Natural and Seamless Image Composition with Color Control

Wenxian Yang, Jianmin Zheng, Jianfei Cai, Susanto Rahardja, Chang Wen Chen

Abstract—While the state-of-the-art image composition algorithms subtly handle the object boundary to achieve seamless image copy-and-paste, it is observed that they are unable to preserve the color fidelity of the source object, often require quite an amount of user interactions, and often fail to achieve realism when there exists salient discrepancy between the background textures in the source and destination images. These observations motivate our research towards color controlled natural and seamless image composition with least user interactions.

In particular, based on the Poisson image editing framework, we first propose a variational model that considers both the gradient constraint and the color fidelity. The proposed model allows users to control the coloring effect caused by gradient domain fusion. Second, to have less user interactions, we propose a distance-enhanced random walks algorithm, through which we avoid the necessity of accurate image segmentation while still able to highlight the foreground object. Third, we propose a multiresolution framework to perform image compositions at different subbands so as to separate the texture and color components to simultaneously achieve smooth texture transition and desired color control. The experimental results demonstrate that our proposed framework achieves better and more realistic results for images with salient background color or texture differences, while providing comparable results as the state-of-the-art algorithms for images without the need of preserving the object color fidelity and without significant background texture discrepancy.

Index Terms—Digital image editing, image composition, image copy-and-paste, Poisson image editing.

I. INTRODUCTION

Image composition is a basic process in digital image editing. Its objective is to enable convenient image object copy-and-paste to generate new images which look natural and realistic. In general, there are two classes of image composition: image cloning and image blending. Image cloning deals with placing opaque image components one over another while image blending or mixing merges semitransparent image components together [1]. We focus on the first type of image composition.

A. Related Work

Recent research has been mainly focused on two aspects of image composition: seamless composition and least user interactions. The Poisson image editing scheme [2] is the most representative framework for seamless image composition. It adopts the gradient domain fusion technique to hide seams between the composed regions through converting high-frequency artifacts that may appear at the boundaries into low-frequency variations that spread across the image.

The Poisson image editing framework ensures only color continuity at the boundary. In comparison, the Photoshop healing brush [3], [4] achieves seamless composition by constructing iterative solutions to a fourth-order partial differential equation (PDE), and therefore the resulting image is continuous not only in color but also in boundary derivatives. Because the solution of high order PDEs is computationally expensive, the healing brush is typically used for removing small patches of defects in images instead of object composition from one image to another.

Another disadvantage of the Poisson image editing framework is that the user has to carefully outline the boundary of the region-of-interest (ROI) to avoid any salient structure confliction in the foreground and the background. In order to relief the user from carefully outlining the image object, the digital photomontage algorithm [5] adopts a graph cut based method [6] to choose good seams to avoid confliction among multiple constituent images. Their method requires user-drawn strokes on each image for seam selection before the gradient domain fusion. When applying the digital photomontage framework to the copy-and-paste image composition between a source image and a destination image, computing a good seam is equivalent to segmenting the object-of-interest (OOI) out from the source image.

The drag-and-drop pasting method [7] also improves the Poisson image editing framework by finding an optimal boundary. In this method, the OOI is first segmented from the ROI using the grab cut image segmentation algorithm [8], and then the optimal boundary is searched within the ROI but outside the OOI. However, without normalizing the energy function by the length of the boundary, the method for finding the optimal boundary opts for shorter ones and thus the advantage of bringing in this optimal boundary is limited.

Recently, the Photo Clip Art work [9] extends the drag-and-drop approach by utilizing some domain / prior knowledge. As the drag-and-drop approach, Photo Clip Art runs in three steps: first object extraction using the grab cut algorithm with shape prior, then blending mask computation and finally composition with the Poisson image editing algorithm. Noticing that the Poisson image editing framework cannot preserve the color of the pasted foreground object, Photo Clip Art does not perform cloning when the mask coincides with the object while cloning in places where the mask is outside the object. In this way, it

Wenxian Yang, Jianmin Zheng and Jianfei Cai are with School of Computer Engineering, Nanyang Technological University, e-mail: {wxyang, asjmzheng, asjfcai}@ntu.edu.sg. Susanto Rahardja is with Institute for Inforcomm Research, e-mail: rsusanto@i2r.a-star.edu.sg. Chang Wen Chen is with Department of Computer Science and Engineering, University at Buffalo, e-mail:chencw@cse.buffalo.edu. Contact author: Jianfei Cai. This research was partially supported by Singapore A*STAR SERC Grant (062 130 0059).

guarantees the preservation of the object's color.

There are some other schemes being proposed to improve, extend or generalize the Poisson image editing framework. For example, in [10], the alpha interpolation is introduced to remove the abrupt edges in the resulting image caused by mixed seamless cloning. In addition, the luminance re-scaling is used to scale the influence of the source image on the result.

Interactive image matting algorithms [11][12][13] can also be used for image composition. Image matting algorithms extract the foreground object with a matte which has fractional values near the object boundaries. When pasting the object to another image, the color values are generated through a weighted average of the foreground object and the new background according to the matte, therefore leading to a natural transition from the background color to the foreground color. Matting algorithms are suitable for editing flurry objects. However, they only handle the color mixing problem and cannot generate realistic and natural images when the source and the destination images have different textures.

B. Motivation

Although the above discussed algorithms [2] [5] [7] [9] [13] have achieved excellent performance in image composition, there still exist severe limitations in the following aspects.

Color control: As pointed out in [7] [9], when the source and destination images differ greatly in color, the Poisson image editing framework changes the color of the pasted foreground object significantly and globally, which is not desired in many situations. One example can be found in Figure 1(c). Although the Photo Clip Art work guarantees the preservation of the object color, it fails in the cases of favoring color change since it does not perform cloning on OOI at all. Therefore, it is needed to find a systematic way to control the coloring effects.

Natural and seamless compositions: In terms of color seamless image composition, the Photomontage approach performs very well. This is because in Photomontage the OOI is segmented out and pasted onto the destination image. In this way, as long as a clean OOI boundary is obtained, the resulting image is perceptually seamless (see Figure 5(d)). However, the resulting image in Figure 5(d) does not look natural / realistic (How can a dog look static in a flowing river?). The main reason is that Photomontage completely removes the background or neighborhood of the object in the source image. The same problem exists for matting algorithms.

What we advocate in this paper is *natural* composition of images. The imaging of an object is not isolated but interrelated with its located environment / scene because of the relative position, the material properties, and more commonly the lighting conditions of the scene being captured. Therefore, for natural composition of images, the neighborhood of the object in the source image should not be simply removed or replaced. For example, the reflection and the ripples in Figure 2(a) are closely related to the object - the swan.

On the other hand, although the Poisson image editing framework does not remove the object background, it fails to achieve seamless results when there exists salient discrepancy between the background textures in the source and destination images. For example, as shown in the result of the Poisson image editing in Figure 2(c), although the color of the pasted region is consistent with that of the new background, the seam is still quite obvious due to the salient differences in texture.

Note that by extending the foreground strokes to the object neighborhood, Photomontage can include part of the object neighborhood (see Figure 2(d)). However, intensive foreground and background strokes are required to be drawn alongside with the desired boundary.

Complexity: All the three state-of-the-art image composition algorithms [5] [7] [9] require to explicitly segment the OOI out using the interactive image segmentation algorithms such as the graph cut [14] and the grab cut [8]. These benchmark interactive image segmentation algorithms typically require the user to define a tri-map in either an explicit (e.g. using foreground and background strokes) or an implicit (e.g. using a lasso input in [7]) way, which separates the image pixels into three sets: foreground seeds, background seeds and unknown pixels. In the cases with low contrast edges or noisy images, it is necessary to carefully paint the initial foreground and background strokes and even perform local editing iteratively to obtain satisfactory results. Such meticulous operation is not welcomed for our image composition task. In addition, in the cases that there do not exist salient edges between foreground and background, it is hard to obtain a good segmentation. One example can be found in Figure 2(a) where the swan, its reflection, and the ripples can hardly be separated.

C. Our Work

In this paper, we provide solutions for the aforementioned limitations of the Poisson image editing framework and we target at natural and realistic image composition with least user interactions. In particular, we first in Section 2 propose a variational model that considers both the gradient constraint and the color fidelity. The proposed model allows users to control the coloring effect caused by gradient domain fusion. Second, instead of explicitly segmenting the foreground object from its background, in Section 3 we propose a distance-enhanced random walks algorithm to generate a weight image that implicitly conveys the image segmentation information and measures the significance of the background. Moreover, in Section 4 we propose a multiresolution framework where we perform subband compositions at different boundaries to enable a more smooth and natural transition between the source and destination images. The experimental results discussed in Section 5 demonstrate that our proposed framework achieves better and more realistic results for images with salient background color or texture differences, while providing comparable results as the state-of-the-art algorithms for images without the need of preserving the OOI color fidelity and without significant background texture discrepancy. Finally, we conclude this paper in Section 6.

II. PROPOSED VARIATIONAL MODEL

Let f_1 and f_2 denote the source and the destination images, respectively. Let f denote the resulting image. Ω denotes the
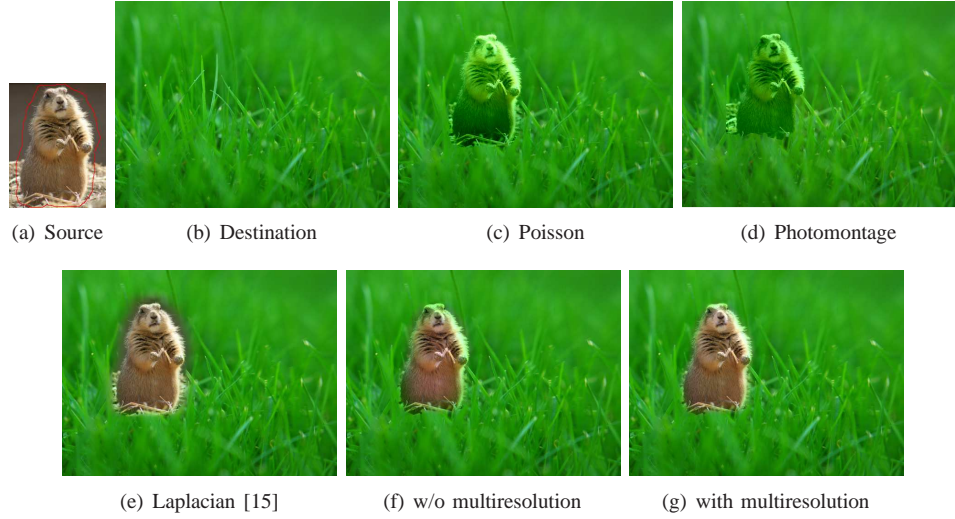
Fig. 1.   The image composition results for *squirrel*: (a) source image with the user inputted lasso in red, (b) destination image, (c) result by Poisson image editing, (d) result by digital Photomontage, (e) result by the classical Laplacian Pyramid blending [15], (f) result by our approach without multiresolution with $\lambda = 0.005$, and (g) result by our approach with multiresolution. The images are best viewed in color.
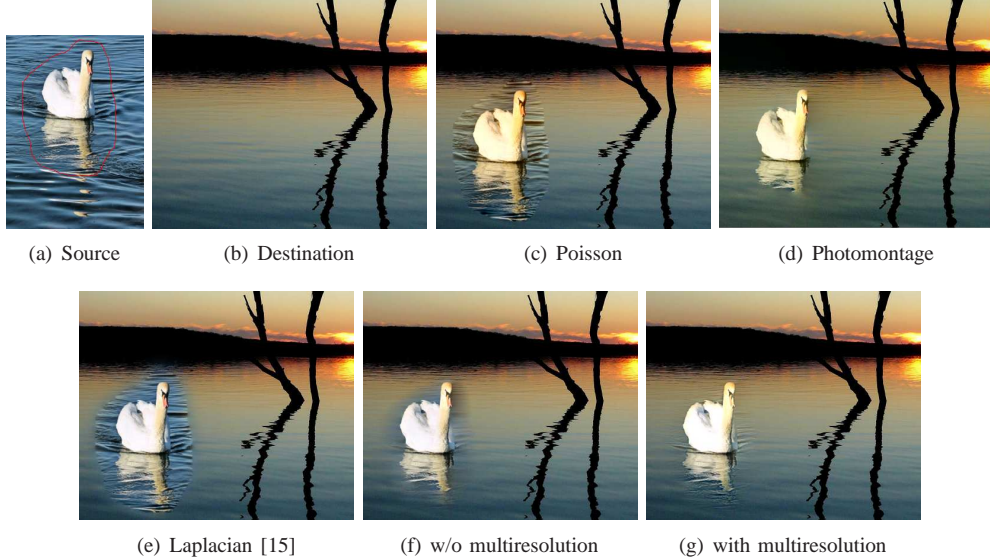

Fig. 2.   The image composition results for *swan*: (a) source image with the user inputted lasso in red, (b) destination image, (c) result by Poisson image editing, (d) result by digital Photomontage, (e) result by the classical Laplacian Pyramid blending [15], (f) result by our approach without multiresolution with $\lambda = 0.005$, and (g) result by our approach with multiresolution. The images are best viewed in color.

TABLE I
DENOTATIONS

Source image	f_1
Destination image	f_2
Resulting image	f
Weight image	w
ROI	Ω
OOI	Ω_o
Exterior boundary of ROI	$\partial\Omega$

user inputted ROI, and $\partial\Omega$ represents its exterior boundary. Ω_o denotes the OOI. The denotations are summarized in Table I.

We formulate the image composition problem as the mini-

mization of energy $E(f)$, where

$$E(f) = \iint_\Omega ||\nabla f - g_v||^2 d\Omega + \lambda \iint_{\Omega_o} ||f - f_1||^2 d\Omega, \quad (1)$$
$$\text{s.t.} \quad f|_{\partial\Omega} = f_2|_{\partial\Omega}.$$

Clearly, the proposed energy function in Eq. (1) combines the gradient term and the color fidelity term through a tradeoff parameter λ. In the gradient term, unlike the classical Poisson image editing, where the guidance vector field g_v is typically the gradient of the source image, we define g_v as the weighted combination of the gradients of the source and destination images, i.e.

$$g_v = w \cdot \nabla f_1 + (1 - w) \cdot \nabla f_2, \quad (2)$$

where w is a weight function and $w \in [0, 1]$. The idea here is to allow a smooth transition in the ROI-but-OOI region

$(\Omega \setminus \Omega_o)$. The weight w should be zero for the region outside the ROI, one for the OOI region and increasing from zero to one from the ROI boundary to the OOI boundary. Note that although the proposed weight function w looks similar to the matte used in image matting from the aspect of its value range, the definition, generation and usage of w are distinct from the matte. We will discuss how to generate w in Section III. For color images, the image composition is performed separately for each of the three color channels in RGB color space.

As aforementioned, the classical Poisson image editing algorithm, which copies the gradient of the source image into the guidance vector field with Dirichlet boundary conditions, may lead to significant global coloring of the object towards the background. Thus, our proposed framework introduces the additional color fidelity term to preserve the color fidelity of the OOI. Figure 1(f) shows the effect of the color fidelity term. It can be seen in Figure 1(c) that the object color is affected by the background color globally for the result generated by the Poisson image editing algorithm. Instead, our proposed algorithm keeps the color fidelity with regard to the input source image (Figure 1(f)).

To our understanding, whether the coloring effect is welcome or not actually depends on high level image semantics. More specifically, it depends on the lighting conditions and the environment when the image content was captured. For example, when pasting an object onto a sun-set image where everything turns rosy (see Figure 5(b)), it is desired that the pasted object be consistent with the specific lighting condition. On the contrary, the coloring effect in Figure 1(c) is definitely not desired. Our proposed framework allows users to control the coloring effect through adjusting the tradeoff parameter λ. When $\lambda = 0$, Eq. (1) is the same as the composition equation in Poisson image editing. In this case, the object pasting is seamless but the color of the object is likely to be changed. The other extreme case is that, when λ is very large, the gradient term only affects the ROI-but-OOI region. In this case, the OOI is unchanged from the source image, but the seams between the OOI, the ROI-but-OOI region, and the ROI can be very obvious. The in-between λ values create a gradual change between these two extreme cases.

III. Distance-enhanced Random Walks

One question we need to answer now is how to generate the weight w to compute the guidance vector field shown in Eq. (2), bearing in mind that our goal is to preserve the salient features of the OOI while introducing a smooth and gradual transition in the ROI-but-OOI region $(\Omega \setminus \Omega_o)$.

One intuitive solution is to use the distance information to determine the weight. This idea of distance based weighting has been used in image stitching [15][16]. However, using the distance information to determine the weight implicitly assumes that the pixels in the user inputted ROI boundary are approximately of the same distance away from the OOI boundary. This is usually not true for a casual user input. Figure 3(a) shows the generated weighting image based on only the distance information [17], which fails to highlight the OOI.
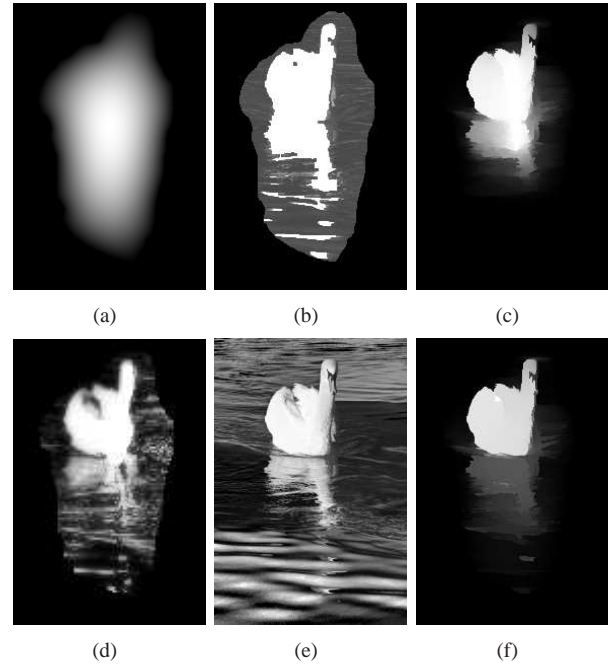

Fig. 3. The weight image for *swan* generated by (a) shape-representing random walks, (b) grab cut image segmentation, (c) random walks for image segmentation, (d) robust matting, (e) spectral matting, and (f) proposed distance-enhanced random walks algorithm. The weights are scaled to $[0, 255]$ for better display.

Another intuitive solution is to use the probability of a pixel belonging to the OOI as the weight. The probability can be derived according to the foreground and background color models of the source image f_1. A common way to compute the probability is to apply the graph cut based image segmentation algorithms, which have proven their strength in interactive image segmentation and have been adopted to generate a good initialization for image stitching [18] and composition [7]. However, the problem with this type of color model based methods is that the weight is solely determined by the color distribution without considering the distance information. Specifically, the graph cut based approaches could generate disconnected object segments and the resulting probability map could contain uniform low values in the ROI-but-OOI region. Figure 3(b) shows the generated weighting image using the grab cut image segmentation algorithm [8], which fails to generate a smooth and gradual transition in the ROI-but-OOI region.

More recently, the random walks algorithm has been adopted for various image processing tasks. It has been demonstrated in [19] that the random walks algorithm can achieve better image segmentation performance than the graph cut algorithm. The random walks algorithm models an image as a graph. By assigning different meanings to graph nodes and defining different relationships among neighboring nodes, the values generated by the random walks algorithm can represent either the probabilities that individual pixels belong to the foreground [19], or the distance of individual pixels to the background [17]. Note that the existing random walks models only consider either the color distribution such as in [19] or the geometric distance information such as in [17].

A. Generating the Weight

From the above discussion, it is clear that the weight should be generated based on both color distribution and distance information. Thus, in this paper, we propose a distance-enhanced random walks scheme, where we employ two ways to embed the distance information in the original random walks framework [19]. Our algorithm only needs to process the ROI region in the source image f_1 once to generate the weight image.

In particular, similar to the original framework, we consider an image as an undirected weighted graph with each pixel being a node. Let p_i represent both an image pixel and the corresponding node in the graph, and let e_{ij} represent both the edge connecting p_i with its neighbor p_j and the corresponding edge weight. The edge weight e_{ij} represents the likelihood that a random walker starting from p_i takes a next step towards p_j. First, we define the edge weight as

$$e_{ij} = \exp(-\beta||g_i - g_j||^2 \cdot ||h_i - h_j||) \qquad (3)$$

where g_i represents a pixel value for gray-scale images or a RGB triplet for color images, h_i denotes the coordinate of pixel p_i, and β is a free parameter (by default $\beta = 300$). Compared with the original random walks framework [19], we use the 2nd-order neighborhood system and the edge weight is enhanced by the distance between the pixels (i.e. $||h_i - h_j||$).

Second, we calculate the time taken for a random walker starting from pixel p_i to hit the region boundary $\partial\Omega$, which is denoted as x_i. The time x_i can be considered as the time for a random walker to reach one neighbor of p_i, say, p_j, plus the time taken from p_j to hit the region boundary, which is x_j. Averaging over all the neighbors, we obtain

$$x_i = \frac{\sum_j e_{ij} x_j}{\sum_j e_{ij}} + 1, \qquad (4)$$

where the time for a random walker starting from p_i to reach one of its neighbors is considered as a constant equal to 1. Obviously, for a pixel q lying on the boundary $\partial\Omega$, $x_q = 0$. Our proposed distance-enhanced random walks algorithm is actually to solve a system of linear equations, which contains the equations derived from Eq. (4) for all the pixels inside the ROI region and the equations for the boundary conditions. After obtaining all the values of x_i, we normalize them to be within $[0, 1]$ and the normalized values are used as the weight w_i. Figure 3(f) shows the generated weight image by our proposed algorithm.

B. Determining the OOI

Another question we need to answer is how to determine the OOI region (Ω_o). One straightforward approach is to apply one of the state-of-the-art interactive image segmentation algorithms such as the grab cut [8] or the random walks [19]. However, as mentioned in Section I-B, in order to achieve very good segmentation results, these algorithms often require extensive local editing, which is not desired for our image composition task. In addition, in the cases that there do not exist salient edges between foreground and background, it is hard to perform a good segmentation.

Different from the interactive image segmentation algorithms, our image composition task here might not need to accurately segment the OOI out. What we need is a good boundary that can closely cover the object-of-interest. Since the weight image generated by the proposed distance-enhanced random walks algorithm has already highlighted the OOI, i.e. the weight image implicitly defines the foreground object, for the sake of maintaining low complexity, we directly determine the OOI by thresholding the weight image. In particular, the OOI region Ω_o is a binary map defined as

$$\Omega_o(i) = \begin{cases} 1, & \text{if } w(i) > T_o \\ 0, & otherwise \end{cases} \qquad (5)$$

where T_o is a threshold and $w(i)$ is the weight for pixel i. To our understanding, the determination of OOI itself is a very subjective process. For the example of Figure 2(a), some people might choose the swan as the OOI while others might also want to include the swan's reflection into the OOI. Therefore, we allow user interactions to determine the OOI by choosing an appropriate threshold T_o.

Compared with the conventional random walks algorithm, our proposed algorithm does not require foreground and background strokes. The only input is the lasso, which can be casually drawn. Although the proposed algorithm requires user interactions to set the threshold T_o, our empirical studies show that $T_o \leq 0.5$ and usually only a few values need to be tried to find a good OOI boundary.

C. Results of the Weight Image

The effectiveness of the proposed distance-enhanced random walks algorithm is shown in Figure 3. We first compare our generated weight image with those obtained by the shape-representing random walks algorithm [17], the grab cut algorithm [8] and the random walks algorithm [19]. The weights are scaled to $[0, 255]$ for better display. It can be seen that the weight image in Figure 3(b) obtained by the grab cut shows abrupt changes from the background to the ROI, and uniform values inside the ROI-but-OOI region. On the other hand, the weight image in Figure 3(f) generated by the proposed algorithm shows good properties considering both color and distance information with regard to the user's input. Specifically, the foreground object is highlighted due to its color properties and the weights in the ROI-but-OOI region are gradually attenuated as approaching the ROI boundary.

Note that the grab cut algorithm only outputs a binary map. The result in Figure 3(b) is generated from the parameters of the two Gaussian Mixture Models (GMM), which are used to model the foreground and the background, respectively. Particularly, the pixels connected to the foreground node are assigned the weight 1 and those connected to the background node are assigned the weight $\frac{p_f}{p_f+p_b}$, where p_f and p_b denote the fitness or probability to the foreground GMM and the background GMM, respectively. In addition, an additional foreground stroke is drawn to obtain the result in Figure 3(c).

We also compare our weight image with the alpha mattes obtained by the two state-of-the-art image matting algorithms: the spectral matting [11] and the robust matting [12]. As

shown in Figure 3(d), the matte obtained by the robust matting algorithm shows a clear seam around the ROI. Although the matte in Figure 3(e) generated by the spectral matting [11] is much better, it does not provide a smooth transition in the ROI-but-OOI region. In addition, compared with the random walk algorithm, the spectral matting method has higher computation complexity due to the necessity of computing eigenvectors.

IV. A MULTIRESOLUTION FRAMEWORK

A. Why multiresolution

Figures 1(f), 2(f) and 4(g) show the image composition results using our proposed variational model. It can be seen that although the proposed model is able to better preserve the color fidelity of the OOI, there exist some unwanted color leaking effects, e.g., the boundary of the squirrel turns green and the neighborhoods of the swan and the dolphin are inconsistent with their environment in color. No matter how we adjust the λ in Eq. (1), it is hard to simultaneously preserve the OOI color fidelity and eliminate the color leaking. Specifically, in order to move towards natural and seamless composition, what we want is to preserve the OOI color fidelity while in the ROI-but-OOI region having smooth texture transition and consistent color with the destination background. Only using the proposed variational model is insufficient to achieve these multiple goals in many situations.

Motivated by the field of texture analysis and synthesis [20], where image texture is usually described by high frequency components while the low frequency component is deemed to convey the average color information, we introduce a multiresolution image composition framework to tackle the limitation of the proposed variational model. Our basic idea is to separate the texture and color components through multiresolution decomposition and only add the color fidelity term into the composition of the color component.

It has long been discovered [15] that better transition can be achieved in image stitching by using a multiresolution representation. The idea is to divide images into bandpass signals and stitch subbands at different levels separately. In this way, the low frequency components are composited over a wide range while the high frequency components are composited over a narrow range. However, in this paper, multiresolution representation is adopted for a totally different purpose.

B. Proposed Multiresolution Image Composition

It is not a trivial task to extend the Poisson image editing framework from single layer to multiple layers. Directly applying the Poisson image editing at different resolution levels with aligned boundaries does not perform better than the original single-layer approach.

Our approach is to divide the ROI-but-OOI region into several embedded subregions, where the smallest subregion is the OOI and the largest subregion is the ROI. Different subregion boundaries are used as the Poisson editing boundary at different subbands. The subregion for the lowest frequency subband only covers the OOI while the subregions for the high frequency subbands increase with the frequency. The color fidelity term is only applied for the composition in

the lowest frequency subband. In this way, the color fidelity of the OOI is controlled through the composition of the lowest frequency subband with no average color information of the source image background being introduced. On the other hand, with no color fidelity term in the compositions of the high frequency subbands, the Poisson image editing framework can well preserve the gradients of the texture of the OOI and its neighborhood. In addition, by using larger subregions for higher frequency subbands, we gradually blend in the frequency components of the source image until all the frequency components of the source image emerge in the OOI region.

In particular, the Laplacian pyramid is adopted to decompose an image into different frequency subbands. For a decomposition level of L (by default $L = 2$), there are $L + 1$ subbands in the Laplacian pyramid denoted as $f(l)$, $l = 0, \cdots, L$, where l is the subband index and $f(0)$ is the lowest frequency subband. For the sake of low complexity, the weight image w is reused to obtain the subregion boundaries since w approximately represents the content distance to the OOI. Specifically, the ROI-but-OOI region is divided in a uniform manner, i.e. calculating the subregion $\Omega[l]$ as

$$\Omega[l](i) = \begin{cases} 1, & \text{if } w(i) > T_o \cdot (L - l)/L \\ 0, & otherwise \end{cases} \quad (6)$$

where T_o is the same as that in Eq. (5).

For subband l, let w_l denote its weight image defined as

$$w_l(i) = \begin{cases} 1, & \text{if } i \in \Omega_o \\ w(i), & i \in \Omega[l] \setminus \Omega_o \\ 0, & i \notin \Omega[l] \end{cases} \quad (7)$$

Note that $w_l(i)$ is re-scaled to $[0, 1]$ in the transition zone $\Omega[l] \setminus \Omega_o$ to allow the smooth transition facilitated by Eq. (2). For subband l, let λ_l denote the tradeoff parameter used in Eq. (1). We set $\lambda_l = 0$ for all the subbands except the lowest frequency subband so as to control the color fidelity of the OOI without introducing the average color information of the source image background. Finally, we process subband l in the same way as processing the original image but with the parameters of $(f_1(l), f_2(l), \Omega[l], w_l, \lambda_l)$. Note that all the subbands $f(l)$ are upsampled into the original image size for composition.

The variational model of Eq. (1) in the discrete form can be written as a system of linear equations, which is solved by the sparse direct linear solver efficiently implemented in the open-source library TAUCS (http://www.tau.ac.il/~stoledo/taucs/).

Figures 1(g), 2(g) and 4(h) show the image composition results using our proposed multiresolution framework. Compared with the one without multiresolution, the multiresolution approach eliminates the color leaking effects, e.g. the neighborhood of the dolphin in Figure 4(h) is consistent with the environment in color, while still well preserving the color fidelity of the foreground object. Figure 4(c) shows the generated subregions for the 3-level decomposition of the image *dolphin*.

V. EXPERIMENTAL RESULTS

We compare the proposed method with the Poisson image editing algorithm [2], the digital Photomontage method [5],

(a) Source (b) Weight image (c) Subregions

(d) Poisson (e) Photomontage (f) Laplacian [15] (g) w/o multiresolution (h) with multiresolution
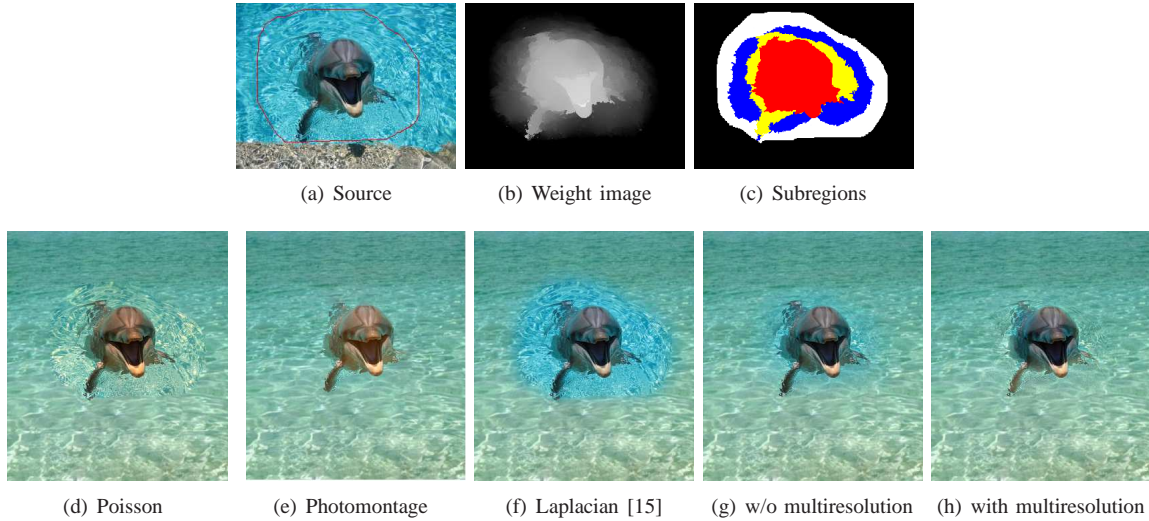
Fig. 4. The image composition results for *dolphin*: (a) source image with lasso, (b) weight image, (c) subregions for different frequency components (The subregions, from the highest frequency subband to the lowest frequency subband, are marked in white, blue, yellow and red, respectively.), (d) result by Poisson image editing, (e) result by digital Photomontage, (f) result by the classical Laplacian Pyramid blending [15], (g) result by our approach without multiresolution, and (h) result by our approach with multiresolution. The images are best viewed in color.

and the classical Laplacian Pyramid blending [15]. Note that in order to obtain satisfactory results for the digital photomontage algorithm, the foreground and background strokes have to be carefully drawn. All the results generated by the classical Laplacian Pyramid blending method use 3-level decompositions.

We first demonstrate the ability of our proposed approach in color control. As shown in Figures 1, 2, 4 and the *dog* example in Figure 5, all the three existing composition algorithms fail to preserve the color of the OOI but our approach can do it.

We then test the performance of seamless composition with different textures. In particular, we consider three different scenarios: from texture to another texture (e.g. *dog* in Figure 5 and *dolphin* in Figure 4), from texture source to smooth destination (e.g. *swan* in Figure 2), and from smooth to texture (e.g. *boat* in Figure 5). As shown in Figures 2(c), 4(d), and 5(i), a clear seam can be seen for the results of the Poisson image editing. This is because of the significant texture differences in the source and destination images. Although the results obtained by the digital photomontage do not have obvious seams, they look unnatural in Figures 4(e) due to the lost of the source image background and in Figures 2(d) due to imperfect segmentation. Moreover, we notice that the digital photomontage slightly changes the color of the destination, which is often undesirable. The classical Laplacian Pyramid blending produces a gradual transition along the ROI boundary with a bit blurring, but fails to provide seamless composition either in color or in texture. On the contrary, our proposed framework avoids these problems and the obtained results look smooth and natural.

Another test is on the image of *pyramid* in Figure 5(a). In this case, the source and destination images have similar texture background and changing the color of the OOI is welcomed. We can see that the result obtained by our proposed algorithm is comparable to that of Photomontage.

We did not preform comparisons with the other two state-of-

the-art approaches: drag-and-drop [7] and Photo Clip Art [9], because there is no publicly available implementation for them. It is also hard to reproduce these two approaches due to the complexity of the techniques. However, it can be sure that the two approaches cannot fully solve the three limitations mentioned in Section I-B. In particular, it is clearly stated in [7] that the drag-and-drop approach generally does not maintain the color fidelity of the OOI and it does not perform well when there exists salient discrepancy between the background textures in the source and destination images. On the other hand, although the Photo Clip Art work guarantees the preservation of the OOI color fidelity, it fails in the cases of favoring color change since it does not perform cloning on OOI at all. In other words, Photo Clip Art does not provide a systematic way in color control. In addition, Photo Clip Art requires that the source and destination images are of high similarity at the region where the blending performs, which limits its applications.

The distance-enhanced random walks algorithm plays an important role for the overall performance of the proposed framework. As this algorithm is based on the random walks algorithm, it does not work well in camouflage images or for objects with thin structure. It is possible for these difficult images that the proposed method generates a weight image where the background has higher weight than the foreground, and therefore some background regions could be considered as part of the OOI while part of the true foreground could be considered as background. In this case, the composition may fail to produce a good resulting image. Figure 6 gives a failure case of the proposed framework, where the background pixels around the horn have high weighting values and the resulting image is not satisfactory. To handle this type of images, more user interaction or even an interactive segmentation procedure to outline the OOI is desirable.

The computational complexity of the system lies in solving sparse linear equations. For a system with L level decompo-
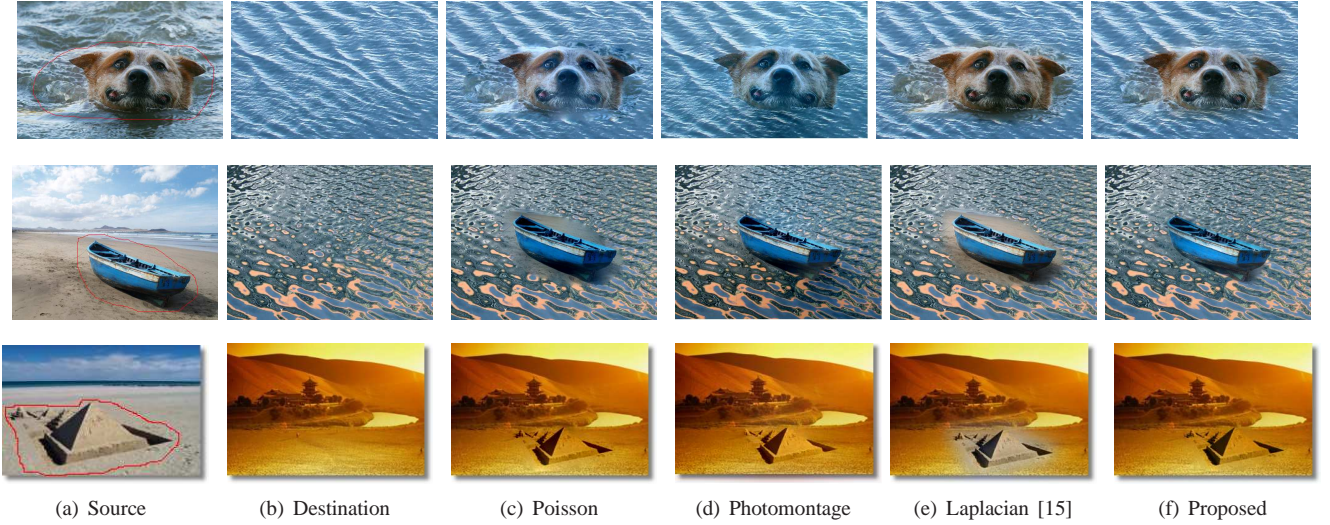
Fig. 5. The image composition results for the images of *dog*, *boat* and *pyramid*. The leftmost column shows the source image with a user-specified lasso. The second column shows the destination image. The third, fourth and fifth columns are the results obtained by the Poisson image editing, the digital photomontage, and the classical Laplacian Pyramid blending [15], respectively. The rightmost column shows our results. The images are best viewed in color.
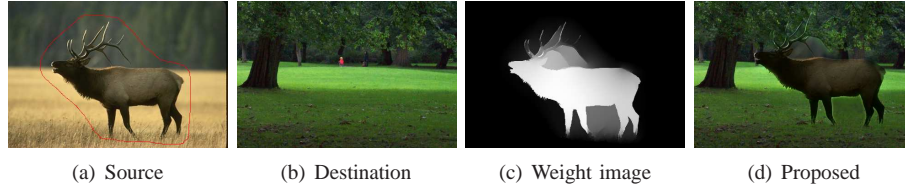

Fig. 6. A failure case. (a) and (b) show the source and destination images, (c) shows the weight image by the distance-enhanced random walks algorithm, and (d) shows the final composition result.

sition, there are one set of sparse linear equations to solve for calculating the weights from distance-enhanced random walks, and $L + 1$ sets of sparse linear equations to solve for $L + 1$ subbands in the multiresolution composition. The dimension of the sparse matrix equals to the number of pixels in the ROI, and that divided by 2^l for level l. Experiments are conducted on a PC with Intel 2.67GHz CPU with 2G RAM. Our experience is that it typically takes about 5 seconds to process the proposed image composition with $L = 2$ for VGA size images. Note that our implementation is for research purpose and is thus without any code optimization. For industrial use, sparse linear systems can be handled more efficiently as parallelized to be processed by GPU [19].

VI. CONCLUSIONS

In this research, we consider natural compositions of images, where we introduce the object background in the source image into the resulting image since the background of the source image is often interrelated with its foreground object. We have proposed three ingredients, the variational model, the distance-enhanced random walks algorithm and the multiresolution framework, to solve the important problems of the Poisson image editing framework. In particular, first, a color fidelity term is introduced in the variational model to control the color of the OOI. In addition, a mixed gradient is used as the guidance vector in the variational model to allow smooth transition in the transition zone. Second, the distance-enhanced random walks algorithm is proposed to avoid the necessity of

accurate image segmentation while still able to highlight the OOI. It also provides the weights for generating the mixed gradient. Third, the multiresolution framework is proposed to separate the blending of the color and texture components so that smooth texture transition and desired color control can be simultaneously achieved. The experimental results show that our method is especially preeminent in the cases where the preservation of the color fidelity of the OOI is required or there exists salient texture difference between the source and destination backgrounds.

There is still a long way to achieve the photorealistic image composition. The images of different objects in a real world scene are interrelated in a complex way. There are many challenges that neither our current framework nor any other existing image editing system can handle, to the best of our knowledge. For example, an object may be affected by both self-shadow and cast-shadow. While the former has already been drawn attention in the research community [21], the cast-shadow is much more difficult to detect and remove. Furthermore, when pasting an object onto a background image taken under directional light, new cast-shadow shall be added with the consideration of the lighting conditions and its surroundings. To solve this type of problems, higher level knowledge and user intervention are required.

REFERENCES

[1] M. Grundland, R. Vohra, G. P. Williams, and N. A. Dodgson, "Cross dissolve without cross fade: Preserving contrast, color and salience in

image compositing," in *Proc. of EuroGraphics*, no. 3, 2006.

[2] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Siggraph*, no. 3, pp. 313–318, July 2003.

[3] T. Georgiev, "Photoshop healing brush: a tool for seamless cloning," in *Workshop on Applications of Computer Vision in conjunction with ECCV*, 2004.

[4] ——, "Covariant derivatives and vision," in *ECCV*, 2006, pp. 56–69.

[5] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *ACM Siggraph*, 2004, pp. 294–302.

[6] Y. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 11, pp. 1222–1239, 2001.

[7] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum, "Drag-and-drop pasting," *ACM Siggraph*, no. 3, pp. 631–636, July 2006.

[8] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": Interactive foreground extraction using iterated graph cuts," in *ACM Siggraph*, 2004, pp. 309–314.

[9] J.-F. Lalonde, D. Hoiem, A. A. Efros, C. Rother, J. Winn, and A. Criminisi, "Photo clip art," *ACM Siggraph*, vol. 26, no. 3, August 2007.

[10] D. Leventhal, B. Gordon, and P. G. Sibley, "Poisson image editing extended," in *ACM Siggraph Research posters*, 2006.

[11] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral matting," in *IEEE Proc. of Computer Vision and Pattern Recognition (CVPR)*, June 2007.

[12] J. Wang and M. F. Cohen, "Optimized color sampling for robust matting," in *IEEE Proc. of Computer Vision and Pattern Recognition (CVPR)*, June 2007.

[13] ——, "Simultaneous matting and compositing," in *IEEE Proc. of Computer Vision and Pattern Recognition (CVPR)*, June 2007.

[14] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. of International Conference on Computer Vision (ICCV)*, 2001, pp. 105–112.

[15] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. on Graphics*, no. 4, pp. 217–236, October 1983.

[16] A. Zomet, A. Levin, S. Peleg, and Y. Weiss, "Seamless image stitching by minimizing false edges," *IEEE Trans. on Image Processing*, vol. 15, no. 4, pp. 969–977, April 2006.

[17] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt, "Shape representation and classification using the poisson equation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 12, pp. 1991–2005, December 2006.

[18] A. Eden, M. Uyttendaele, and R. Szeliski, "Seamless image stitching of scenes with large motions and exposure differences," in *IEEE Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 2498–2505.

[19] L. Grady, "Random walks for image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 11, pp. 1768–1783, November 2006.

[20] J. S. DeBonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," in *ACM Siggraph*, 1997, pp. 361–368.

[21] T.-P. Wu, C.-K. Tang, M. S. Brown, and H.-Y. Shum, "Natural shadow matting," *ACM Siggraph*, no. 2, 2007.

Jianmin Zheng received the BS and PhD degrees from Zhejiang University, China. He is currently an assistant professor in the School of Computer Engineering at Nanyang Technological University, Singapore. Previously, he was a faculty member at Zhejiang University and a research faculty at Brigham Young University in Provo, Utah. His research interests include computer aided geometric design, CAD/CAM, computer graphics, animation, digital imaging, and visualization.

Jianfei Cai (S'98-M'02-SM'07) received his PhD degree from University of Missouri-Columbia in 2002. Currently, he is an Associate Professor with Nanyang Technological University, Singapore. His major research interests include digital media processing, multimedia compression, and multimedia networking technologies. He has published more than 80 technical papers in international conferences and journals. He has been actively participated in program committees of various conferences. He s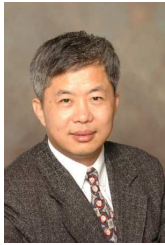erved as one of the track co-chairs for IEEE ICME 2006 & 2008 & 2009, the technical program co-chair for Multimedia Modeling (MMM) 2007 and the conference co-chair for Multimedia on Mobile Devices 2007. He is also an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT). He is a senior member of IEEE.

Susanto Rahardja received his Ph.D. degree from the Nanyang Technological University (NTU, Singapore) in Electrical & Electronic Engineering. Currently, he is a Principal Scientist, Director of Personal 3D Entertainment System program and Head of Signal processing Department at the Institute for Infocomm Research (I^2R). He is also a Program Director of Science & Engineering Research Council of A*STAR. Dr Rahardja is a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE). He was the recipient of the IEE Hartree Premium Award for the best journal paper published in IEE Proceedings in 2002. In 2003, Dr Rahardja received the prestigious Tan Kah Kee Young Inventors Gold award in the Open Category, for his contributions to scalable to lossless audio compression technology. From 2002-2006, Dr Rahardja participated in the international ISO/IEC JTC1/SC29/WG11 (Moving Picture Expert Group, or MPEG) where he contributed to the development of MPEG-4 Scalable to Lossless System (SLS) in which his technology was adopted and published as an international standard in June 2006 (ISO/IEC 14496-3:2005/Amd.3:2006). In recognition of his significant contributions to the national standardization program and advancement of digital audio signal processing and its adoption to the MPEG, Dr Rahardja was awarded the Standards Council Merit Award by SPRING Singapore and the National Technology Award in 2006 and 2007 respectively. He also received the A*STAR Most Inspiring Mentor Award in 2008.

Dr Rahardja has served in several boards, advisory and technical committees in various IEEE and SPIE related professional activities in the areas of multimedia. He is elected members of the Technical Committee of the Visual Signal Processing and Communications, Circuits and Systems for Communications and Multimedia Systems and Applications of the IEEE Circuits & Systems Society. He is currently serving as Associate Editors for IEEE Transactions on Audio, Speech and Language Processing, the Journal of Visual Communication and Image Representation and IEEE Transactions on Multimedia. He served as the Conference Chair of Multimedia Systems and Applications at SPIE OpticsEast Symposium from 2006 to 2007 as well as Symposium Co-chair of the IEEE International Symposium on Multiple-Valued logic in 2006. He was also the Industry and Government Advisor of 1st ACM SIGGRAPHASIA 2008 and the General Chair of 7th ACM SIGGRAPH VRCAI 2008. He is currently the President of SIGGRAPH Singapore Chapter (SSC) and Southeast Asia Graphics (SEAGRAPH) society.

Wenxian Yang received the B.Eng. degree from Zhejiang University, China, in 2001, and the PhD degree from Nanyang Technological University, Singapore, in 2006. She is currently a research fellow in the School of Computer Engineering at Nanyang Technological University.

She was a Postdoctoral Researcher with the French National Institute for Research in Computer Science and Control (INRIA-IRISA), France from 2005 to 2006, and a Postdoctoral Fellow with the Chinese University of Hong Kong from 2006 to 2007. Her research interests include image and video processing, video compression and 3-D video.

Chang Wen Chen (F'04) is a full professor at the Department of Computer Science and Engineering, University at Buffalo, the State University of New York. Previously, he was the Allen S. Henry Distinguished Professor of Electrical and Computer Engineering at the Florida Institute of Technology from 2003 to 2007, on the faculty of Electrical and Computer Engineering Department at the University of Missouri-Columbia from 1996 to 2003, on the faculty of Electrical Engineering Department at the University of Rochester from 1992 to 1996. He also served as the Head of Interactive Media Group at David Sarnoff Research Labs from 2000 to 2002. He is a Fellow of IEEE.

Currently, he is serving as Editor-in-Chief for IEEE Trans on Circuits and Systems for Video Technology and an Associate Editor for IEEE Trans on Multimedia. He has served as Technical Program Committee Chair for ICME2006 held in Toronto, Canada. His research interests include image and video coding, joint source and channel coding, wireless and Internet video, wireless sensor networks, and multimedia communication and networking. His research is supported by NSF, DARPA, NASA, Whitaker Foundation, and Kodak.

He received his BS from University of Science and Technology of China in 1983, MSEE from University of Southern California in 1986, and Ph.D. from University of Illinois at Urbana-Champaign in 1992.