanning the contribution map column by column. It can be seen that $\mathcal{T}$ is a sparse and non-negative vector with less than 5 % non-zero elements clustered in several local groups and $\|\mathcal{T}\|_1 = 1$.

4 Solving environment matting problem

Now we consider how to solve the environment matting problem. According to our model of (5), for each foreground pixel, there are three variables (scalar S, ρ and vector $\mathcal{T}$) for each color channel and totally nine variables for RGB channels. It is hard to solve them all at one time. Thus, we choose a two-step approach to recover the unknowns. Once S, ρ and $\mathcal{T}$ are recovered, given any new background image, we can



(a) Contribution map **(b)** Corresponding light transport vector

Fig. 1 Simulation of light refraction, reflection and transmission and the corresponding light transport vector. The background resolution is 256 × 256 and there are three separate parts contributing to certain foreground pixel with oriented Gaussian distribution for the weights of each region. The corresponding light transport vector is sparse with only 2,806 non-zero elements out of 65,536 elements

easily synthesize the resulting image of the transparent object under the new background.

4.1 Recovery of compensation and light absorption index

At the first step, we aim to recover the compensation term S and the light absorption index ρ. Inspired by Zongker's methods, we project some constant image to the transparent object so as to eliminate the influence of light refraction and reflection. Since the background plan has constant color, $\mathcal{L}$ can be written as the product of an intensity value b with an identity vector I. Further considering $\|\mathcal{T}\|_1 = 1$, we obtain $\mathcal{T} \cdot \mathcal{L} = b$. Thus, Eq. (5) becomes

$$C = S + \rho b. \tag{6}$$

In this way, we can easily recover S and ρ for each color channel of every object pixel by capturing several sample images of the transparent object in front of constant background with different intensity values of b (see Fig. 2 for some examples).

4.2 Light transport vector recovery

Since $\mathcal{T}$ is a sparse vector, it can be faithfully reconstructed with small number of measurements using compressive sensing algorithms. Consider that $\mathcal{T}$ is composed of the contribution coefficients of all the background regions to one foreground object pixel. By capturing a foreground pixel color result C under a given background pattern $\mathcal{L}$, we obtain one measurement of $\mathcal{T}$. If we construct an appropriate matrix $\mathcal{L}_{\mathcal{M}}$ with RIP, where each column is one background image pattern, to measure $\mathcal{T}$ many times and record all the resulted foreground pixel color C, $\mathcal{T}$ can then be recovered accurately through compressive sensing.

4.2.1 Background pattern generation

As pointed out in [1], only some special random matrices can be used as measurement matrix in compressive sensing. In our implementation, similar to [26], we choose the

Fig. 2 Sample images of a transparent object in front of constant background images with different colors

Bernoulli matrix, a binary matrix (containing either $+1$ or -1) drawn from a Bernoulli distribution. Considering a resolution of N for the background, we generate an $N \times M$ Bernoulli matrix $\mathcal{L}_{\mathcal{M}}$, where M is the number of measurement. For each column of $\mathcal{L}_{\mathcal{M}}$, a corresponding background pattern image is generated by mapping individual $+1$ and -1 elements to white and black colors, respectively. Figure 3 shows the captured images of a transparent object in front of the random background patterns with different resolutions.

4.2.2 Recovery by compressive sensing

Under different background patterns, a series of sample images of the transparent object can be captured (see Fig. 3). For each foreground object pixel, we can rewrite Eq. (5) into a matrix form:

$$\mathcal{C} = S \cdot I + \rho \mathcal{T} \mathcal{L}_{\mathcal{M}}, \tag{7}$$

where $\mathcal{C}$ now is a $1 \times M$ vector recording all the captured colors of the pixel under different background patterns, and I is a $1 \times M$ identity vector. Note that the $N \times M$ matrix $\mathcal{L}_{\mathcal{M}}$ is not the originally generated Bernoulli matrix, but the corresponding background image patterns, where each background image pattern is sampled to the same resolution as the captured images. Then, the target sparse vector $\mathcal{T}$ can be recovered by solving the following optimization according to the compressive sensing theory:

$$\min \|\mathcal{T}\|_0, \quad \text{s.t. } \mathcal{T}\mathcal{L}_{\mathcal{M}} = \frac{\mathcal{C} - S \cdot I}{\rho}. \tag{8}$$

However, such a general solution is not cost-effective since it does not consider the properties of the light transport vectors.

Therefore, to make sure $\mathcal{T}$ is recovered faithfully and efficiently, during the recovery process we further add in the constraints based on the properties of $\mathcal{T}$ described in Sect. 3.1:

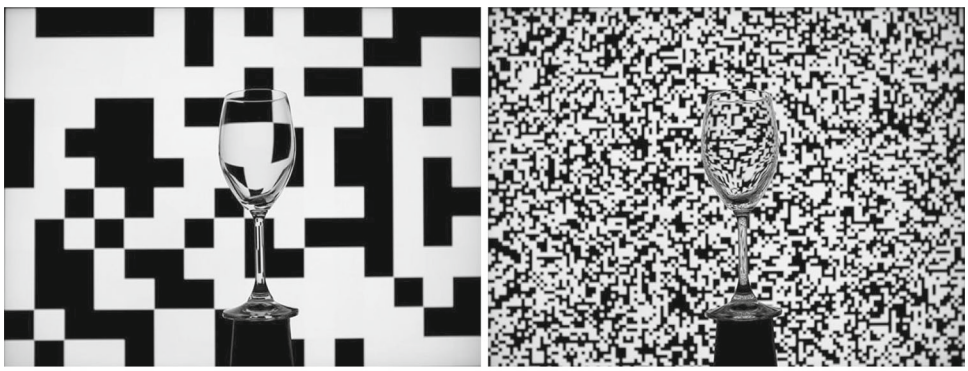
$$\min \|\mathcal{T}\|_0, \quad \text{s.t.} \begin{cases} \mathcal{T}\mathcal{L}_{\mathcal{M}} = \frac{\mathcal{C} - S \cdot I}{\rho} \\ \forall i,\ t_i \geq 0 \\ \|\mathcal{T}\|_1 = 1 \\ \mathcal{T} \sim \mathcal{N}(c_x, c_y, \sigma_x, \sigma_y, \theta) \end{cases} \tag{9}$$

where t_i is the i-th element of $\mathcal{T}$, and $c_x, c_y, \sigma_x, \sigma_y, \theta$ are center, variance and rotation angle parameters of 2D Gaussian functions. Note that the last constraint in (9) means that the non-zero elements of $\mathcal{T}$ should have 2D Gaussian distributions, which comes from the Gaussian prior and is important for noise removal. Most of the current compressive sensing algorithms deal with the problem of computing a sparse vector from "clean" measurement samples. However, for practical applications such as environment matting, the measurements often contain complex noise, which is hard to be identified or modeled. It is necessary to utilize the prior, i.e. the Gaussian prior here, to suppress the measurement noise by removing the non-zero elements in $\mathcal{T}$ that do not satisfy the Gaussian distribution property.

Moreover, the group clustering prior is directly used when solving (9). Particularly, we treat all the non-zero elements as some local groups instead of several separate ones. As pointed out in [14], for a k-sparse vector $\mathcal{T} \in \mathbb{R}^n$, $\mathcal{O}\left(k\log(\frac{n}{k})\right)$ measurements are needed; but if k non-zero elements can be grouped into q groups, the number of required measurements can be reduced drastically to $\mathcal{O}\left(k + q\log(\frac{n}{q})\right)$. This is exactly the case in our problem, where q is typically no more than three based on our experiments. Using this group clustering prior, we are able to reduce the time complexity of the data acquisition and the recovery process significantly.

4.2.3 Hierarchical measurement and recovery scheme

As we know, for the environment matting problem, to generate a photo-realistic composition result, the resolution of the background pattern N should be large enough, ideally the same as the resulting image resolution. However, this will lead to an extremely large dimension for the measure-



(a) Coarse level **(b)** Fine level

Fig. 3 Sample images of a transparent object in front of the random background patterns with different resolutions

ment matrix $\mathcal{L}_{\mathcal{M}}$ and the light transport vector $\mathcal{T}$, which subsequently results in a very expensive computational cost to solve the constrained optimization problem (9). Therefore, to accelerate the whole optimization process, we choose a hierarchical measurement and recovery scheme. The basic idea is to recover the light transport vector $\mathcal{T}$ progressively in a coarse-to-fine manner. Similar idea has also been exploited in [21,24,32]. In particular, as shown in Fig. 1, only small parts of the background area contribute to certain foreground pixel color while all the other parts are useless. Thus, an intuitive and efficient hierarchical scheme is designed as follows:

1. At the coarse level (see Fig. 3), we use compressive sensing to solve the constrained optimization problem (8) with a very low-dimensional vector $\mathcal{T}$. The recovered $\mathcal{T}$ gives a rough estimation about where the contribution regions are, i.e. the rough location of those non-zero elements.
2. At the fine level (see Fig. 3), we delete those useless dimensions in both $\mathcal{L}_{\mathcal{M}}$ and $\mathcal{T}$, increase the resolutions for the non-zero locations, and then resolve the optimization problem.

Using this hierarchical recovery method, the dimension and the computational cost of the optimization problem are reduced significantly. We would like to point out that this light transport vector recovery process is different from the measurement process, where the former is adaptive but the latter is non-adaptive. For the measurement process, the measurement matrix or the corresponding background patterns are generated and stored in advance, and thus it is non-adaptive. For the vector recovery process, we first recover the light transport vector at the coarse level, and then recover the fine-level vector at those regions which have non-zero values at the coarse level vector. Thus, the vector recovery process is adaptive.

5 Experiments

In this section, we conduct experiments to verify the effectiveness of the proposed compressive environment matting algorithm. In particular, we use Canon 5D Mark II digital camera to capture images of transparent objects and a 24" SAMSUNG LCD monitor to project background images. Unlike the measurement process in [21,25], which generates adaptive background pattern in each iteration based on analyzing the previously captured sample images, our measurement process is non-adaptive, where the measurement matrix or the corresponding background patterns are generated and stored in advance. In this way, our data acquisition is simplified into a process of capturing a video of the transparent object in front of the monitor that projects various background patterns continuously. Since there is no synchronization between the LCD and the camera, at each second only two background patterns are projected on the LCD so as to ensure the video recording can capture correct sample images.

Note that during the data acquisition process, the positions of the camera, the transparent object and the monitor are fixed so as to make sure that the unknowns in Eq. (7) remain unchanged. In addition, for easy setup, no individual calibration is performed for the LCD and the camera. Instead, we use S and ρ in (5) to implicitly include the calibration of the LCD and the radiometric calibration of the camera.

5.1 Implementation of light transport vector recovery

As described in Sect. 4.2, we use compressive sensing with hierarchical structure to recover the light transport vector. In particular, for the monitor with a resolution of $1,280 \times 960$, we set the resolutions to be 16×12 and 128×96 for the coarse level and the fine level, respectively. This corresponds to a block size of 10×10 at the fine level. Reducing the block size will improve the resolution, but at the cost of significantly increasing the dimension of the light transport vector, for which the common computer memory cannot cope. So, empirically we find that 10×10 block size is a good tradeoff between performance and memory and computation cost.

During the sparse vector recovery process, we adopt the dynamic group sparsity recovery algorithm (DGSR) [14] rather than the common ROMP method used in [24,26] due to the following two reasons.

1. For a sparse vector with group clustering prior, the DGSR algorithm can recover it accurately with less computational cost (see Table 2) and fewer measurements [14].
2. Most of the existing compressive sensing algorithms need to know the sparsity value of the vector before the recovery process starts. However, for our environment matting problem, it is impossible to know the exact sparsity value of the light transport vector $\mathcal{T}$ in advance. In addition, the sparsity values for distinct foreground points are usually different because of different shape, refraction and reflection properties. Fortunately, DGSR is not very sensitive

Table 2 Average time for the light transport vector recovery using the non-negative dynamic group sparsity recovery algorithm (NN-DGSR) with different settings

Hierarchical (seconds per 1,000 pixels)	Direct (seconds per 1,000 pixels)
NN-DGSR	
306 ± 5	538 ± 5

to the input sparsity value and can recover the group clustering information dynamically without any prior.

Considering that a sparse vector can be steadily and accurately recovered if the number of measurements is above five times of the vector sparsity, we set the sparsity value of $\mathcal{T}$ to be $1/5$ of the total measurements. Furthermore, to ensure the non-negative and sum constraints in (9), we make some modifications to the original DGSR method. Specifically, in each iteration we solve the non-negative least square problem instead of the general least square problem. The iteration stops when the L_1 norm of the currently recovered vector is close to one.

The Gaussian prior is estimated on-the-fly during the optimization. Particularly, in each iteration step of the optimization, after computing the light transport vector, we use the vector to fit an oriented Gaussian function to calculate the Gaussian parameters, and then use the computed oriented Gaussian to filter out the noise and refine the light transport vector. Since the lighting properties for different points on the transparent object may change drastically, we do not assume any Gaussian parameters in advance and instead we compute them on-the-fly.

5.2 Visual results of environment composition

We first verify the proposed algorithm using a real glass cup. In particular, we capture 24 images at coarse level (see Fig. 3a) and 100–400 sample images at fine level (see Fig. 3b) for the designed hierarchical measurement. Then, the refractive light transport vector $\mathcal{T}$ of each object pixel is recovered using our proposed algorithm and subsequently composited with a new background image. Note that since it is captured in real environment, the measurements inevitably contain some noise.

Figure 4 shows the composition results of our methods with and without Gaussian prior. It can be seen that the results with Gaussian prior are always better than that without the prior. Moreover, with the decreasing number of image samples, the gap between the two methods becomes larger. At the case with only 100 sample images, the results of the method without Gaussian prior are full of artifacts. This is mainly because the number of measurements is insufficient to recover the sparse vector $\mathcal{T}$. In contrast, the method with Gaussian prior achieves much better performance since the additional Gaussian prior is able to effectively suppress the recovery errors in $\mathcal{T}$. Note that when zooming into the results with relatively small numbers of samples, there are quite some visible tiny artifacts such as color shifts, which is mainly caused by the inaccurately recovered light transport vectors under insufficient samples. Since our method recovers light transport vectors for each pixel and each color channel independently, the wrongly recovered light transport vectors would appear randomly. In addition, there are also some regular grid-pattern artifacts exhibited in the results, which we believe is mainly due to the minimum 10×10 unit size used in the background patterns. We have also measured the mean square errors (MSE) of the results for our methods with and without Gaussian prior. With 100 sample images, our method with Gaussian prior can reduce MSE by more than 25 %.

We also conduct some experiments in simulated environment. In particular, we use some complex 3D models and obtain all the sample images through rendering the 3D models using 3Ds Max. Such a scenario can be considered as an ideal case, since the capturing environment is well controlled and the measurement noise can be ignored. Figure 5 shows the composition results of a glass elephant with a high-frequency checker box as the backdrop. It can be seen that the composition results of the method with Gaussian prior outperform those of the method without Gaussian prior. Note that although the backdrop is black and white and the elephant model is colorless, the results under small numbers of samples contain severe color noise, which becomes apparent under a close-up view. This is mainly because the light transport vectors are recovered independently for each of the RGB channels, which results in color noise at the cases with insufficient samples.

Figure 6 shows the composition results using 200 sample images for the colorful 3D elephant and budda models, which represent general transparent objects. The two colorful objects have varying colors at different parts, which means their light reflection and refraction effects are wavelength-dependent. The results of our method with Gaussian prior are very close to the ground truth, which demonstrates that our method can handle general cases as well. Figure 7 shows the composition results for some additional models.

5.3 More comparisons

To better illustrate the advantages of the proposed compressive environment matting framework, we compare our algorithm with the existing representative environment matting methods in Table 1. The information about the existing environment matting methods is directly copied or estimated from the corresponding publications.

From Table 1, we can see that the first two methods require less number of sample images but can only recover single-region or single-point mapping for each foreground pixel, which is certainly limited for high-quality environment matting and composition. Moreover, the real-time method [4] can only handle colorless and purely specular transparent objects. Among the other methods that can recover multiple-region mapping effects, our approach is the best one in terms of the number of required sample images and the data acqui-

Fig. 4 Composition results of a real cup using our proposed methods with and without the Gaussian prior, where the former significantly outperforms the latter in terms of visual quality at the same number of sample images. Note that the zoom-in views of the *red box regions* are displayed at the *last row*

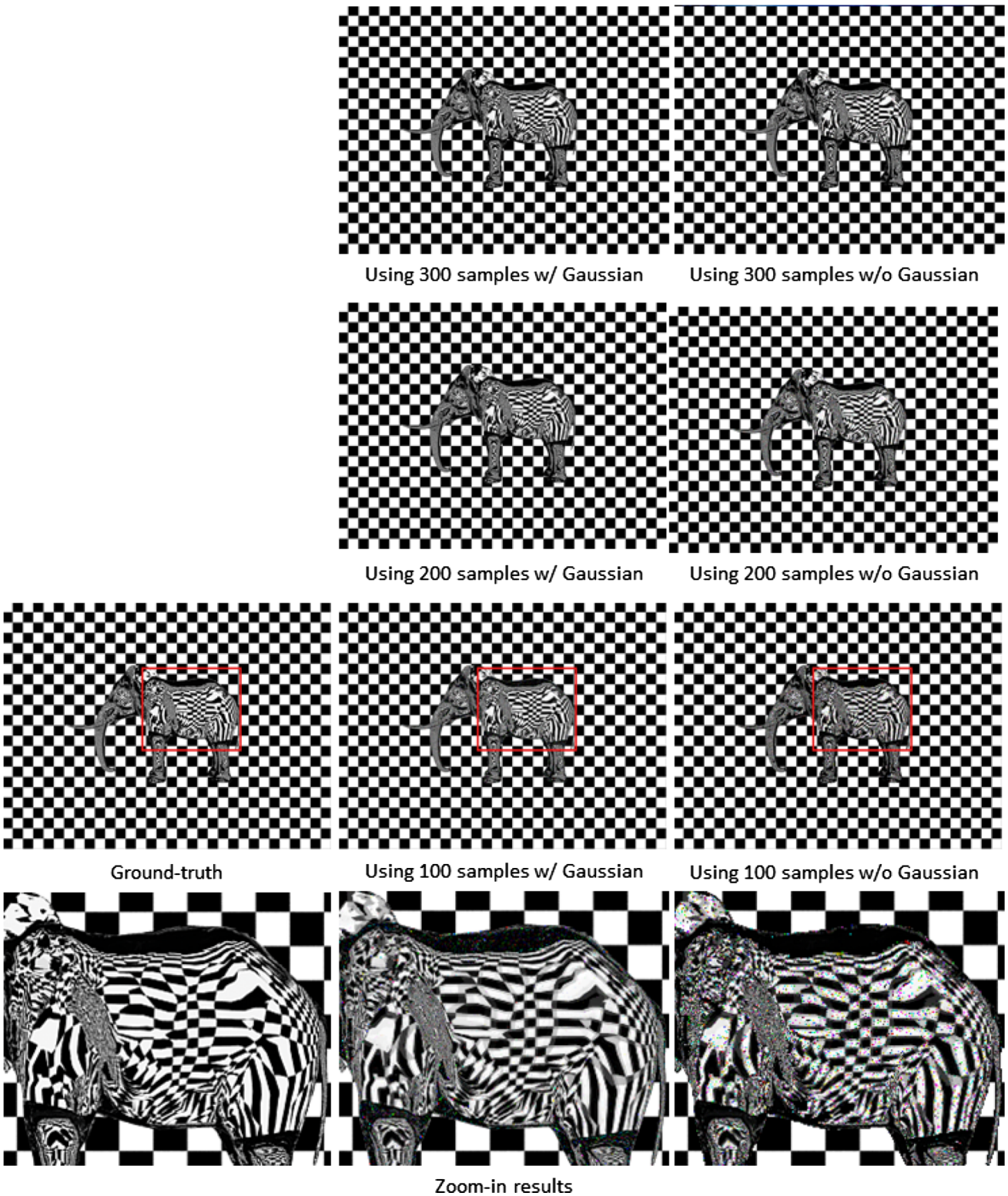


Fig. 5 Composition results of a colorless 3D elephant model using our proposed methods with and without the Gaussian prior. Note that all sample images and ground truth are rendered using 3Ds Max. The zoom-in views of the *red box regions* are displayed at the *last row*

sition complexity. Note that the wavelet noise method in [22] requires similar number of sample images as ours. However, compared with [22], our algorithm is more practical for general scenarios because we use only simple binary patterns and can handle limited dynamic range situation to improve SNR of the results while [22] needs to exploit wavelet com-

Fig. 6 Composition results of colorful 3D elephant and budda models with 200 sample images. Note that all sample images and ground truth are rendered using 3Ds Max. The second row and the fourth row are the zoom-in views of the corresponding *red box regions*

pressibility of the light effects, generate complicated wavelet basis pattern and be restricted to capture HDR sample images, which might require capturing several images to generate one corresponding HDR image. Moreover, [22] may not guarantee the best recovery accuracy because their energy splitting strategy is suboptimal and may ignore some possible sub-coefficients. In addition, we can easily remove some coefficients in our hierarchical scheme to accelerate the recovery process while [22] cannot. The scale of the number of sample images for our method only depends on the sparsity of the light transport vector, which is much smaller than the resolution of the background image.

Note that the data acquisition process in our algorithms becomes a simple video capturing process that can be done

(a) Ground Truth **(b)** with Gaussian Prior

Fig. 7 Composition results of the colorless bunny and dragon models

in a few minutes while other methods take at least half an hour for data acquisition. Compared with our method without Gaussian prior, the one with the prior can reduce the required image samples from 400 to about 250–300 and save at least 25 % data acquisition time while producing visually accurate composition results.

In terms of the computational cost to extract the environment matte data, the major burden in our algorithm comes from the process to recover the light transport vector $\mathcal{T}$ in each color channel for every foreground point. Table 2 gives the average time taken to recover the light transport vector using the non-negative dynamic group sparsity recovery algorithm (DGSR) with different settings for the case in Fig. 4 with 400 sample images, where the direct method solves the constrained optimization problem with high dimension for once while the hierarchical one solves the optimization problem with low dimension for twice. The algorithms are running on a HP xw4400 workstation with Intel Core2 6700 CPU and 2G memory. Note that the time values given in Table 2 are the average results obtained by running the algorithms for many times over different parts of the transparent object. It can be seen that the hierarchical method is faster than the direct one and the matte extracting time is about 1–2 h in our experiments depending on the object size, which is acceptable for practical applications. Considering that the light transportation vectors of different foreground pixels are independent, we can use parallel processing such as GPU to accelerate the processing speed.

5.4 Comparison with compressive image relighting

Since environment matting can be considered as a special case of the image-based relighting problem, in this section we give some discussions to compare our method with the state-of-the-art compressive image relighting algorithms [24, 26]. In general, all these methods use compressive sensing technique to reduce the number of sample images, and both our method and [24] use the hierarchical scheme to accelerate the measurement and recovery process.

However, there are some significant differences. First, we make use of the unique properties of the environment matting problem including the oriented Gaussian distribution and the group clustering property, which might not exist in the general image relighting problem that considers diffusion scattering effects. By utilizing these important priors, we are able to reduce the number of image samples significantly and thus accelerate both the data acquisition process and the matte data recovery process. In particular, to recover a 128×128 sparse light transport vector, [24] uses about 1,000 non-adaptive illumination samples and [26] requires to capture 1,990 pattern images, while we only need about 250–300 samples to recover a 128×96 vector $\mathcal{T}$. Second, to fight with the common issue of measurement noise, [24] utilizes the similarity among neighboring pixels while we use the Gaussian prior to suppress the effects of the noise, which is more effective since it complies with the light BRDF property. Third, the hierarchical scheme in [24] is used to explore the similarity

among neighboring pixels but the size of the measurement matrix was fixed all the time, while in our method the size of the measurement matrix in the finer level is much smaller since it is based on the results of the coarse level. Thus, the computational cost of our method is much less than that of [24].

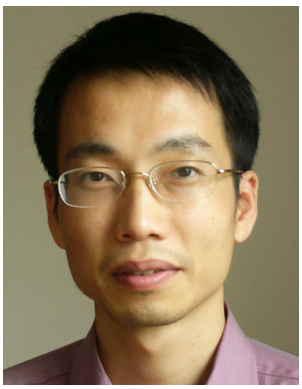
6 Conclusion

In this paper, we have proposed a framework to solve the environment matting problem through compressive sensing. Particularly, our method incorporated important properties of the light transport vector such as group clustering and Gaussian prior, with a hierarchical sampling design. The experiment results show that, compared with our method without Gaussian prior, the one with Gaussian prior can recover the reflection and refraction property of transparent objects more accurately with less number of sample images.

References

1. Baraniuk, R., Davenport, M., DeVore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. Approx. **28**(3), 253–263 (2008)
2. Candès, E.J., Romberg, J.K., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
3. Candès, E.J., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inf. Theory **52**(12), 5406–5425 (2006)
4. Chuang, Y.-Y., Zongker, D.E., Hindorff, J., Curless, B., Salesin, D., Szeliski, R.: Environment matting extensions: towards higher accuracy and real-time capture. In: SIGGRAPH '00, pp. 121–130. ACM Press, New York (2000)
5. Duan, Q., Cai, J., Zheng, J., Lin, W.: Fast environment matting extraction using compressive sensing. In: 2011 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2011)
6. Duarte, M., Davenport, M., Takhar, D., Laska, J., Sun, T., Kelly, K., Baraniuk, R.: Single-pixel imaging via compressive sampling. IEEE Signal Process. Mag. **25**(2), 83–91 (2008)
7. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. In: SIGGRAPH '00, pp. 145–156. ACM Press, New York (2000)
8. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
9. Fuchs, M., Blanz, V., Lensch, H.P., Seidel, H.-P.: Adaptive sampling of reflectance fields. ACM Trans. Gr. **26**(2), 10 (2007)
10. Goyal, V., Fletcher, A., Rangan, S.: Compressive sampling and lossy compression. IEEE Signal Process. Mag. **25**(2), 48–56 (2008)
11. Gu, J., Nayar, S., Grinspun, E., Belhumeur, P., Ramamoorthi, R.: Compressive structured light for recovering inhomogeneous participating media. In: Proceedings of the 10th European Conference on Computer Vision, ECCV 2008, pp. 845–858. Springer, Berlin (2008)
12. Garg, G.G., Talvala, E.-V., Levoy, M., Lensch, H.P.A.: Symmetric photography: Exploiting data-sparseness in reflectance fields. In: Anenine-Möller, T., Heidrich, W., (eds.) Rendering Techniques 2006: Eurographics Symposium on Rendering, pp. 251–262. Eurographics Association, Nicosia (2006)
13. Haldar, J., Hernando, D., Liang, Z.P.: Compressed-sensing MRI with random encoding. IEEE Trans. Med. Imaging **30**(4), 893–903 (2011)
14. Huang, J., Huang, X., Metaxas, D.N.: Learning with dynamic group sparsity. In: IEEE 12th International Conference on Computer Vision, ICCV 2009, pp. 64–71. Kyoto (2009)
15. Ihrke, I., Ziegler, G., Tevs, A., Theobalt, C., Magnor, M.A., Seidel, H.-P.: Eikonal rendering: efficient light transport in refractive objects. ACM Trans. Gr. **26**(3), 59 (2007)
16. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn. Reson. Med **58**(6), 1182–1195 (2007)
17. Levoy, M., Hanrahan, P.: Light field rendering. In: SIGGRAPH, pp. 31–42 (1996)
18. Matusik, W., Loper, M., Pfister, H.: Progressively-refined reflectance functions from natural illumination. In: Proceedings of the 15th Eurographics workshop on rendering, Eurographics Association, pp. 299–308 (2004)
19. Masselus, V., Peers, P., Dutré, P., Willems, Y.D.: Relighting with 4D incident light fields. In: SIGGRAPH '03, pp. 613–620. ACM, New York (2003)
20. Matusik, W., Pfister, H., Ziegler, R., Ngan, A., McMillan, L.: Acquisition and rendering of transparent and refractive objects. In: Proceedings of the 13th Eurographics workshop on rendering, Eurographics Association, pp. 267–278 (2002)
21. Peers, P., Dutré, P.: Wavelet environment matting. In: Proceedings of the 14th Eurographics workshop on rendering, Eurographics Association, pp. 157–166 (2003)
22. Peers, P., Dutré, P.: Inferring reflectance functions from wavelet noise. In: Proceedings of the 16th Eurographics workshop on rendering, Eurographics Association, pp. 173–182 (2005)
23. Peyre, G.: Best basis compressed sensing. IEEE Trans. Signal Process. **58**(5), 2613–2622 (2010)
24. Peers, P., Mahajan, D., Lamond, B., Ghosh, A., Matusik, W., Ramamoorthi, R., Debevec, P.E.: Compressive light transport sensing. ACM Trans. Gr. **28**(1), 1355–1363 (2009)
25. Sen, P., Chen, B., Garg, G., Marschner, S.R., Horowitz, M., Levoy, M., Lensch, H.P.A.: Dual photography. ACM Trans. Gr. **24**(3), 745–755 (2005)
26. Sen, P., Darabi, S.: Compressive dual photography. Comput. Gr. Forum **28**(2), 609–618 (2009)
27. Sen, P., Darabi, S.: Compressive estimation for signal integration in rendering. Comput. Gr. Forum **29**(4), 1355–1363 (2010)
28. Sen, P., Darabi, S.: Compressive rendering: a rendering application of compressed sensing. IEEE Trans. Vis. Comput. Gr. **17**(4), 487–499 (2011)
29. Wang, J., Dong, Y., Tong, X., Lin, Z., Guo, B.: Kernel nyström method for light transport. In: SIGGRAPH '09: ACM SIGGRAPH 2009 papers, , pp. 1–10. ACM, New York (2009)
30. Wexler, Y., Fitzgibbon, A.W., Zisserman, A.: Image-based environment matting. In: Rendering Techniques, pp. 279–290 (2002)
31. Wakin, M., Laska, J., Duarte, M., Baron, D., Sarvotham, S., Takhar, D., Kelly, K., Baraniuk, R.: An architecture for compressive imaging. In: 2006 IEEE International Conference on Image Processing, pp. 1273–1276 (2006)
32. Zongker, D.E., Werner, D.M., Curless, B., Salesin, D.: Environment matting and compositing. In: SIGGRAPH, pp. 205–214 (1999)
33. Zhu, J., Yang, Y.-H.: Frequency-based environment matting. In: Pacific Conference on Computer Graphics and Applications, pp. 402–410. IEEE Computer Society (2004)
34. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Towards robust and effective shape modeling: Sparse shape composition. Med Image Anal **16**(1), 265–277 (2012)

Qi Duan received his Ph.D. degree in Computer Engineering from Nanyang Technological University, Singapore, under the supervision of Prof. Jianfei Cai and Prof. Jianmin Zheng. His major research focus is on Computer Graphics and Image Processing. Before that, he got his Master degree from Shanghai Jiao Tong University in 2008 and his Bachelor degree from National University of Defense Technology in 2004, both in Computer Science. During July 2010 to August 2010, he was a visiting student in Massachusetts Institute of Technology. Currently, he is the manager of the visualization group in Shanghai United-Imaging, focusing on Medical Imaging Visualization and related researches.



Jianfei Cai received his Ph.D. degree from the University of Missouri-Columbia. He is currently an Associate Professor and has served as the Head of Visual & Interactive Computing Division and the Head of Computer Communication Division at the School of Computer Engineering, Nanyang Technological University, Singapore. His major research interests include visual computing and multimedia networking. He has published more than 140 technical papers in international conferences and journals. He has been actively participating in program committees of various conferences. He has served as the leading Technical Program Chair for IEEE International Conference on Multimedia & Expo (ICME) 2012 and the leading General Chair for Pacific-rim Conference on Multimedia (PCM) 2012. He was an invited speaker for the first IEEE Signal Processing Society Summer School on 3D and high definition / high contrast video process systems in 2011. He is currently an Associate Editor for IEEE Trans on Image Processing (T-IP) and has served as an Associate Editor for IEEE Trans on Circuits and Systems for Video Technology (T-CSVT) from 2006 to 2013. He is also a senior member of IEEE.

Jianmin Zheng is an Associate Professor in the School of Computer Engineering at Nanyang Technological University, Singapore. He received the BS and Ph.D. degrees from Zhejiang University, China. His recent research focuses on T-spline technologies, digital geometric processing, interactive digital media and applications. He has published more than 100 technical papers in international conferences and journals. He was the conference co-chair of Geometric Modeling and Processing 2014 and has served on the program committee of several international conferences.