# An Image Completion Algorithm Using Occlusion-Free Images from Internet Photo Sharing Sites

Hanieh AMIRSHAHI[†a], *Nonmember*, Satoshi KONDO[††], Koichi ITO[†††], *and* Takafumi AOKI[†††], *Members*

**SUMMARY**    In this paper, we propose an image completion algorithm which takes advantage of the countless number of images available on Internet photo sharing sites to replace occlusions in an input image. The algorithm 1) automatically selects the most suitable images from a database of downloaded images and 2) seamlessly completes the input image using the selected images with minimal user intervention. Experimental results on input images captured at various locations and scene conditions demonstrate the effectiveness of the proposed technique in seamlessly reconstructing user-defined occlusions.
*key words:*  *image completion, image analysis, image composting, image processing, database searching*

## 1.  Introduction

It is not unusual for a photograph taken at a scenic spot or a famous monument to be full of occlusions. Getting the perfect shot with no occlusion is a challenging and time consuming task.  The occlusions in the image can be manually removed using various interactive tools like Photoshop clone stamp. However, the process is time consuming and the final result depends on user skills and may show signs of duplication.  Image completion is a technique for filling-in or replacing an image region with new image data. Its objective is to create seamless and coherent completed images, so that an observer who is not familiar with the original input image cannot detect the modifications.  Automatic and seamless image completion is a problem under research.

Conventional image completion algorithms can be classified into two categories:  1) those which use the image data available in the *same* input image and 2) others which stitch sections from *different* images together.  Image inpainting [1]–[4] and exemplar-based image completion techniques [5]–[8] are representative of the first category. Both techniques assume that the necessary image data to fill-in the occlusion is somewhere in the same given image.  This assumption can only hold for either images with

very simple backgrounds or small low textured occlusions. The result may still show signs of duplication. There is also the fact that the single image at hand does not contain sufficient image data for coherent completion of the occlusion.

Image stitching algorithms [9]–[16] are representative of the second category.  They can further be classified into two groups based on their objective of 1) accurately reconstructing or 2) finding acceptable substitutes for the occluded region.  Refs. [9]–[12], which belong to the first group, assume that the stack of input images to be stitched together are all globally registered or taken under the same scene/camera parameters. Refs. [13]–[16], which belong to the second group, stitch together semantically valid regions from multiple images to produce composite images which usually portray completely different or to-some-extent altered scenes from that of the input image.  Hays et al. [13] present an algorithm which uses a huge database of 2 million images to find images containing semantically valid scenes which could be used to replace a user-specified region in an input image.  Not particularly aiming to complete occluded objects and scenes with their true selves from the database, Hays et al. use other image content to insert entirely novel objects and textures into the original image.  Refs. [14] and [15] demonstrate the use of multiple images from an annotated database for image synthesis.  However, in Ref. [14], the 100 or so image regions available in the database are annotated manually with semantic labels and image characteristics.  In Ref. [15], the database labeling process is done semi-automatically by allowing a human to make the final decision among the labeling results automatically produced by the system. Finally, Ref. [16] builds on Agarwala et al.'s work [10] and uses a few manually selected images with useful image regions to complete user-specified regions in an input image.

Addressing these limitations, we propose an image completion algorithm which automatically completes occlusions using the occlusion-free regions selected from the image database.  Our approach is to use the images available on Internet photo sharing sites to remove and reconstruct user-identified occlusions in an input image.  While being among one of the first works which utilizes the sea of images available on the Internet for image completion, the main contribution of this work is in introducing a technique for automatic selection of occlusion-free image regions from a database of downloaded images. Automatic selection of images which satisfy certain selection criteria from a database of images has been done before. In conventional algorithms,
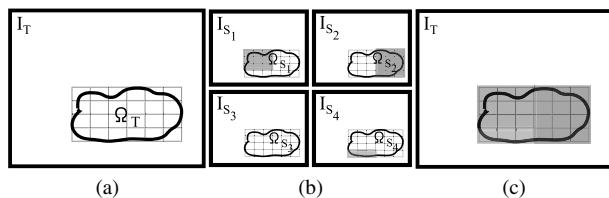
Hays et al. [13] take a data-driven approach and use the gist scene descriptor [17], [18] to find images depicting similar scenes as the input image. Brown et al. [19] use the SIFT features to automatically discover matching relationships between images and recognize panoramas in an unordered dataset. However, neither of these works tackles the issue of occlusion in their automatically selected images which is discussed in this paper. On the other hand, in our proposed algorithm, the user only needs to (i) input the original image to be completed, (ii) identify the occluded area to be removed and (iii) provide a keyword which describes the input image. The algorithm would then *automatically* select the most suitable images which would seamlessly remove the occlusion and output the completed image without any further user intervention. The effectiveness of the proposed technique has been demonstrated through experimental results using various input images.

The structure of this paper is as follows. Section 2 presents an overview of the proposed algorithm. Section 3 presents our image selection and alignment technique. Section 4 introduces our occlusion-free image selection method. Section 5 describes how the occlusion-free image regions are stitched together to fill-in the target fragment. Section 6 presents our experimental results and discussion. In Sect. 7, we end with the conclusion and future work plans.

## 2. Overview

Figure 1 illustrates the image completion problem being addressed in this paper. Let the *target image* $I_T$ be the input image to be completed and the *target fragment* $\Omega_T$ be the user-identified occlusion in $I_T$. Our assumption is that the scene depicted in the target image can be identified with a unique keyword. We use this keyword to tag-search and download images from Internet photo sharing sites. In this paper, we download images from Flickr [20]. Images in the database which satisfy the selection criteria (to be explained in Sect. 3) are referred to as *source images* $I_S$. Let the *source fragment* $\Omega_S$ be the region in an aligned source image which corresponds to $\Omega_T$. In this paper, we propose a technique to automatically detect occlusion-free regions in $\Omega_S$ and use them to complete $\Omega_T$ as shown in Fig. 1(c).

We briefly describe the proposed image completion algorithm as presented in Fig. 2. The proposed algorithm consists of three stages: 1) source image selection and alignment, 2) occlusion-free image selection and 3) target image completion. The only user-inputs required in this algorithm are (i) the target image $I_T$ to be completed, (ii) the roughly indicated target fragment $\Omega_T$ and (iii) a keyword which identifies the scene depicted in the target image. We first create a local database of related images using the provided keyword. In the source image selection and alignment stage (Stage 1), the algorithm searches through the local database and automatically retrieves images which satisfy certain selection criteria. These selection criteria, which would be explained in detail in Sect. 3, ensure that the selected source images $I_S$ depict the same physical scene as the target image from as close a viewpoint as possible. In the occlusion-free

---

**Input**

1. Target image $I_T$
2. Target fragment $\Omega_T$
3. Keyword identifying $I_T$

**Output**   Completed target image

**Algorithm**

**Stage 1: Source Image Selection and Alignment**

**1.1**   Create local image database from Internet photo sharing sites using provided keyword.

**1.2**   Select suitable source images $I_S$ from the local database.

**1.3**   Accurately align source images $I_S$ with the target image $I_T$.
Let $\Omega_S$ be the region in the aligned source images which corresponds to $\Omega_T$.

**Stage 2: Occlusion-Free Image Selection**

**2.1**   Create gradient images for all aligned source images $I_S$. Divide gradient $\Omega_T$ and $\Omega_S$ into small patches.

**2.2**   For every patch in $\Omega_T$:

- Calculate the Sum of Squared Differences (SSD) between all pairs of gradient source fragments.
- Connect all pairs of gradient source fragments whose SSD value is less than a threshold.
- Choose the biggest group of connected source gradient fragments.
- Select the gradient source fragments in the above group whose total number of connections is more than a threshold.

If there exists a patch in $\Omega_T$ for which no source fragment patches were selected, change the thresholds and repeat Step 2.2 once.

**2.3**   Sort the source images $I_S$ according to the total number of selected patches from Step 2.2.

**Stage 3: Target Image Completion**
Set $i = 1$.

**3.1**   Find the best seams along which to cut the source fragment $\Omega_{S_i}$.

**3.2**   Align the intensity values between the target image $I_T$ and the source fragment region from Step 3.1.

**3.3**   Update the target image using results from Step 3.2. Increment $i$ and repeat Steps 3.1 and 3.2 while there still exists uncompleted target fragment patches and unused source images.

---



**Fig. 1**   Image completion problem: (a) the target image $I_T$ with target fragment $\Omega_T$ to be completed, (b) the aligned source images $I_S$ with the source fragments $\Omega_S$ which correspond to $\Omega_T$, and (c) the completed target image where the colored regions indicate the section used from the corresponding source fragment $\Omega_S$.

**Fig. 2**   The proposed image completion algorithm.

image selection stage (Stage 2), the algorithm further analyzes the selected source images to automatically find the occlusion-free regions in them. Finally, in the target image completion stage (Stage 3), the algorithm stitches the occlusion-free regions from multiple source images together to seamlessly complete the target fragment.

The main part of the proposed algorithm is the occlusion-free image selection stage (Stage 2) which distinguishes this work from others. Without this stage, regardless of the number of selected source images, we would not be able to guarantee that the occlusion in the target image has been faithfully completed. The occlusion-free image selection stage makes it possible to reduce the required user-intervention. The idea of using multiple images to remove occlusions in an input image has been presented before in the literature [9], [10], [12], [14], [16]. However, this paper is one of the first works which proposes a technique to *automatically* retrieve the most suitable occlusion-free image regions from a database and seamlessly stitch them together to complete the target image.

## 3. Source Image Selection and Alignment

The first stage of our image completion algorithm involves creating a local database of related images, choosing the most suitable source images from the database, and accurately aligning the selected images with the target image.

*Step 1.1*

This step focuses on creating a local database by downloading images from Internet photo sharing sites using metadata. In this research, we assume that the scene depicted in the target image can be uniquely identified with a keyword. We use the provided keyword for creating the database. Image retrieval using tag-search technique is inaccurate because the keyword given to an image might be generalized, incorrect or vary based on the user's choice of words. Thus, we need to select those images which depict the same physical scene as the target image.

*Step 1.2*

The aim of this step is to automatically retrieve suitable images from the database which can later be used to complete the target image. For this purpose, we use an object matching approach based on Scale Invariant Feature Transform (SIFT) keypoint detectors [21], [22]. The detected SIFT keypoints are invariant to changes in illumination, rotation and scaling. We detect and match the SIFT keypoints between the target image and every image in the local database. The number of matched keypoints between two images depicting the same object decreases when there is a considerable change in viewpoint. For example, despite the fact that the two images in Fig. 3 depict the same monument, the variation in viewpoint between the images has caused the total number of matched SIFT keypoints to decrease. Hence, similar to Ref. [23], we use the total number of matched SIFT keypoints as a criterion for automatically selecting images which depict the same physical scene as the target image from as close a viewpoint as possible, with-

out considering the extent of image transformation. Based on our experimental results, images with at least 20 matched SIFT keypoints satisfy the above condition.

Other criteria for selecting suitable images include the differences in the image size and the scale of the objects between the downloaded images and the target image. Both of these criteria affect the quality of the final completed image. Hence, under the assumption of same image resolution, we use the coordinates of the matched keypoints to estimate the object scale ratio. The scale ratio $\rho$ is calculated as the mean value of the ratio of the distance between any pair of matched SIFT keypoints in the target and their corresponding points in the query image as follows:

$$\rho = \frac{1}{N} \sum_{\substack{i,j \\ i \neq j}} \frac{(x_i' - x_j')^2 + (y_i' - y_j')^2}{(x_i - x_j)^2 + (y_i - y_j)^2}, \tag{1}$$

where $N$ is the total number of matched keypoint pairs, $i$ and $j$ are the index numbers of the matched keypoints, $(x_i, y_i)$ and $(x_i', y_i')$ are the coordinates of matched keypoints in the target and the selected query image, respectively. The ratio in Eq. (1) is a convenient way to compare the size of the objects depicted in the query image versus the target image. We refer to this as the object scale ratio between the two images. We only select those images whose object scale ratio is within a set threshold. This prevents the need for extreme zooming in and out of our selected source images when aligning them with the target image in Step 1.3. Also, to prevent excess enlarging, we discard images whose size is too small compared to the target image. Thus, our criteria and selection order is as follows: images with

1. at least 20 matched keypoints,
2. object scale ratio $\rho$ between 0.7 and 1.4 and
3. image size of at least 80% of that of the target image,

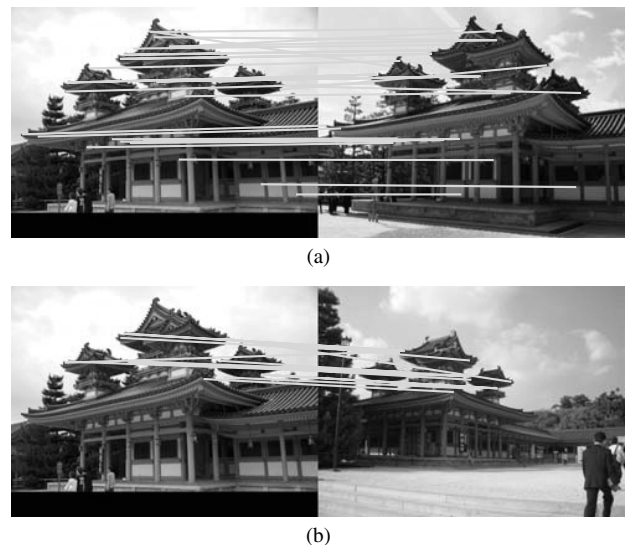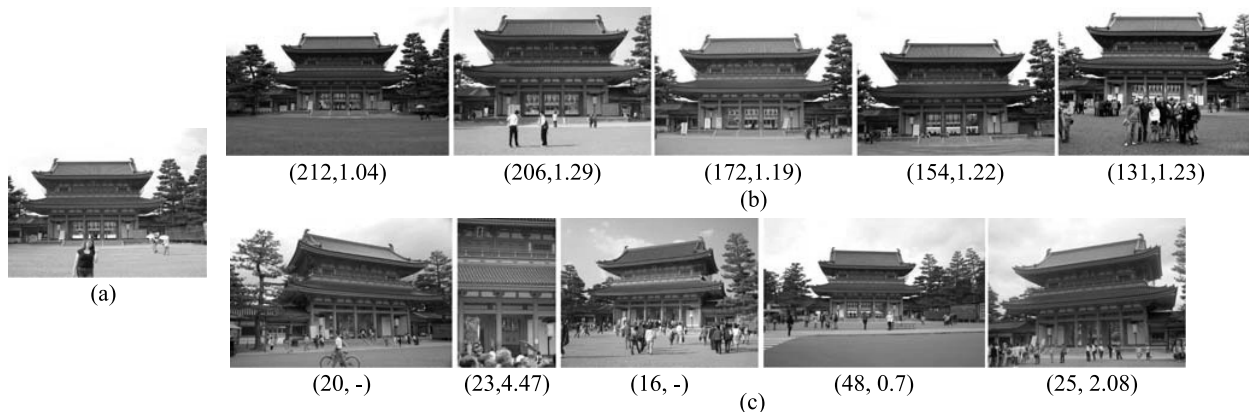are selected as suitable source images. Figure 4 presents a



(a)



(b)

**Fig. 3** Example of corresponding SIFT keypoints between two images where the total number of matched points is 31 in (a) and is 15 in (b).

(212,1.04)　(206,1.29)　(172,1.19)　(154,1.22)　(131,1.23)
(b)

(a)

(20, -)　(23,4.47)　(16, -)　(48, 0.7)　(25, 2.08)
(c)

**Fig. 4** Results of source image selection for "Heian Shrine" database: (a) the target image $I_T$, some of the (b) selected source images $I_S$, and (c) unselected images. The values in the brackets are the total number of matched SIFT keypoints and the object scale ratio. "-" means that no object ratio could be calculated due to insufficient matched keypoints.

sample of our image selection results.

*Step 1.3*

To ensure a coherent completion result, the source images selected in the previous step have to be accurately aligned with the target image. This is done under a full planar perspective motion model in two stages: 1) global alignment using the matched SIFT keypoints and 2) local alignment focusing around the target fragment $\Omega_T$ and its corresponding region in the globally aligned source image $I_S$. For the first alignment, we use the matched SIFT keypoints to estimate a homography matrix to globally transform the source image. For homography estimation, we use the RANSAC [24] algorithm with a maximum of 1,000 iterations and an error threshold of $10^{-4}$. Figure 5 (a) shows the result of completing the target image using a globally aligned source image. In this case, the majority of the detected SIFT keypoints are concentrated on the Buddha statue with little or no matching keypoints from the area around the target fragment $\Omega_T$. Therefore, the alignment of the source and target image around the target fragment is inaccurate. The result is noticeable edge discontinuities in the image as shown in Fig. 5(a). Thus, we need more locally oriented alignment which focuses on the region around the target fragment $\Omega_T$.

For the second alignment, we employ the sub-pixel correspondence search algorithm using Phase-Only Correlation (POC) introduced in Ref. [25]–[27]. The POC-based algorithm provides us with a dense correspondence relationship and can detect minute translation between the target image and the globally aligned source image. We define the search region as all the pixels within a 30-pixel radius of the boundary of the target fragment $\Omega_T$ and its corresponding area around the source fragment $\Omega_S$ in the globally aligned source image. The dense POC-based correspondence means that even if minor occlusions occur in either the area around the target fragment $\Omega_T$ or the source fragment $\Omega_S$, there would still be sufficient corresponding points to estimate a second homography matrix. We use correspond-
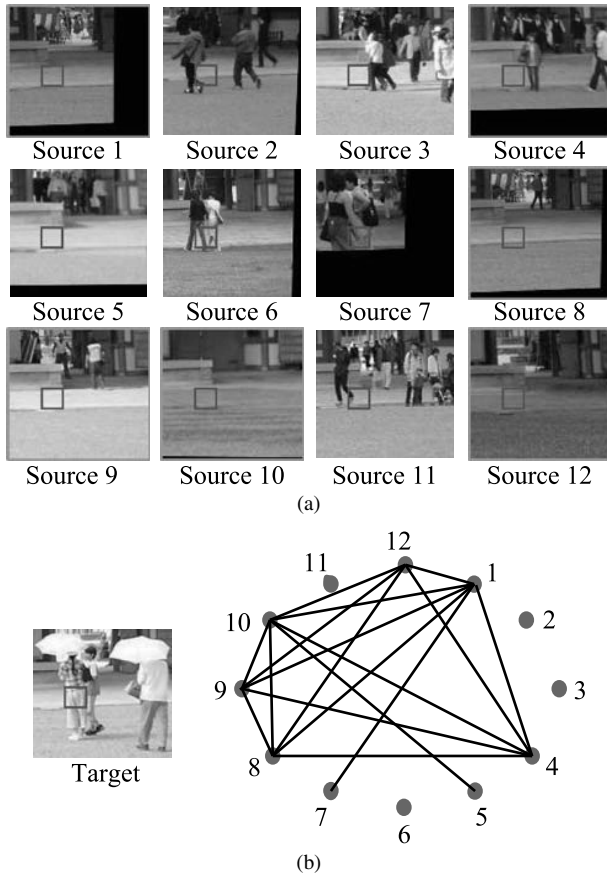


(a)　(b)
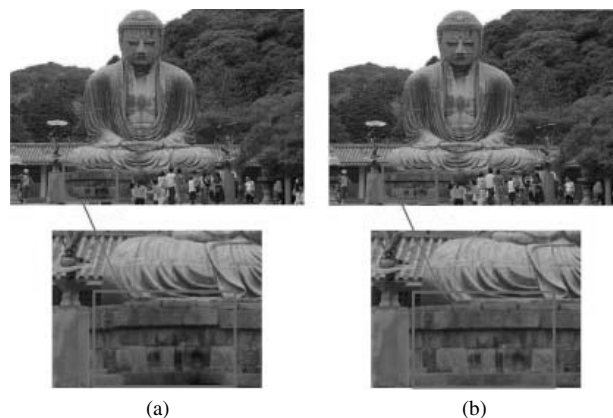
**Fig. 5** Effect of local alignment: the completed target image using (a) globally and (b) locally aligned source fragments.

ing points with a POC peak value of over 0.7 to estimate a second homography matrix to locally align the source image. The effect of this second alignment can be noticed by comparing the images in Figs. 5(a) and (b). Thus, at the end of Stage 1, we obtain a list of suitable source images $I_S$ which are accurately aligned with the target image $I_T$. In the rare case, when the search area is completely occluded or there is insufficient texture (as when matching patches of over-exposed grass or sky) for satisfactory POC-based correspondence, we rely only on the results of the first global alignment.

## 4. Occlusion-Free Image Selection

The aim of Stage 2 of the algorithm is to find occlusion-free image regions from the source images $I_S$ selected in Stage 1. The regions in the aligned source images, which correspond to the target fragment $\Omega_T$, are extracted as source fragments $\Omega_S$. However, the extracted source fragments $\Omega_S$ may sometimes contain partially occluded regions. We

Source 1    Source 2    Source 3    Source 4

Source 5    Source 6    Source 7    Source 8

Source 9    Source 10    Source 11    Source 12

(a)

Target

(b)

**Fig. 6** An example of occlusion-free patch selection: (a) aligned source fragments with the query patch denoted by a small square box and (b) graph $G_\kappa$ representing the SSD-wise relationship between the gradient source fragments for the query patch indicated in the target fragment. The source fragments with the frame border are selected as suitable for completing the query patch in the target fragment.

propose a patch-based verification method to automatically identify occlusion-free source fragment regions needed for accurate target image completion. The following explains the details of Stage 2.

*Step 2.1*

Comparison of the source fragments together is one way to find the common occlusion-free regions among them. Pixel color value is an unsuitable comparison metrics because the source images are captured under various camera/scene parameters. Therefore, we compare the source fragments' gradients instead of their pixel color value. The comparison is done in a patch-wise basis as represented by the small squares in Fig. 1. In this step, we create gradient images of all aligned source images and divide the gradient source fragments into square patches. In our experiments, we set the patch size to $20 \times 20$ pixels. We need to find the occlusion-free regions in the source fragments and use them to complete the target fragment $\Omega_T$.

*Steps 2.2*

In this step, we perform the patch-wise comparison of the gradient source fragments to detect their occlusion-free image regions. For *every patch* $\kappa$ in $\Omega_T$, we represent each



(a)     (b)     (c)     (d)
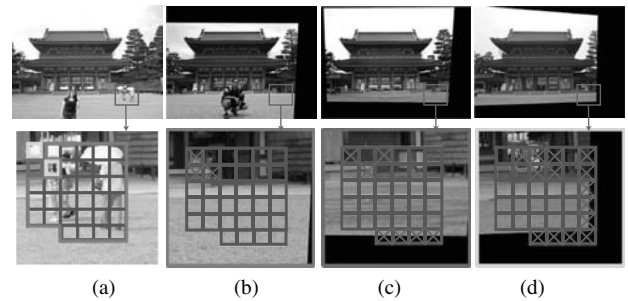
**Fig. 7** Results of occlusion-free patch selection: (a) the target image and its fragment, (b)-(c) the top-3 source images and their fragments, respectively. × denotes occluded patches. The color coding represents how the target fragment patches are assigned to different source fragments.

gradient source fragment as a node on a graph $G_\kappa(V_\kappa, E_\kappa)$, where $V_\kappa$ is the set of nodes and $E_\kappa$ is the set of edges. We calculate the Sum of Squared Differences (SSD) between all pairs of gradient source fragments over the query patch $\kappa$. If the SSD value is small, we connect the corresponding nodes on the graph. Figure 6 (a) shows an example of the source fragments with the query patch indicated. Figure 6 (b) shows the SSD-wise relationship between the source fragment patches in (a). We focus on the biggest connected subgraph $G'_\kappa(V'_\kappa, E'_\kappa)$. Let $|V_k|$ be the size of set $V_k$. Nodes in $G'_\kappa$ with a minimum degree of $\frac{|V_k|}{3}$ are selected as suitable for completing the target fragment patch $\kappa$. After each iteration, the selected nodes are output as $V_{select}(\kappa)$. If there exists a target fragment patch which has not been assigned any source fragment yet, we increase the SSD threshold and change the nodes' degree threshold to $\frac{|V_k|}{4}$, and repeat Step 2.2. Based on the proposed method, the source fragments in Fig. 6(a) marked with a frame border are chosen as suitable for completing the indicated target fragment patch in Fig. 6(b).

*Step 2.3*

In this step, based on the $V_{select}$ results, we rank the source images according to the total number of patches which they can complete in the target fragment. Figure 7 shows the top 3 source images for the "Heian Shrine" database. The enlarged sections in this figure present the result of the occlusion-free patch detection from the previous step. The crossed out patches are believed to be occluded. Thus, at the end of Stage 2, we obtain a reordered list of source images with their occlusion-free source fragment regions clearly indicated.

## 5. Target Image Completion

In Stage 3, we seamlessly stitch the occlusion-free source fragment regions onto the target image.

*Step 3.1*

In this step, we start with the highest ranked source image and use the occlusion-free patches from this source image to complete the corresponding patches in the target fragment. As shown in Fig. 8(a), simply cutting and pasting the patches from the source fragment onto the target im-
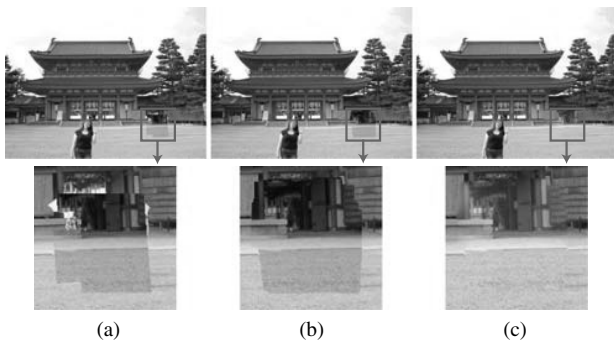
**Fig. 8** An example of the completed target image using (a) cut and paste, (b) graph cut, and (c) graph cut and Poisson blending, respectively.
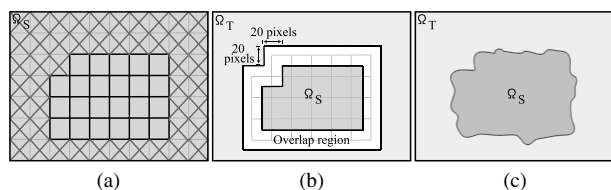


**Fig. 9** The constraints used for the minimum cost graph cut algorithm: (a) the source fragment $\Omega_S$ with the occluded patches crossed out with ×, (b) target fragment $\Omega_T$ with the 20 pixel wide graph cut overlap region indicated, and (c) graph cut selection results.

age leaves behind obvious seams. Addressing this problem, we use the minimum cost graph cut algorithm [28]–[30] to find the best seams along which to cut the source fragment. The idea here is to construct a graph and find the minimum cost cut which divides the graph into two disjoint subgraphs. Each of these subgraphs are then assigned to either the target or the source fragment. Figure 9 shows the constraints used for the minimum cost graph cut algorithm. The area surrounded by the inner border of the overlap region is constrained to the source fragment $\Omega_S$. The area surrounding the outer border of the overlap region is constrained to the target fragment $\Omega_T$. We use the minimum cost graph cut algorithm to assign a label of $I_T$ or $I_S$ to each of the pixels in the overlap region. The solution to the labeling problem would identify the best seams along which to cut the source fragment.

The most decisive factor in minimum cost graph cut algorithm is the cost function being minimized. Depending on how we define the cost function, the final seam selection outcome can change dramatically. Human eye is more sensitive to edge discontinuity than to smooth color transition in an image. Thus, we focus on the object edges in the image and employ a cost function such as that introduced in Ref. [31], which avoids breaking object edges. Let $p$ and $q$ be two adjacent pixels in the overlap region. The cost $C(p, q)$ between pixels $p$ and $q$ is defined for each channel in the RGB color space as:

$$C(p,q) = \sum_{R,G,B} \{(1 - \beta)C_s(p,q) + \beta C_d(p,q)\}, \qquad (2)$$

where $C_s$ and $C_d$ are cost functions which measure the ex-
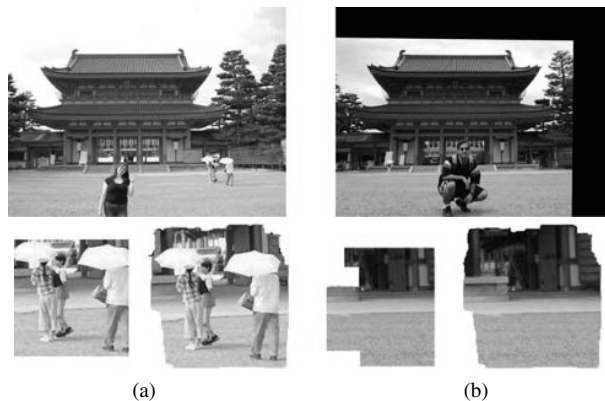


**Fig. 10** An example of graph cut selection results: (a) the target image and its fragments before and after graph cut and (b) the highest ranked source image from Fig. 8(b) with its fragments before and after graph cut.

tent of gradient smoothness and gradient similarity between neighboring pixels, respectively. $\beta$ is a weight factor to determine the relative influence of the two costs and is set to 0.3 in this paper. $C_s$ avoids breaking object edges between $p$ and $q$ and is defined as:

$$C_s(p,q) = \qquad (3)$$
$$\|\nabla I_T(p)\| + \|\nabla I_T(q)\| + \|\nabla I_S(p)\| + \|\nabla I_S(q)\|,$$

where $\|\nabla I_T(p)\|$ and $\|\nabla I_S(p)\|$ refer to the norm of the image gradient at pixel $p$ in the target and source images, respectively. $C_d$ ensures pixel similarity in the gradient domain along the horizontal ($x$-axis) and vertical ($y$-axis) directions and is defined as:

$$C_d(p,q) = \qquad (4)$$
$$\|\nabla_x I_S(p) - \nabla_x I_T(p)\| + \|\nabla_x I_S(q) - \nabla_x I_T(q)\|+$$
$$\|\nabla_y I_S(p) - \nabla_y I_T(p)\| + \|\nabla_y I_S(q) - \nabla_y I_T(q)\|,$$

where each term represents the gradient-level similarity at the same pixel location in the overlap region. As presented in Fig. 8(b), this cost function helps us find the optimal seams along which to cut the source fragment $\Omega_S$ so that it seamlessly blends in with the rest of the target fragment $\Omega_T$.

The minimum cost graph cut results takes all the object edges in the overlap region into consideration. On the other hand, the occlusion-free patch assignment result from Stage 2 only considers the object edges in the query patch. Thus, as is presented in Fig. 10, it is very likely that the graph cut selection result would expand to include patches which were originally assigned to other source fragments. Allowing this reassignment of patches ensures a more seamless completion result.

*Step 3.2*

In this step, we modify the intensity of the source fragment region selected in the previous step to ensure that it smoothly blends in with the target image. For this purpose, we use Poisson blending [5]. In Poisson blending, we compute a function whose gradient is closest in the $L_2$-norm to

some prescribed vector field referred to as the guidance vector field. Simply said, Poisson blending is an image interpolation technique using the guidance vector field. Here, our aim is to compute the pixel intensity values in $\Omega_T$. For this purpose, we set the pixel gradients in the target fragment $\Omega_T$ equal to the pixel gradients in the source fragment $\Omega_S$, and interpolate the intensity values of the pixels on the border of $\Omega_T$, inwards. As presented in Fig. 8(c), the combination of graph cut and Poisson blending allows a seamless completion of the target fragment.

*Step 3.3*

In the final step, we paste the Poisson blending results from the previous step onto the target fragment and update the target image. If there exists any uncompleted patches in the target fragment, we repeat Steps 3.1 and 3.2 using the next highest ranked source image. At the end of Stage 3, we obtain the completed target image.

## 6. Experimental Results and Discussions

We evaluate our proposed algorithm using a variety of input images captured at different locations and scene conditions. The images in our databases are downloaded from Flickr [20] using the given keyword under the *full text* option. In order to prevent excess downloading, we initially download approximately the first 300–600 images found using the given keyword. Due to the inaccuracy of keyword search technique, it is very likely that among the images downloaded there also exists many images captured at completely different locations to that of the input image. Thus, we run our robust source image selection program to determine if we have obtained enough suitable images in our downloaded batch. In the rare unfortunate cases when only a few suitable images have been downloaded, we download the next 300–600 images and repeat the process. The images are all down-sampled so that their maximum dimension is 800 pixels before running the SIFT based object recognition.

Table 1 summarizes the result of our image selection algorithm. The accuracy of the automatic selection was verified by individually checking the selected source images. They all depicted the same physical scene as the target image from very close viewpoints. While it is very possible that the user would have selected all the images in Fig. 4 as source images, the automatic selection process has ensured that *only* the most suitable images are retrieved. This proves

the accuracy of the proposed selection method. Some of the selected source images from each database are presented in Fig. 11(b).

As indicated in Fig. 7, the proposed occlusion-free image selection technique is effective in screening out the partially occluded regions in the source fragments. Based on this result, the algorithm selects only the top few source images needed to completely replace the occlusion in the target fragment. The occlusion-free image selection results for other databases are indicated in Fig. 11(b). Careful analysis of the results shows that the selected images are in fact the most suitable for completing the target image. The effectiveness of Stage 2 of the algorithm is the distinguishing factor which allows us to minimize the human intervention required in our proposed image completion algorithm.
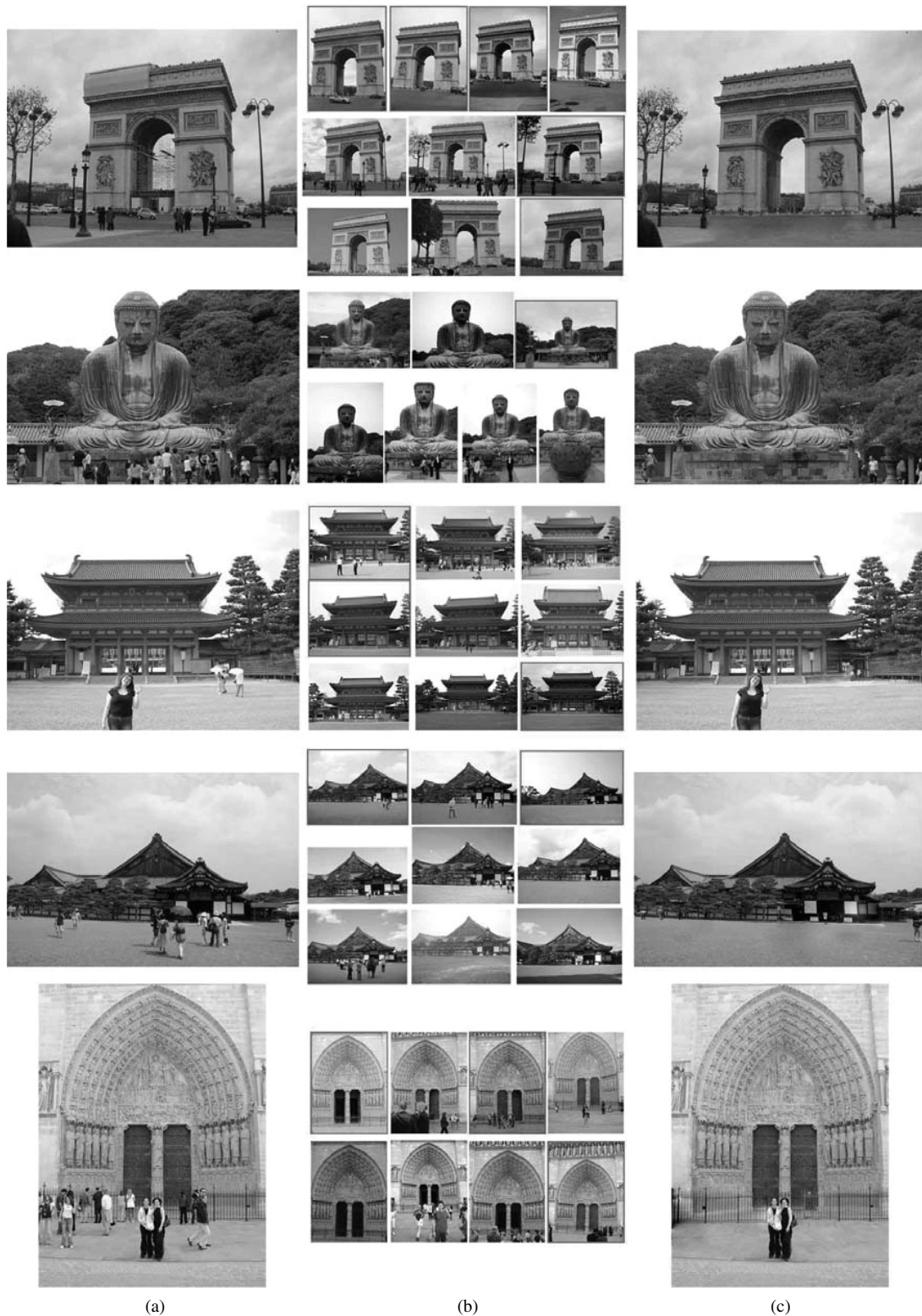
Figure 11(c) presents our completion results for the input images presented in Fig. 11(a). As it can be seen, using the proposed method, we have seamlessly replaced the occlusions and coherently completed the target image. In particular, the "Arc de Triumph" image completion result portrays the power of the proposed technique. It would be quite challenging to accurately reconstruct the section of the arc under renovation using exemplar-based image completion techniques or algorithms which use multiple images taken on the same day.

Figure 12 presents two examples of the limitations of our algorithm. In Fig. 12(a), we were able to remove the green net and seamlessly complete the construction on St. Basil Cathedral. On the other hand, due to the lack of suitable source images, we could not remove the pedestrians from the image and some visible artifacts were left regardless of our efforts. Merging patches from images captured by completely different cameras from various viewpoints, causes a challenge to the alignment of the images together. In Fig. 12(b), the estimation error of the transformation matrix used for the alignment and the limitations of the full planar perspective motion model assumption has resulted in distinct misalignments. In this image, the majority of the feature points used for estimating the transformation matrix for the first alignment were from the shrine building. The second local alignment failed due to the lack of texture and over exposure in the area around the target fragment. This resulted in the visible misaligned stairs in the final completed result.

Thus, the advantage of our image completion algorithm is that we can replace and reconstruct occlusions of random size which appear in images with random background complexity, with minimal user intervention. Also, the ability of the algorithm to automatically select the most suitable occlusion-free image regions from a database of downloaded images, is the distinguishing factor which signifies the robustness of the proposed technique.

## 7. Conclusion & Future Work

In this paper, we presented an image completion algorithm which takes advantage of the sea of images available on

**Table 1** Results for the proposed image selection method: (a) the total number of images available with the given keyword on Flickr (accessed on 12 December 2007), (b) the number of images in our local database, and (c) the number of source images satisfying all the selection criteria.

| Database Name | (a) | (b) | (c) |
|---|---|---|---|
| Arc De Triumph | 5,665 | 600 | 12 |
| Kamakura Daibutsu | 3,848 | 250 | 13 |
| Heian Shrine | 5,100 | 265 | 12 |
| Nijo Castle | 8,535 | 600 | 14 |
| Notre Dame Entrance | 1,302 | 997 | 25 |

(a)                  (b)                  (c)

**Fig. 11** Experimental results for "Arc De Triumph," "Kamakura Daibutsu," "Heian Shrine," "Nijo Castle" and "Notre Dame Entrance" databases: (a) the target image, (b) the some selected source images, where the source images used for image completion are indicated with a frame border, and (c) the completed target image.

the Internet to complete the user-identified occlusions in an input image. Unlike conventional algorithm, the proposed technique does not impose any restriction on the size of the occlusion nor on the background complexity of the input im-
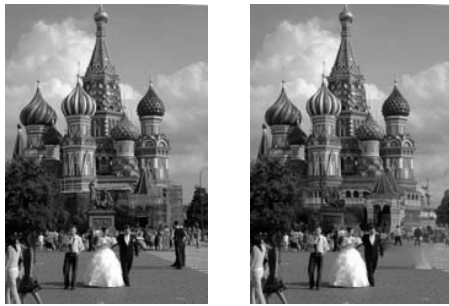
(a)



(b)

**Fig. 12** Examples of the limitations of our algorithm: (a) visible artifacts in the image due to the lack of suitable source images and (b) visible misalignment in the stairs due to the lack of texture and over exposure in the area around the target fragment.

age. It is aimed at any input image which can be identified with a keyword and for which suitable source images can be found on the Internet. We demonstrated the success of the proposed technique using various input images captured with different camera/scene parameters. For future work, we are planning to expand the algorithm to allow video completion. Our idea is to be able to remove an occlusion from all frames of a video sequence.

## References

[1] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles, "Image inpainting," Proc. SIGGRAPH, pp.417–424, July 2000.

[2] C. Ballester, V. Caselles, J. Verdera, M. Bertalmio, and G. Sapiro, "A variational model for filling-in gray level and color images," Proc. Int. Conf. Computer Vision, vol.1, pp.10–16, June 2001.

[3] M. Betalmio, A. Bertozzi, and G. Sapior, "Navier-stokes, fluid dynamics, and image and video inpainting," Proc. Int. Conf. Computer Vision Pattern Recognition, vol.1, pp.355–362, Dec. 2001.

[4] M. Oliveira, B. Bowen, R. McKenna, and Y.S. Chang, "Fast digital image inpainting," Proc. Int. Conf. on Visualization, Imaging and Image Processing, pp.261–266, Sept. 2001.

[5] P. Perez, M. Gangnet, and B. Blake, "Poisson image editing," ACM Trans. Graph., vol.23, no.3, pp.313–318, Aug. 2003.

[6] J. Sun, L. Yuan, J. Jia, and H.Y. Shum, "Image completion with structure propagation," ACM Trans. Graph., vol.24, no.3, pp.861–868, July 2005.

[7] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," IEEE Trans. Image Process., vol.13, no.9, pp.1200–1212, 2004.

[8] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," ACM Trans. Graph., vol.22, no.3, pp.303–312, 2003.

[9] R. Raskar, A. Ilie, and J. Yu, "Image fusion for context enhancement and video surrealism," Proc. Int. Conf. Non-Photorealistic Animation and Rendering, pp.85–95, 2004.

[10] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," ACM Trans. Graph., vol.23, no.3, pp.294–302, Aug. 2004.

[11] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," Proc. 29th Annual Conf. Computer Graphics and Interactive Techniques, pp.249–256, 2002.

[12] C. Herley, "Automatic occlusion removal from minimum number of images," Proc. Int. Conf. Image Processing, vol.2, no.I 1-14, pp.II 1046–1049, Sept. 2005.

[13] J. Hays and A. Efros, "Scene completion using millions of photographs," ACM Trans. Graph., vol.26, no.3, pp.4:1–4:8, 2007.

[14] N. Diakopoulos, I. Essa, and R. Jain, "Content based image synthesis," Proc. Int. Conf. Image and Video Retrieval, pp.299–307, 2004.

[15] M. Johnson, G. Brostow, J. Shotton, O. Arandjelovic, V. Kwatra, and R. Cipolla, "Semantic photo synthesis," Comput. Graph. Forum, vol.25, no.3, pp.407–413, 2006.

[16] M. Wilczkowiak, G. Brostow, B. Tordoff, and R. Cipolla, "Hole filling through photomontage," Proc. British Machine Vision Conference, pp.492–501, 2005.

[17] A. Torralba, K.P. Murphy, W.T. Freeman, and M.A. Rubin, "Context-based vision system for place and object recognition," Proc. Int. Conf. Comp. Vision, pp.273–280, 2003.

[18] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," Progress in Brain Research, vol.155, pp.23–39, 2006.

[19] M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," Int. J. Comput. Vis., vol.74, no.1, pp.59–73, 2007.

[20] Flickr Photo Sharing, http://www.flickr.com/

[21] D. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol.60, no.2, pp.91–110, 2004.

[22] D. Lowe, "Demo software: Sift keypoint detector (ver. 4)," http://www.cs.ubc,ca/lowe/keypoints/

[23] M. Brown and D.G. Lowe, "Recognising panoramas," Proc. Int. Conf. Computer Vision, pp.1218–1225, 2003.

[24] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography," Commun. ACM, pp.381–395, 1981.

[25] K. Takita, M.A. Muquit, T. Aoki, and T. Higuchi, "A sub-pixel correspondence search technique for computer vision application," IEICE Trans. Fundamentals, vol.E87-A, no.8, pp.1913–1923, Aug. 2004.

[26] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," IEICE Trans. Fundamentals, vol.E86-A, no.8, pp.1925–1934, Aug. 2003.

[27] M. Muquit, T. Shibahara, and T. Aoki, "A high-accuracy passive 3d measurement system using phase-based image matching," IEICE Trans. Fundamentals, vol.E89-A, no.3, pp.686–697, March 2006.

[28] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: Image and video synthesis using graph cuts," ACM Trans. Graph., vol.22, no.3, pp.277–286, 2003.

[29] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.11, pp.1222–1239, 2001.

[30] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.9, pp.1124–1137, Sept. 2004.

[31] J. Jia and C. Tang, "Eliminating structure and intensity misalignment in image stitching," Proc. Int. Conf. Computer Vision, 2005.

**Satoshi Kondo** received his B.S., M.S. and Ph.D. degrees at Osaka Prefecture University in 1990, 1992 and 2005, respectively. He joined Matsushita Electric Industrial Co., Ltd. in 1992. His research interests are in the fields of video coding and image/video signal processing. He was also a participant in international standardization activities such as AV/C command sets in IEEE1394 Trade Association and H.264/MPEG-4 Advanced Video Coding in ITU-T VCEG and ISO/IEC MPEG."



**Koichi Ito** received the B.E. degree in electronic engineering, and the M.S. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2000, 2002 and 2005, respectively. He is currently an Assistant Professor of the Graduate School of Information Sciences at Tohoku University. From 2004 to 2005, he was a Research Fellow of the Japan Society for the Promotion of Science. His research interests include signal and image processing, and biometric authentication.



**Takafumi Aoki** received the B.E., M.E., and D.E. degrees in electronic engineering from Tohoku University, Sendai, Japan, in 1988, 1990, and 1992, respectively. He is currently a Professor of the Graduate School of Information Sciences at Tohoku University. For 1997.1999, he also joined the PRESTO project, Japan Science and Technology Corporation (JST). His research interests include theoretical aspects of computation, VLSI systems, multiple-valued logic, digital signal processing, image sensing, computer vision, biometric authentication, and secure embedded systems. Dr. Aoki received the Outstanding Paper Award at the IEEE International Symposium on Multiple-Valued Logic (ISMVL) in 1990, 2000, 2001 and 2006, the Outstanding Transactions Paper Award from the IEICE of Japan in 1989 and 1997, the IEE Ambrose Fleming Premium Award in 1994, the IEICE Inose Award in 1997, the IEE Mountbatten Premium Award in 1999, the Best Paper Award at the IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) in 1999, and the Best Paper Award at the Workshop on Synthesis And System Integration of Mixed Information technologies (SASIMI) in 2007.



**Hanieh Amirshahi** received her B.E. and M.S. degrees in Information Sciences from Tohoku University, Sendai, Japan, in 2006 and 2008, respectively. She is currently working toward the D.S. degree of the Department of Media Technology at Helsinki University of Technology, Finland. Her research interests include signal and image processing, and arbitrary view synthesis.