This paper proposes an example-based method to increase the resolution of a low-resolution image. In the proposed method, we generate example images by a novel view synthesis technique using 3D geometry reconstruction and camera pose estimation from a video or images capturing the same scene. We then increase the resolution by minimizing an energy function by searching for the optimal example from the generated example images. The proposed method has less limitations on camera positions and geometry of the target scene than those in conventional methods. Experiments demonstrate the effectiveness of the proposed method by qualitatively comparing the results of the proposed and conventional methods.
point of the target image using the reconstructed 3D geometry, and the warped images are used as example candidates \( \{I_k^e | k = 1, \cdots, K\} \) for increasing the resolution (2). The candidate images are then narrowed down using camera poses (3). After that, target image \( I_t \) is enlarged by bi-cubic interpolation to the size of the target resolution for giving initial values to generated image \( I_s \) (4). The final result of \( I_s \) is obtained by energy minimization using the example images (5).

### 2.1. 3D reconstruction and image warping

In the proposed method, we estimate camera poses of input images including both target and reference ones and reconstruct 3D geometry of the scene by applying Structure from Motion (SfM) [11, 12] and Multi View Stereo (MVS) [13] to input images. We then generate a depth map with a target resolution for each input image from the reconstructed 3D geometry as shown in Figs. 2(a) and (b). Next, the reference images (Fig. 2(c)) are warped to the viewpoint of the target image by projecting pixel values of the reference images using the depth map of the target image and the estimated camera poses. In warping images, we check the consistency of depths between the target and reference images, and the regions of inconsistency caused by occlusions or estimation errors of 3D model are set as unusable ones as shown in the red regions in Fig. 2(d).

### 2.2. Narrowing down of reference images

Generally, the warped image that is generated from the input image captured near objects includes many higher frequency components than that captured far from them as shown in Fig. 3. Based on this fact, we narrow down the warped reference images so that we can use example images with high-frequency components in the energy minimization process.

Specifically, we determine the pixels in the warped images corresponding to a pixel in \( I_s \). We then select top \( T \) warped images of the smallest depth values of the corresponding pixels. By this way, we independently select \( T \) example images for each pixel in \( I_s \).

### 2.3. Resolution enhancement by energy minimization

The resolution-enhanced image \( I_s \) of target image \( I_t \) is generated by minimizing an energy function using the example images and the original target image. It should be noted that we do not have pixel values in the warped images for the pixels of which depth values do not exist in the depth map of the target image. For these pixels, we just leave initial values generated by bi-cubic interpolation. In this process, the input image is transformed from RGB to YCbCr. After that, the resolution enhancement process is applied to Y (intensity) channel, and Cb and Cr (chromatic) channels are interpolated by bi-cubic interpolation as similar to most of conventional methods for super-resolution, which transform RGB channels to intensity and chromatic ones and use the intensity one for super-resolution. In the following, we first give the definition of the energy function and then describe the minimization process.
2.3.1. **Definition of energy function**

Energy function $E$ is defined using two different kinds of energy terms as follows:

$$E = \sum_{x_i \in I_s} \{E_{sr}(x_i, x_j, k) + \beta E_{data}(x_i)\},$$  \hspace{1cm} (1)

where $E_{sr}$ represents the pattern dissimilarity between generated image $I_s$ and example image $I^k$, and this term gives the effect of increasing the resolution of generated image $I_s$ using the texture including high-frequency components in example image $I^k$. $E_{data}$ represents the intensity difference between generated image $I_s$ and original target image $I_t$, and the term gives the effect of preserving the structure of target image $I_t$ onto generated image $I_s$. $\beta$ is a weight for balancing the two terms. $E_{sr}$ and $E_{data}$ are defined as follows, respectively:

$$E_{sr}(x_i, x_j, k) = \omega(x_i, k) \sum_{\mathbf{p} \in W} \{I_s(x_i + \mathbf{p}) - I^k(x_j + \mathbf{p})\}^2,$$

$$E_{data}(x_i) = \{I_s(x_i) - I_t(Dx_i)\}^2.$$ \hspace{1cm} (2) (3)

Here, $x_i$ and $x_j$ denote pixels in $I_s$ and $I^k$, respectively. $I_s(x_i)$, $I^k(x_j)$ and $I_t(x_i)$ represent the intensities of pixels in images $I_s$, $I^k$ and $I_t$, respectively. $\mathbf{p}$ is a shift vector to indicate a pixel in a square window $W$. $D$ transfers a pixel position $x_i$ in $I_s$ to the corresponding pixel in $I_t$, $\omega_{x_i, k}$ is a reciprocal of the total of the power spectrum values that is larger than a threshold obtained by Fourier transforming the window region centered at $x_i$ in example image $I^k$. This term enables to select an example image with high-frequency components from the selected $T$ images, which often include motion blurs and defocuses even after the example images are narrowed down using the camera poses.

2.3.2. **Iterative energy minimization**

Energy function $E$ is minimized by iterating the following two processes: (i) search for similar textures in the selected example images and (ii) update pixel values in $I_s$.

In the process (i), we determine two parameters, pixel position $x_i$ and example image index $k$, by searching $T$ example images for the position $x_i$ around which the pattern is most similar to that around $x_i$ so that we minimize Eq. (2). The searching region is a certain range of $L \times L$ pixels around the coordinate $x_i$ in $T$ example images because the example images are roughly aligned by image warping. The searching compensates for the misalignment caused by geometric errors. The selected example image index and pixel corresponding to $x_i$ are represented by $n(x_i)$ and $f(x_i)$.

In the process (ii), pixel values $I_s(x_i)$ in the generated image are updated in parallel so as to minimize the energy function $E$ while keeping all the similar texture pairs fixed. Energy function $E$ is resolved into element energy $E(x_i)$ for each pixel $x_i$ in $I_s$:

$$E(x_i) = \sum_{\mathbf{p} \in W} \omega_{f(t)\cdot n(t)}(I_s(x_i) - I^e(t)(f(t) - \mathbf{p}))^2$$

$$+ \beta \{I_s(x_i) - I_t(Dx_i)\}^2,$$

$$t = x_i + \mathbf{p}.$$ \hspace{1cm} (4) (5)

$E$ can be minimized by minimizing each element energy $E(x_i)$ because energy $E$ consists of the sum of all element energies. $I_s(x_i)$ that minimizes $E(x_i)$ can be calculated by differentiating $E(x_i)$ with respect to $I_s(x_i)$, and is

$$I_s(x_i) = \frac{\sum_{\mathbf{p} \in W} \omega_{f(t)\cdot n(t)} I^e(t)(f(t) - \mathbf{p}) + \beta I_t(Dx_i)}{\sum_{\mathbf{p} \in W} \omega_{f(t)\cdot n(t)} + \beta}.$$ \hspace{1cm} (6)

### 3. EXPERIMENTS AND RESULTS

In this section, we evaluate the performance of the proposed method by subjectively comparing the results of the proposed method with those by the conventional methods. In the experiment, we captured two videos consisting of 60 frames with 640×480 pixels while moving a camera in indoor and outdoor environments as shown in Fig. 4. We selected one image from the input images as the target image, and the target resolution is set to $1280 \times 960$ (magnification factor is 2). We experimentally determined the parameters of the proposed method as: $W = 25 \times 25$, $\beta = 0.0005$, $L = 10$, and $T = 5$.

Figure 5 shows the results by bi-cubic interpolation (a), example-based method [9] (b) and the proposed method (c). From these results, we can confirm that the high-frequency components are successfully generated by the proposed method, and the results of the proposed method are much clearer than those of both bi-cubic interpolation and conventional example-based method [9]. However, in our method, we cannot enhance the resolution of the regions that have no depth values because of the limited area of reconstructed geometry. Therefore, the unnatural change in the resolution appears on the boundary of the enhanced and the other regions as shown in Fig. 6.

### 4. CONCLUSION

In this paper, we have proposed a method to increase the resolution of a low-resolution image using example images generated by a novel view synthesis technique using 3D geometry. Our contribution is to have less limitations on camera positions and geometry of the target scene than those in the conventional methods. Our experimental result has demonstrated that our proposed method successfully generate high-resolution images. In future work, we should attempt to enhance the resolution of the regions that have no depth values.

**Acknowledgment** This research was partially supported by JSPS KAKENHI Nos. 23240024 and 25540086.
Fig. 4. Examples of input images including a target image (upper left) and reference images.

Fig. 5. Experimental results

Fig. 6. Example of unnatural change in resolution caused by the missing of depth values in the target image
5. REFERENCES


