# AR Marker Hiding Based on
# Image Inpainting and Reflection of Illumination Changes

Norihiko Kawai      Masayoshi Yamasaki      Tomokazu Sato      Naokazu Yokoya*

Graduate School of Information Science, Nara Institute of Science and Technology, Japan

## ABSTRACT

This paper proposes a new method of diminished reality which removes AR markers from a user's view image in real time. To achieve natural marker hiding, assuming that an area around a marker is locally planar, the marker area in the first frame is inpainted using the rectified image to achieve high-quality inpainting. The unique inpainted texture is overlaid on the marker region in each frame according to camera motion for geometric consistency. Both global and local luminance changes around the marker are separately detected and reflected to the inpainted texture for photometric consistency.

**Index Terms:** H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

## 1 INTRODUCTION

In augmented reality (AR) applications, marker-based geometric registration is often used because of its robustness and easiness. However, such marker-based AR has the limitation that seamless fusion between virtual objects and the real world is not achieved due to the existence of markers. To solve this problem, diminished reality methods, which remove real objects by overlaying background texture onto the object regions, have been proposed [2, 3, 5]. Among the methods, we focus on an approach which generates textures on objects using image inpainting instead of capturing the real background. In the inpainting-based approach, three factors should be considered to achieve natural marker hiding; (1) naturalness of texture generated on a marker area, (2) geometric consistency between frames, (3) photometric consistency between the marker area and its surrounding. We describe conventional methods and the proposed method while focusing on the three factors.

Herling et al. [3] have applied an exemplar-based inpainting method to each frame of an input video. Just applying an exemplar-based inpainting method to an original input image often generates unnatural textures because of its weakness for perspective distortion. There exists another problem caused by inpainting for every frame; although the method tries to reduce texture flickering by temporally propagating patch correspondences for inpainting, unnatural geometric changes between frames are still observable. To overcome these problems, in the proposed method, perspective distortion is corrected for input of inpainting for natural texture generation (1), and the unique inpainted texture is used in all the frames without trying to re-inpaint for temporal geometric consistency (2),

On the other hand, Korkalo et al. [5] have focused on photometric consistency in diminished reality. The method generates a low-resolution texture for a current frame and combine it with high-resolution texture inpainted using the first frame. While this method can treat only global luminance changes, the proposed method considers both global and local ones for photometric consistency (3).

---

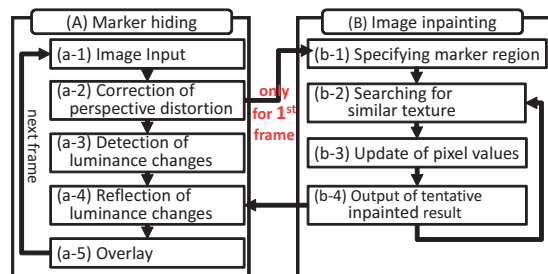*e-mail:{norihi-k, tomoka-s, yokoya}@is.naist.jp
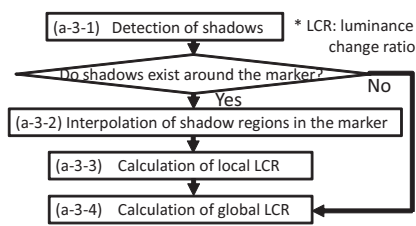
Figure 1: Flow of the proposed method.
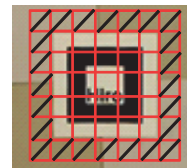


Figure 2: Detailed steps in step (a-3).



Figure 3: Divided grids.

## 2 REAL-TIME AR MARKER HIDING

Figure 1 shows the proposed pipeline. The method consists of two processes: (A) marker hiding by reflecting luminance changes and overlaying the inpainted textures on markers for every frame, and (B) image inpainting for marker areas. Although process (B) is not performed in real-time for generating high-quality texture, users can use the AR system without waiting time by performing the two processes concurrently. In process (B), we employ exemplar-based image inpainting method [4] with searching method [1]. The inpainting method is applied to the first frame image of which the perspective distortion is corrected. In process (A), after inputting an image (a-1), the image is rectified in step (a-2). By correcting the perspective distortion, the size and shape of a texture pattern are unified. That results in generating high-quality texture by image inpainting in process (B). In addition, pixel correspondences between different frames of input images can be easily determined.

### 2.1 Detection and reflection of luminance changes

The luminance of the texture generated by process (B) is adjusted by considering local luminance changes (e.g. by cast shadows) and global luminance changes (e.g. by ambient illumination changes and camera parameter changes). It should be noted that we do not directly use the amplitude of luminance changes in the marker but estimate them from the marker's surrounding because photometric properties of markers and the surrounding materials are different.

Step (a-3) calculates local luminance change ratio $\beta$ for shadow region and global one $\alpha_i$ for each pixel $\mathbf{x}_i$ in the remainder of the marker. The ratios of RGB are separately calculated. Then step (a-4) adjusts the luminance of the texture generated by process (B) by multiplying $\beta$ to the interpolated shadow region and $\alpha_i$ to the remainder. Figure 2 shows the detailed steps of step (a-3). In these steps, a marker and the surrounding region $\Psi$ (grid with a diagonal line in Figure 3) are divided into grids for efficient processing.
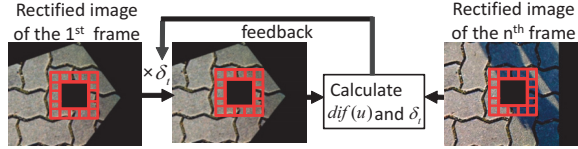
Figure 4: Detection of cast shadows.

### 2.1.1 Calculation of local luminance change ratio

In step (a-3-1), we assume that shadows do not exist in the first frame and the intensities of shadow regions are much smaller than the other regions. As shown in Figure 4, provisional global luminance change ratio $\delta_t$ for the regions in which cast shadows do not exist in $\Psi$ is first calculated in iterative manner as follows:

$$\delta_t = \sum_{u\in\Psi} \omega(u)\bar{I}_n(u) / \sum_{u\in\Psi} \omega(u)\bar{I}_1(u), \qquad (1)$$

where $\bar{I}_n(u)$ is the average of pixel values in grid $u$ in the $n$-th frame rectified image. $t$ is number of iterations. We define $\omega(u)$ as $1/(\mathrm{dif}(u)+1)$ where

$$\mathrm{dif}(u) = max(\delta_{t-1}\bar{I}_1(u) - \bar{I}_n(u), 0), \qquad (2)$$

because weight $\omega(u)$ should be small in non-shadow regions. After several iterations, the grid whose dif value by Eq. (2) is larger than a threshold is determined as cast shadow region. The above processes are then performed pixel-wise for only two grids which include the shadow boundary, and each pixel is classified into shadow and non-shadow regions.

In step (a-3-2), the shadow boundaries in region $\Psi$ are first detected. To decide pairs of the boundaries, cost $|\theta_i - \theta_{ij}| + |\theta_j - \theta_{ij}|$ is used. Here $\theta_{ij}$ is the gradient direction of shadow boundary in connecting pixel $\mathbf{x}_i$ and $\mathbf{x}_j$. $\theta_i$ and $\theta_j$ are the gradient directions of shadow at pixel $\mathbf{x}_i$ and $\mathbf{x}_j$. The pair which gives the smallest cost is decided, and after removing the pair, a next pair is decided. This process is iterated until all pairs are decided, and the area surrounded by shadow is set as shadow region. Local luminance change ratio $\beta$ is determined by the ratio of sum of pixel values in the shadow region in region $\Psi$ between the first and current frames.

### 2.1.2 Calculation of global luminance change ratio

In step (a-3-4), luminance change ratio $\hat{\alpha}_u$ in each grid $u$ in region $\Psi$ is first calculated with the ratio of sum of pixel values in a grid of the first and current frames. To exclude the influence of cast shadows, ratio $\hat{\alpha}_u$ of the grid $u$ which includes cast shadows in region $\Psi$ is reset by linear interpolation using change ratios of non-shadow grids. $\hat{\alpha}$ of each grid in the marker is then calculated so as to minimize cost $E$.

$$E = \sum_{(u,v)\in R} (\hat{\alpha}_u - \hat{\alpha}_v)^2, \qquad (3)$$

where $u$ and $v$ are indexes of grids. $R$ is a set of pairs of grids indexes which are in eight neighbor. Change ratio $\alpha_i$ for each pixel $\mathbf{x}_i$ in the marker is then calculated by bilinear interpolation.

## 3 EXPERIMENTS

We performed experiments by putting markers on several environments. In experiments, we used a PC with Core i7 3.47 GHz CPU and 12 GB of memory for input images with resolution $640 \times 480$ captured by a USB camera with automatic white balance and gain.

We compared results by four approaches using a sequence including cast shadows. In the results by inpainting without rectification in Figure 5(b), unnatural textures are generated due to the perspective distortion. The results in Figures 5(c), (d) and (e) are generated with rectification and different reflections methods. Without reflection of luminance changes, photometric consistency is not achieved as shown in Figure 5(c). Reflecting only global



(a) Input Image     (b) No rectification     (c) No reflection



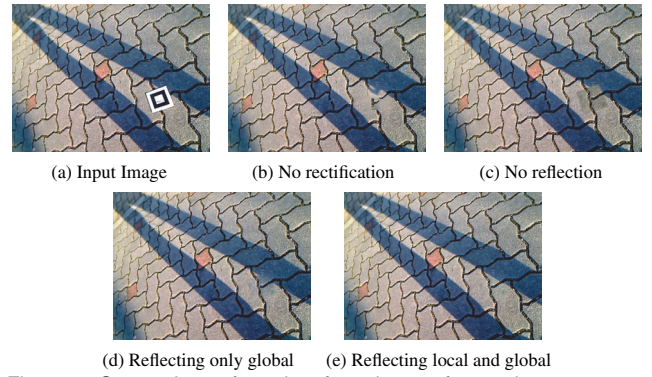(d) Reflecting only global     (e) Reflecting local and global

Figure 5: Comparison of results of continuous frames in a sequence including cast shadows.
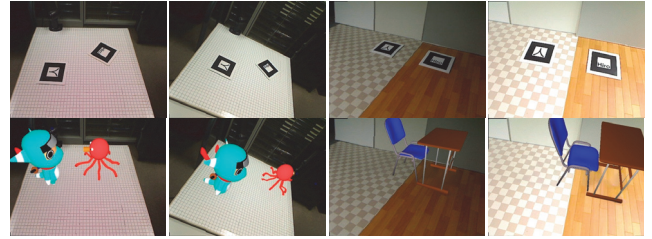


Figure 6: Examples of AR applications using the proposed marker hiding (Top: input images with markers, bottom: output AR images).

luminance changes gives blurry shadow edges as shown in Figure 5(d). Reflecting both local and global changes, clear cast shadows are successfully given as shown in Figure 5(e). The frame rate was 18.8 fps and the processing time for image inpainting in process (B) was eight seconds. Therefore, we can experience AR applications with completely inpainted results in about eight seconds after starting.

Figure 6 shows prototype AR systems using the proposed method considering scenarios of hiding markers in playing games and simulating furniture arrangement. We can enjoy AR applications without being concerned about the existence of markers.

## 4 CONCLUSION

We have proposed a new diminished reality method for removing AR markers in real time. Correction of the perspective distortion of images enables high-quality image inpainting. Geometric and photometric consistencies are achieved by reflecting both local and global luminance changes separately and overlaying the texture on the marker according to camera motion. Future work involves developing a marker hiding method for a non-planer background.

### REFERENCES

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patch-Match: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):1–11, 2009.

[2] A. Enomoto and H. Saito. Diminished reality using multiple handheld cameras. In *Proc. ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing*, pages 130–135, 2007.

[3] J. Herling and W. Broll. Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In *Proc. ISMAR*, pages 207–212, 2010.

[4] N. Kawai, T. Sato, and N. Yokoya. Image inpainting considering brightness change and spatial locality of textures and its evaluation. In *Proc. PSIVT*, pages 271–282, 2009.

[5] O. Korkalo, M. Aittala, and S. Siltanen. Light-weight marker hiding for augmented reality. In *Proc. ISMAR*, pages 247–248, 2010.