MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Entire shape scan system with multiple pro-cams using texture information and accurate silhouette creating technique

Yoshinori Oki
Kagoshima University

Marco Visentini-Scarzanella
Kagoshima University

Tomohito Wada
National Institute of Fitness and Sports

Ryo Furukawa
Hiroshima City University

Ryusuke Sagawa
AIST

Hiroshi Kawasaki
Kagoshima University

Abstract

Recently, many approaches have been proposed for complete 3D reconstruction of moving objects. Particularly, one-shot reconstruction using a projector-camera system (pro-cam) has attracted considerable attention as it can reconstruct the full 3D shape from a single image and it is therefore suitable for capturing moving objects. However, one-shot measurement methods have the issue that the projection pattern is easily affected by the texture of the target object and its background. Furthermore, background subtraction can be inaccurate due to the shadow cast by the target object. In this paper, we propose a method to solve these issues using texture estimation and online background synthesis techniques, by relying on a dictionary of pre-captured images. By applying these methods to our multi-view pro-cams, we can demonstrate the robustness of our approach for one-shot reconstruction of moving objects, with reduced sensitivity to texture and cast shadows.

1 Introduction

Recently, dense shape acquisition has seen an increase in popularity due to its applications in many fields, and various techniques have been proposed in the literature, including Shape-from-Silhouette, Multi-view Stereo and active techniques. Shape from Silhouette [3] is one of the typical methods to acquire the entire shape in dynamic scenes, however surface details usually cannot be efficiently recovered.

For static scenes, Multi-View Stereo (MVS) [2] can generally yield accurate reconstructions, however its reliance on feature matching results in unreliable results whenever textureless objects are considered.

On the other hand, active 3D scanning systems are not influenced by the lack of distinctive features. However, the texture on the object can still negatively affect the reconstruction, since the projected pattern may be absorbed by the object texture and because of cast shadows, thus resulting in decoding issues.

In this paper, we propose a multi-view active stereo scanning system which adapts to the object's texture characteristics for increased robustness. Specifically, the pattern is first projected on the object, from which the object texture is recovered through the Texture Recovery technique proposed in [7]. Finally, by compensating the color information of the projected pattern with the recovered object texture information, we are able to acquire the shape of object with complex texture patterns.

Another severe issue for active multi-view stereo system is that the projected light influences the other non-target objects (*e.g.*, wall, ground, etc.), which results in problems for precisely defining the object silhouette. To solve this, we propose an accurate silhouette creating technique, using a database of backgrounds with projected structured light patterns.

Without the two techniques, our active multi-view acquisition system was able to operate only in dark room conditions with no external interference, however, following the integration of the techniques, it was possible to achieve measurements even under natural ambient light conditions, which drastically facilitate the scanning. The experiments show good reconstruction quality using six cameras and six projectors even in a brightly lit room.

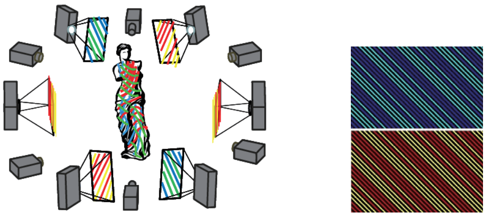
2 System configuration and basic algorithm of multi-view projector camera system

For our proposed method, we use the active MVS system proposed by Furukawa *et al.* [1], where multiple cameras and projectors are used. In the setup, devices are placed so that they encircle the target scene, and the cameras and the projectors are put in alternating order with known position and orientation. The projected pattern is static, thus, no synchronization is necessary between the projectors and cameras. An example of the system setup is shown in Fig 1(a).

In order to capture the geometry of the scene, each of the projectors projects parallel single directional lines (shown in Fig 1(b)) on the object, while the cameras capture the projected line patterns as 2D curves on the captured images. Since multiple patterns are overlapped on the object, each of the detected curve is identified to the projector with which the pattern is generated by using color and curve orientation information. This enables to determine the correspondences between the detected curves and the projected line-patterns, and consequently to obtain a 3D reconstruction of the scene.

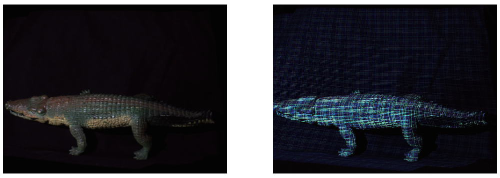
3 Color compensation by texture information

We first focus on reducing the effect of the overlapping between the projected patterns and the underlying object texture by estimating the latter from the acquired images. Because of the overlaid pattern from



(a)System configuration (b)Projection pattern

Figure 1. A setup example to reconstruct the entire shape using six projectors and six cameras.



(a)Texture of object (b)Projected pattern image

Figure 2. Capture images.

the active 3D system, the underlying texture cannot be observed directly; therefore, we use the texture estimation method proposed by Thibault *et al.* [7]. Having gained information on the object texture, we can correct the perceived color of the projected pattern.

3.1 Texture Recovery method using database

In this section, we briefly explain the texture recovery technique employed in our approach [7]. This is an exemplar-based technique: the first set S_1 holds snapshots of the object lit with the projected pattern in a dark environment, while the other set S_2 holds their equivalent snapshots under uniform white light projection. A dictionary is then created consisting of these images.

For recovery, the target image is subdivided into smaller regions of a predefined size, with each subregion treated independently. Assisted by the the images in S_1, a potential subset of candidates with high correlation is chosen for each subregion, where each of the selected candidates has a counterpart with "clean" texture information associated in the S_2. In order to avoid mismatches which can ruin the texture recovery, we select the best fitting matches within a BP framework, yielding a smooth texture map with reduced noise and artifacts between subregions.

3.2 Correction of capture image using texture

In this section, we propose a method to reduce the effect of the interaction between the structured light pattern and the texture of the target object. As an example we show in Fig 2 and Fig 3, the projected pattern the resulting 3D reconstruction are affected by the object texture, respectively. As shown in the example, the reconstruction area is decreased significantly because of the effect from the object texture (Fig 3Red circle). We therefore correct each pixel of the captured

Figure 3. Reconstruction result.

image using the following equation:

$$O(i,j) = \frac{P}{T(i,j)} C(i,j), \tag{1}$$

where $O(i,j)$ is the pixel value after RGB compensationCP is the original RGB value of the pattern that was projected, while $T(i,j)$ is the estimated RGB value of the texture RGBCand $C(i,j)$ is the RGB value of the structured pattern in the captured image, for image pixels i, j. While the SN ratio did not improve following this linear transformation, the detection accuracy did improve for the following two reasons:

First, there is an improvement of the detection accuracy of the line from the projected pattern. In the pattern detection algorithm used in this method, every pixel is labeled as Positive(P), No change(0) or Negative(N), depending on the derivative of the luminance values, and a label change from N to P or P to N, is detected as a line with sub-pixel accuracy [5]. Often in areas of dark texture, light absorption from the texture causes line misdetections. Following the application of our proposed method, it is possible to successfully detect the projected pattern lines. The RGB compensation scheme allows to correctly label each pixel. However, the noise is also amplified with this technique, hence we rely on our reconstruction with silhouette constraints in order to remove most of the noise present.

Secondly, there is a marked improvement on the detection accuracy of each line ID. The reconstruction algorithm employed uses a De Bruijn ID pattern in order to prevent false detection, using two colors and three window sizes. Because of the dependency on color information, the object texture can result in incorrectly perceived color which may result in decoding failures. With the proposed method, we are able to successfully separate the projected pattern from the object texture, resulting in correct decoding of the De Bruijn ID.

4 Creating silhouette method based on database

In general, in order to be able to successfully define the object silhouette, it is necessary to separate the image of the target object from its background. However, in our 3D scanning environment, the 6 projectors are all active at the same time, thus projecting their pattern on the surrounding background, which will be affected by the shadows cast by the object (Fig. 4 and Fig. 5). Therefore, simple background subtraction will not work in our scenario. We therefore propose a new silhouette creation technique to solve this problem. The basic idea of our technique is to create all possibilities for the image areas affected by cast shadows as shown in Fig. 6. Therefore, we capture multiple images of the background illuminated by different subsets

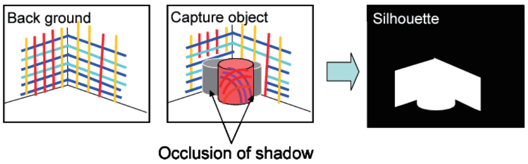


Figure 4. Reason of silhouette creation failure

(a) (b) (c)

Figure 5. Example of silhouette failure (a) Background image, (b) Input image, (c) Silhouette

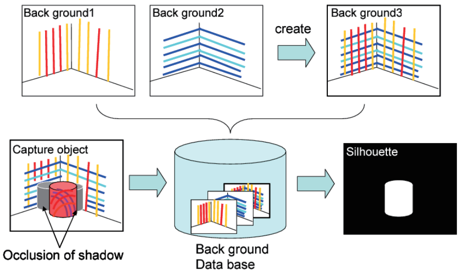


Figure 6. Workflow for the proposed technique.

of the projectors present in the scene. All these images are put in a database, which is used to calculate the similarity with each region of the target image, thus allowing to find correspondences of background regions even in the presence of cast shadows.

We start by capturing images of the background with the projected pattern from each projector. Given a setup with N projectors, we capture N images. Secondly, we generate combined images in which subsets of the N original images are superimposed. This results in a total number of combined images of $2^N - 1$, which constitute our database.

Finally, during the shape acquisition procedure, similarity is calculated for each patch of the object image against the database using ZNCC (Zero-mean Normalized Cross-Correlation), which allows for linear changes in luminance. However, whenever both cast shadows and clean background are present in the candidate patch at the same time, the similarity score is negatively affected. Similarly, at the boundary of the projected lines the perceived color is affected by the demosaicing algorithm, creating instabilities and false matches. In order to alleviate these issues, we consider directly using Bayer pattern images as the input and use an adaptive support-weight [8] when calculating the similarity. Adaptive support-weights work by setting a weight to each pixel of the patch based on color similarity between the center and outer pixels, which we have found experimentally to solve our problem. The patch size is determined depending on the projected pattern line on the captured image. If the patch size is too small, it is impossible to obtain discriminative features, while if the patch size is too large, it can include many areas where multiple shadows are present. In all our experiments, we use a 5x5 patch size. It is shown that there was the region similar to the background image. We use this similarity as the data section of the graph cut, to separate the background area and the foreground area.

Table 1. Evaluation results in synthetic data.

	RMSE[*m*]	*Number of points*
previous method	0.001849	27,745
proposed method	0.001427	90,919

5 Experimental Results

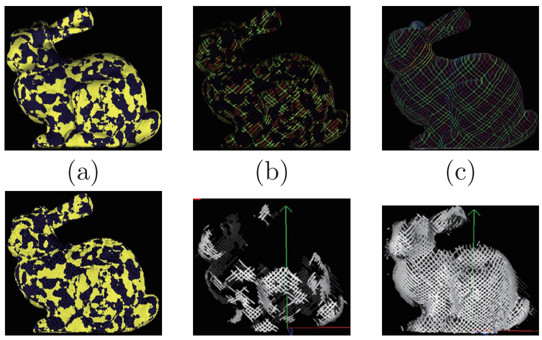
5.1 Synthetic data experiment

Given the Stanford bunny dataset, we generate six virtual cameras and projectors surrounding the target object and render the captured images using POV-Ray [4] for a physically accurate rendering.The result is shown in Fig 7, which shows our performance in correcting the projected pattern on the target object successfully. In this case, PSNR is 23.05dB, reconstruction result evaluated shown in Table 1.

5.2 Real data experiment

In our experiment, we use six Point Grey Research Grasshopper cameras (1600x1200 pixels), and six LCD video projectors at WXGA resolution. The cameras are synchronized and calibrated prior to the experiment and can capture images at 20 fps.

For our experiment, we prepare an object with complex colored texture, onto which a gray code ID is projected from each projector. We evaluate the reconstruction accuracy by calculating RMSE of reconstructed points. The results for our improved technique of color correction and silhouette detection are shown in Fig. 8 and 9. The first figures show that the projected color is successfully compensated with recovered texture of object. The latter figures show that



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

Figure 7. Effect of our technique. (a) Texture image. (b) Projected pattern image. (c) After correction. (d) Estimated texture. (e,f) Reconstructed image of (b,c).

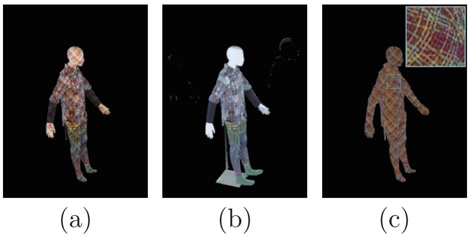


Figure 8. Color correction: (a) Captured image. (b) Estimated texture. (c) After correction.

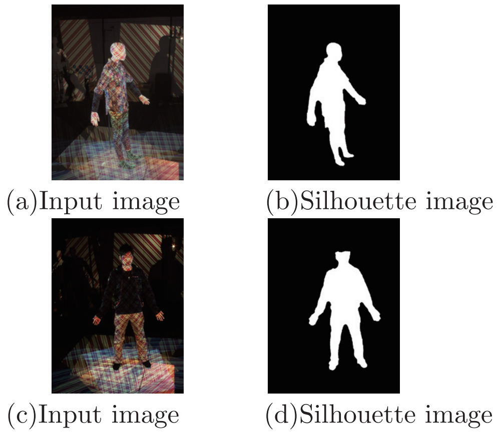


Figure 9. Silhouette creation results.

Table 2. Evaluation results.

	RMSE[m]	*Number of points*
previous method	0.008214	136,871
proposed method	0.008241	176,406

the wall and floor are successfully recognized as background and foreground, despite the complexity of the projected pattern on the wall and floor, and with the several cast shadows from the object present. However, there are some cases of silhouette detection failure as shown in Fig 9. In this case, there is a high correlation between the dark color of the head and the wall, which results in mismatches. An improvement for such failure is our future work. In Fig 10, we show the qualitative results of our entire work-flow, while numerical accuracy results are shown in Table 2. As both the figures and the table clearly show that the reconstruction area has increased using the proposed method with no detrimental effects on the accuracy. However, RMSE is not better than before. The reason is that since the correction algorithm is a linear transformation, small noise occurs when a new pattern curves are detected. This will be investigated in our future research. Fig.11 shows other results of moving humans playing several sports. We can confirm that the actively moving humans are successfully captured with our system.

6 Conclusion

We proposed a method to reconstruct the entire shape of a moving object with an active scanning method, while reducing the influence of texture and cast shadows on the projected pattern with a novel algorithm for silhouette detection. We show that our

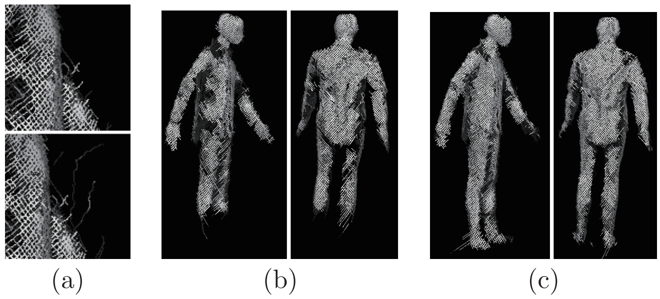


Figure 10. Results comparison. (a) With (top) and without (bottom) silhouette information. (b) Previous and (c) proposed method.

Figure 11. Reconstruction result of moving human (see also supplemental movies).

method allows to increase the reconstruction area without affecting its quality. In our future work, we will improve color compensation for dark regions and silhouette creation accuracy near the cast shadow boundary.

References

[1] R. Furukawa, R. Sagawa, A. Delaunoy, and H. Kawasaki. Multiview projectors/cameras system for 3d reconstruction of dynamic scenes. In *Proc. 4DMOD Workshop on Dynamic Shape Capture and Analysis*, pages 1602–1609, 2011.

[2] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *CVPR*, 2007.

[3] A. Laurentini. The visual hull concept for silhouette-based image understanding. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(2):150–162, 1994.

[4] Persistence of Vision Pty. Ltd. POV-Ray, 2004.

[5] R. Sagawa, Y. Ohta, R. Furuakwa, H. Kawasaki, and Y. Yagi. 3d reconstruction method with detection of dense pattern using the de bruijn sequence and belief-propagation. In *The Institute of Electrical Engineers of Japan (IEEJ) SIG-GI*, pages 47–52, 2009.

[6] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3D. In *ACM SIGGRAPH*, pages 835–846, 2006.

[7] Y. Thibault, H. Kawasaki, R. Sagawa, and R. F. and. Exemplar based texture recovery technique for active one shot scan. In *Proceedings of The 13th IAPR Conference on Machine Vision Applications(MVA2013)*, pages 331–334, 2013.

[8] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.