Exemplar based texture recovery technique for active one shot scan

Thibault Yohan, Kawasaki Hiroshi
University of Kagoshima, Department of Engineering
Kagoshima, Japan
y.thibault,kawasaki@ibe.kagoshima-u.ac.jp

Furukawa Ryo
N. Institute of Advanced Industrial Science and Technology
Ibaraki, Japan
ryo-f@hiroshima-cu.ac.jp

Sagawa Ryusuke
Faculty of Information Sciences, Hiroshima City University
Hiroshima, Japan

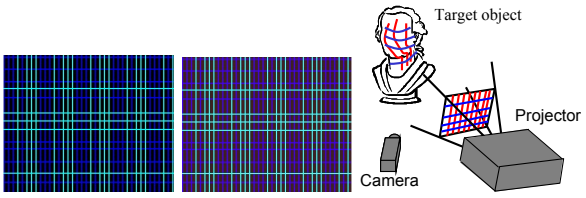
Abstract

Range scanners based on camera-projector configurations have been attracting the attention of many researchers and computer vision practitioners due to its ability of performing 3D scanning with high accuracy and frame rates. This class of scanners is more effective under dark capturing environments by projecting bright colored patterns on objects. As a consequence, the projected pattern is prone to interfere with the reflectance and texture characteristics of the surfaces of the objects. This paper introduces a novel technique for recovering texture information from objects obfuscated by projected patterns. To that end, prior to 3D scanning our method acquires a small amount of pairs of images of the object with and without the projected pattern. These pairs of images are then used to build a dictionary that establishes a relationship between the actual textures and the projected patterns. Following that, 3D scanning is performed normally using a camera-projector setup, and the video frames captured by the camera are matched against the dictionary to filter candidates that resemble the original textures. Regularization is applied on the potential set of candidates to cull out poorly detected one, resulting in smooth texture-aware images.

1 Introduction

Dense shape acquisition of moving objects is a rich research topic, with features that are attractive to many fields such as medical imaging, video games, to cite a few. Recently, active 3D scanning systems have received special attention because of their accuracy and fidelity. These methods perform dense 3D reconstruction of animated objects by projecting a single pattern onto objects, which are then captured by a camera and the resulting video frames are analyzed and processed. Typically, infrared light [3, 6] or visible light [2, 5, 8, 4] is used for the projected pattern. Infrared light is advantageous since the human visual system is unaware of it. However, most readily available light sources work in the visible spectrum; the same applies to image sensors. Therefore, general active scanners, especially for inspection purpose, are bound to visible light. On the other hand, visible light is likely to conceal relevant texture information from the scanned object, further complicating a faithful 3D reconstruction.

In this paper, we propose a simple and effective method removing the projected pattern of active 3D scanning systems operating on visible light. It infers an original textures of the objects that were hidden by the projected pattern.Our approach adopts an exemplar-based strategy to achieve this goal: a dictionary containing two sets of images is compiled; one with and another without the projected pattern on the object. The frames captured during the regular scanning process are, in a first step, matched against the dictionary to select an initial set of potential candidates for subregions of a particular frame. Finally, a belief propagation algorithm, abbreviated as BP hereafter, is applied to qualify the most fitting candidate for each subregion in order to synthesize a smooth texture-recovered image.



(a) (b) (c)

Figure 1. (a) The regular pattern for lines detection [5], (b) The modified pattern with brighter background, (c)Scanning system: multiple lines are projected and their intersections are detected and used for reconstruction.

The remainder of this paper is structured as follows: Section 2 is an overview of the proposed method and the system configuration; Section 3 details each step of the method; Section 4 shows some experiments and a discussion. Finally, Section 5 concludes the paper.

2 Overview

2.1 System configuration

Our technique operates on the raw output of a 3D measurement system consisting of a camera and a projector as shown in Fig. 1(c). Both the camera and the projector are assumed to be calibrated (i.e., the intrinsic parameters of the devices and their relative positions and orientations are known beforehand). The projected pattern remains fixed throughout the entire scanning process so that no device synchronization is required. A grid pattern of vertical and horizontal lines is extruded from the projector and recorded by the camera, as depicted in Fig. 1(a).

For 3D reconstruction, we use the technique proposed by Sagawa et al. [5]. In their method, first the

projected grid pattern is extracted from the recorded image frame. Note that the straight lines of the pattern may appear curved in the recorded images depending on the geometric characteristics of the scanned object. For stable grid pattern detection, BP and de Bruijn sequences are used. The intersections of the detected čurvesäre referred to as grid points, from which simple linear constraints about the crossing pattern planes can be determined. Finally, by solving a set of linear equations established from the grid points, solutions for the pattern planes are obtained. Since there exists an 1-DOF ambiguity in the solution, a final decision is made by matching the actual positions of the pattern planes with those of estimated planes, ultimately resulting in reconstructed 3D curves.

2.2 Overview of the proposed method

Prior to the texture recovery process, a dictionary of images has to be compiled, consisting of two sets of images: the first set holds snapshots of the object lit with the projected pattern in a dark environment, while the other set holds their equivalent snapshots under uniform white light projection. Following that, video frames output by the scanning process can be processed individually.

In a first phase, the frame of interest is subdivided into smaller regions of a given constant size and each subregion is treated independently. Assisted by the set of "pattern-projected images" of the dictionary, a potential subset of candidates is chosen for each subregion. Each of the selected candidates has the counterparts with "clean" texture information associated with them in the "other half" of the dictionary. The rationale is that since pattern-projected candidates are expected to share similarities with the subregion being texture-recovered, it is likely that their associated texture-aware analogues also share similarities with the actual texture of the subregion being concealed by the projected pattern.

However, some candidates of the potential set are better suits than others, and a poorly chosen candidate is prone to spoiling the texture reconstruction. Therefore, in the second phase of our technique, we employ a BP algorithm over the lists of candidates to prune less fitting candidates. In the end, each subregion is left with a single best candidate to shade itself. This yields to make smooth texture recovered images with reduced noise and artifacts.

3 Recovering textures from pattern-projected image

This section details the proposed solution to restore texture data from images polluted by projected patterns in projector-camera scanning systems, as briefly introduced in Section 2.2.

The following subsection, Section 3.1, describes the process of compiling the dictionary. Following that, Section 3.2 explains how the initial potential set of candidates is obtained. Section 3.3 then describes how the BP algorithm can be tailored to our problem for selecting favorable patches.

3.1 Acquisition of the data and creation of the dictionary

Creating the dictionary requires one or several pairs of images of the object to be captured: images with white projected light, and their counterparts with the projected pattern, as illustrated on Fig. 2(a). These pairs of images are obtained during a training session. In the case of still objects, these pairs of images can be easily obtained by switching on and off the projected pattern. For dynamic entities (e.g., a human actor), extra caution must be taken since even small disturbances can lead to artifacts in the final result, thus proper image alignment must be ensured.

Once the pairs of images are acquired, the dictionary is assembled by processing each pair of images in batches. For each corresponding pair, pixels containing valid projected pattern information, i.e. pixels that do not belong to the "black" background, are selected from the image, along with a small neighborhood of pixels around them. Hereafter, we refer to these small squared subregions as *patches*, whose sizes are typically 3×3 or 9×9 pixels. For each such patches, the same region is retrieved in the corresponding white light textured image and a logical link is established between them. The final dictionary thus contains two lists of patches, $\mathcal{L}_1$ holding the patches with projected pattern information and $\mathcal{L}_2$ with patches holding texture information, and a bijective mapping between them.

A simple extension consists of cloning patches using rotation factors. Since the size of patches is allegedly small, only rotations by 90°, 180° and 270° could be considered. However, this ultimately increases the overall computational time, storage requirements and search time of the dictionary.

3.2 Selection of candidates

Once the dictionary is available, we can begin the process of recovering textures for each frame recorded during the scanning session. Given one such recorded frame $\mathcal{I}$ spoiled by the projected pattern and the dictionary obtained in the previous section, the process begins by subdividing $\mathcal{I}$ into smaller subregions whose sizes coincide with that of the patches within the dictionary. These subregions then become queries against the dictionary.

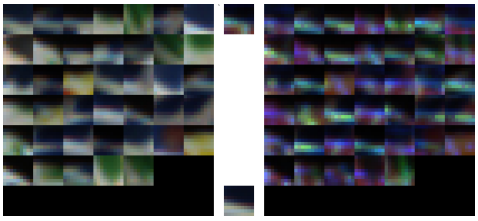
The dictionary lookup algorithm retrieves a list of n patches in $\mathcal{L}_1$ that better resemble a given query subregion $q \in \mathcal{I}$. The search is linear, and the similarity between a patch $p \in \mathcal{L}_1$ and a subregion $q \in \mathcal{I}$ is evaluated using the cumulative pixel-wise squared color distance between them (in the linear RGB color space), or more formally:

$$D(p,q) = \sum_{p_i \in p, q_j \in q} d(c,i,j) \tag{1}$$

$$d(c,i,j) = |I_{p_i}(c) - I_{q_j}(c)|^2, \; c \in \; R,G,B \tag{2}$$

where p_i and q_i are the pixels in p and q, respectively. The lower the outcome of $D(p,q)$ is, the more similar q and p are; zero would be a perfect match. Fig. 2 illustrates the result of the first step. The patch in the top-center represents an area of $\mathcal{I}$ to be recovered, with the ground truth underneath it. The set of patches in the right are the n most similar patches extracted from $\mathcal{L}_1$ (sorted in ascending order of $D(p,q)$) and the patches in the left are their textured counterparts in $\mathcal{L}_2$. In the ideal case, the first patch in the leftmost set should be the same as the ground truth.

At this point, it is already possible to infer texture information for each subregion $q \in \mathcal{I}$ by using the patch with the smallest distance and "blitting" the subregion



(a) (b) (c)

Figure 2. A list of patch to recover a given part of the image. (middle top) patch to recover, (middle bottom) ground truth, (right) selected patches from $\mathcal{L}_2$ and (left) equivalent of (right) from $\mathcal{L}_1$.

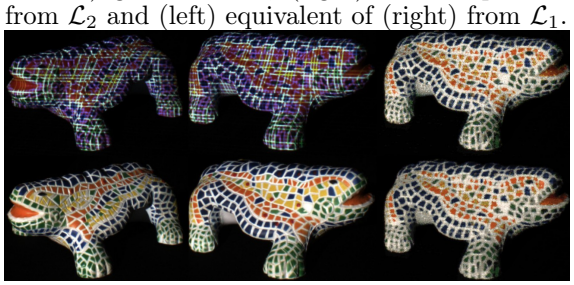


Figure 3. Recovery of a lizard inspired by Gaudí's sculptures.(left) Input image for the dictionary, (middle) Image to recover and the ground truth and, (right) Recovered image before BP (top) and after BP (bottom).

q with the appropriated textured patch in $\mathcal{L}_2$; the result is depicted in Fig. 4(bottom left). Although the result is close to the ground truth in a coarser scale, the synthesized textured image still has noticeable noise and banding artifacts that can be further mitigated.

3.3 Regularization using belief propagation

In the next step the best patches are found among the list of the selected n patterns for each area of $\mathcal{I}$. Fig. 2(a) shows that most candidates contain the proper texture and only few are poorly selected. The best candidate for each area is then defined as the most consistent patch within the list of candidate's in the neighboring areas. The goal is to produce a smooth output where borders between areas are not visible. This problem is a Markov Random field problem and can be solved using a BP algorithm. BP is well-known, thus the idea of this method will not be recalled here but can be found in [1, 7]. BP starts with an initial solution and is mainly oriented by the data cost and the regularization cost. The data cost represents the cost to be replaced, in a given area, the patch selected as initial solution by another candidate. Because the initial solution is noisy, the data cost should have a low influence and let the regularization cost lead BP algorithm. The regularization cost between two patches in adjacent area is defined as the distance between their adjacent pixels. The distance is computed by using equations similar to equations (1) and (2). The typical use of BP requires a large number of iteration to spread the messages all over the images. For our case, experiments show that after 20 to 30 iterations, the minimum is reached. The final output is created by using the patch that minimizes the regularization of each area after k iterations.

4 Experiments

This section presents the results of our method through two types of experiments. Experiments with static objects are done to evaluate the quality of the output and the effectiveness of our method on various textures. Experiments with moving objects captured by a high speed camera shows the effectiveness of our method on video data.

All experiments are done in a dark environment where the only light is provided by the projector. The creation of the dictionary is done by using two pictures of a given subject: one with the projection of white light and the other one by projection of the pattern presented in Figure1(b). Images to be recovered are acquired after the one for the dictionary using similar lightning condition.

4.1 A new projected pattern

The regular pattern presented in Fig. 1(a), which consists of only blue and green color, allows an accurate detection of dense pattern projected on colorful texture for shape reconstruction. However, since the red channel is not used in the pattern, it is problematic for the texture recovery algorithm. For example, without any red channel information, green and yellow textures become identical. Thus a new pattern containing a dark red background is used (Fig. 1(b)) to replace the original one (Fig. 1(a)).

4.2 Experiments using static objects

Fig. 3 illustrates the result of our method. The two left images are used to create the dictionary; the image in the middle top is the one to be recovered and its ground truth (bottom). Right images illustrate our result before BP (top) and after BP (bottom). Fig. 4 illustrates another result with the similar algorithm.

Fig. 3 and 4 show that the proposed method is effective and gives good results. BP algorithm smooth the initialization's result. The mouth of the plastic toy is noisy and contains a lot of poorly selected patches before regularization. This kind of noise appears mostly with red and yellow colors since they are missing in the projected pattern. Similarly with the lizard, yellow tiles are often recovered as orange tiles. The small size of these tiles makes a proper recovery difficult. The small demarcation of the eyes of plastic toy and its mustache are also poorly recovered. When the size of a detail is smaller than the patches' size, their recovery usually fails, especially after application of BP.

In the blue area of plastic toy's head, the lines of the pattern are still visible. This effect results the lack of color's information between line's pattern and leads to the selection of darker candidates. Thus, application of BP cannot smooth the first result and remove these lines since the proper patches are not selected.

Fig. 5 illustrates the result of the single pass method applied to other 3D active scanner patterns. The presented method was originally designed for the 3D scanner presented by Sagawa et al. in [5]. However, it was designed to be compatible with other 3D scanning systems such as [2, 8, 4] illustrated in Fig. 5(c), (d) and (e) respectively. All the results are visually convincing but slightly worse than the results obtained with the pattern specially adapted for our method (Fig. 5(b)). However, it proves that the presented method can be applied to several one shot scanning techniques.

4.3 Evaluation using moving objects

The main purpose of our method is for objects in motion. Fig. 6 illustrates the reconstruction of a single

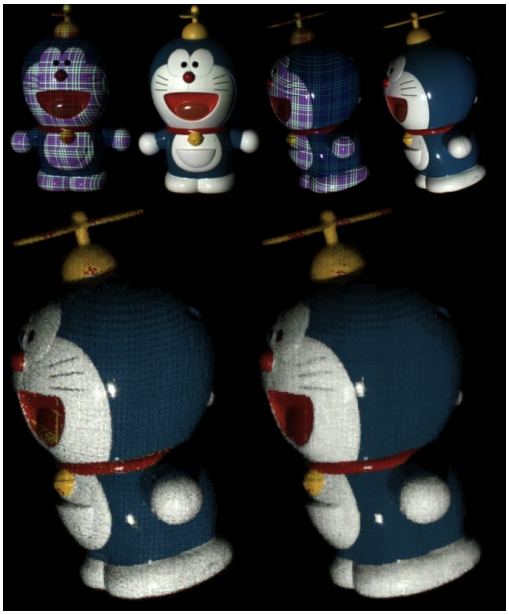
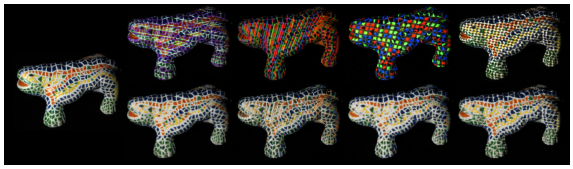


Figure 4. Recovery of a plastic toy.(top left) Input image for the dictionary, (top right) Image to recover and the ground truth and, (bottom) Recovered image before BP (left) and after BP (right).



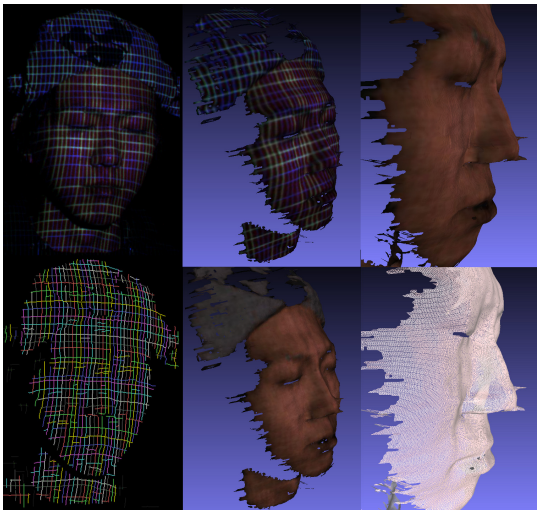
(a) (b) (c) (d) (e)

Figure 5. Recovery of the lizard using different projected pattern. (a) ground truth, (b) modified pattern originally proposed in [5], (c) pattern proposed in [2], (d) pattern proposed in [8], (e) pattern proposed in [4].

frame extracted from a video of a human face doing grimaces captured with high speed camera set at 200 fps. Unlike the static object, human cannot stay perfectly still during the capture of images for the dictionary. These small differences produce some artifacts in the results. Fig. 6(a) is the input image (top) and its detected lines (bottom). Fig. 6(b) and (c) illustrate the reconstruction results. Fig. 6(b)-(top) is the result with the original image and (c)-(top) is without texture. Fig. 6(b) and (c) (bottom) is the result with the texture recovered with our method. Both, shape reconstruction and texture recovery, give good results. Fig. 6(c) top shows that the modified pattern still allows a precise reconstruction and Fig. 6(b) and (c) bottom show that details such as lips color or eyebrow can be recovered.

5 Conclusion

In this paper, a solution to recover the texture for active 3D scanning systems has been presented. We applied the efficient solution derived from the exemplar based method to effectively recover the dense 3D shape and convincing textures even for high speed motion by using a simple pair of images acquired before or after the data acquisition. Our solution can be utilized to offer the less expensive and more effective alternative



(a) (b) (c)

Figure 6. Recovery of a human face including 3D reconstruction. (a) input image and the result of the line detection[5], (b) 3D reconstruction with the input image and with the recovered textures, (c) another point of view with and without the texture.

to the widely used system like kinect with more dense and precise results. Our future work is a extention to temporal direction to achieve smooth texture recovery on video.

References

[1] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *Int. J. Comput. Vision*, 70(1):41–54, 2006.

[2] Changsoo Je, Sang Wook Lee, and Rae-Hong Park. High-contrast color stripe pattern for rapid structured-light range imaging. In *EECCV*, pages 95–107, 2004.

[3] Microsoft. Xbox 360 Kinect, 2010. http://www.xbox.com/en-US/kinect.

[4] Sagawa Ryusuke, Sakashita Kazuhiro, Kasuya Nozomu, Kawasaki Hiroshi, Furukawa Ryo, and Yagi Yasushi. Grid-based active stereo with single-colored wave pattern for dense one-shot 3d scan. In *3DIMPVT*, pages 363–370, 2012.

[5] Sagawa Ryusuke, Ota Yuichi, Yagi Yasushi, Furukawa Ryo, Asada Naoki, and Kawasaki Hiroshi. Dense 3d reconstruction method using a single pattern for fast moving object. In *ICCV*, 2009.

[6] Kazuhiro Sakashita, Yasushi Yagi, Ryusuke Sagawa, Ryo Furukawa, and Hiroshi Kawasaki. A system for capturing textured 3d shapes based on one-shot grid pattern with multi-band camera and infrared projector. In *3DIMPVT*, pages 49–56. IEEE Computer Society, 2011.

[7] Yair Weiss and William T. Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory*, 47(2):736–744, 2001.

[8] L. Zhang, B. Curless, and S. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *3DPVT*, pages 24–36, 2002.