AI Guided Panoramic Image Reconstruction

Zheng ZHANG, Ecole polytechnique de Bruxelles, ULB, Belgium
Arnaud SCHENKEL, Laboratories of Image Synthesis and Analysis / PANORAMA, ULB, Belgium
Olivier DEBEIR, Laboratories of Image Synthesis and Analysis / PANORAMA, ULB, Belgium

Keywords: *panoramic imaging — moving object detection — moving object removal — image stitching*

CHNT Reference: Z. Zhang, A. Schenkel and O. Debeir. 2020. AI Guided Panoramic Image Reconstruction. *Proceedings of the International Conference on Cultural Heritage and New Technologies in Vienna,* 2020.

Introduction

The panoramic image gives an extensive angle of view, which is widely used in immersive environment generation and virtual reality. Used to create virtual tours of museums, art events, or world heritage sites, 360 panoramic view must take care of human presence during the acquisition to avoid any disturbance in the final product. The construction of the panoramic image is accomplished by stitching multiple photos. However, when the photos are captured in a high-traffic place, the moving objects will degrade the visual quality and cause ghosting artifacts. This project aims to solve the acquisition and the moving foreground interference problem in crowded places where it is never possible to have such a museum of world heritage and generate a high-quality panorama of pure background (Fig. 1). like popular city places such as the Grand Place of Brussels where there are always people, even late at night. Several model-based and artificial intelligence-based approaches are proposed and evaluated. Best results are obtained combining both approaches: Mask R-CNN for object detection and MOG2 for shadow detection.

Fig. 1. Grand Place in Brussels, based on limited time acquisitions. Objects and visitors, stationary during all the acquisitions, could not be correctly removed. (© Arnaud Schenkel).

Panoramic Photography

Since always, crowned and fortunate people liked paintings in general or panoramic view, carried out on walls, tapestries or panels, showing wars or hunting scenes. Its emergence with a wider public dates rather from the 18-19th century with paintings or with pieces of art. The birth of photography did not profoundly change the style of panoramas, as they often remained a succession of images rather than a continuous image. Specialized panoramic camera designs were being patented and manufactured for making panoramas. Several approaches are thus developed:

- taking a series of images, which were then shown placed next to each other to create one image;
- using a specialized rotating lens camera and a curved filming plate (e.g. Joseph Puchberger in 1843);
- using a specialized camera with a wide field of view, up to 180° (e.g. Kodak in 1899).

Since the invention of digital photography, it is easier and much less expensive. Modern solutions are based on the same ideas. Software solutions embedded in modern cameras or smartphones allow panoramic photographs to be taken simply by rotating the device. The result obtained is often limited in terms of resolutions and makes it difficult to deal with the presence of obstacles in the scene. The 360° cameras have the main advantage of capturing the whole space at one time. However affordable 360° cameras are limited to two wide angle cameras, producing images of limited quality in terms of distortion and resolution; while devices composed of a larger number of cameras allow higher quality, also at a higher cost.

The current most common method for producing HD panoramas is to take a series of pictures by turning the camera between each shot, considering an overlap between two consecutive shots, covering the full spherical environment (360-degree panorama), and stitch them together. The number and the angular positions of shots thus depends on the characteristics of the camera (sensor size, orientation), the lens type (rectilinear or fisheye, and the focal length), the coverage and the overlap ratio.

The presence of moving objects or people in the acquisitions leads to defects in the result: presence of artefacts, occlusions, ghosting effects, able to ruin the final composition or to hide some important details in an architecture or artifacts in a museum room... Simple stitching processing of these images does not allow to deal with the whole problem, allowing only to partially solve the issues in the overlapping parts.

Approach to Issues with Moving Objects

The global solution of the problem consists in analysing the photographs in parallel with their acquisition to know whether they should be taken again. In animated site contexts, on the one hand, it is difficult to obtain an image without any moving object; on the other hand, each new contribution makes it possible to obtain new small background areas. The combination of all of these areas therefore makes it possible to obtain a completely cleaned panoramic image. The three major steps are:

- Acquisition: datasets are captured pose by pose under control of the robotic head, using multiple passes, producing an image sequence for each camera pose. The robotic system has slight inaccuracies; for a “same” camera pose, different epochs of an image sequence can present small pose variations.
- Moving objects removal: Assuming that the pixels of the same coordinate are not covered by the foreground objects in all photos, the moving object removal approach can extract background information from the image sequences. The presence of a foreground object in the same picture’s part in all the sequences implies the need for a new acquisition. This step is subdivided in the detection and in the removal of the foreground object.
- Stitching: After removing the moving foreground objects, there will be pure background images having overlapping parts, and then these images can be stitched to a panorama.

Moving Foreground Objects Detection

A basic approach to detect moving objects in an image sequence is to calculate the difference between two images, considering a third one allows to identify the persistent elements and therefore to differentiate the foreground and the background; this is the three-frame differencing method. The MOG2 (Mixture of Gaussians v2, background subtractor with shadow detection described by Zivkovic (2004)) and KNN (K-nearest neighbors-based foreground segmentation described by Zivkovic and Van der Heijden (2006)) are background modelling methods, which are usually applied to videos and need hundreds-frames initialization. Three cold-start initialization methods are proposed: the median image initialization, the gamma adjusted initialization, and the image pre-feeding. YOLO (You Only Look Once, real-time and end-to-end object detection convolutional neural network described by Redmon et al. (2016)) and Mask R-CNN (Region-based Convolutional Neural Network method with object mask prediction, described by He et al. (2017)) are deep neural networks-based (DNN) approaches. These two object detection networks can detect foreground objects directly with knowing the class of the foreground objects, like pedestrians and vehicles.

Three-frame differencing, MOG2, and KNN methods require dynamics in the scene to identify moving objects; when they move little, they will not be able to be properly identified. Deep learning approaches make it possible to theoretically identify in a single image all the foreground objects. However, that does not consider the shadows they produce. A combination of model-based and DNN-based methods solves this problem.

Moving Foreground Removal Approaches

The classical approach is the median of images. However, this approach has ghosting artifacts in the area where the foreground objects are denser. A better approach is the foreground mask-based approach. The pixel values in the output are the median of unmarked pixels at the corresponding coordinates.

Depending on the possibilities of acquisition; several scenarios are possible:

- the combination of the images allows to correctly and completely remove the foreground objects, the process then continues with the next step;
- the images acquired are insufficient to treat the problem: either it is possible to make new acquisitions, a new cycle of acquisition and foreground removal is carried out; either a new acquisition is not possible (due to limit of the acquisition time for example, depends on functional or lighting conditions), an image is then selected to present an entire obstacle in order to give a tangible result.

Panoramic Stitching

The determined best moving foreground objects removal algorithm is integrated into the panoramic reconstruction pipeline. After removing the moving foreground objects, there will be pure background images having overlapping parts, and then these images can be stitched to a panorama. The desired output is a 360-degree panorama, which is an immersive image containing information of the full enclosed-sphere scene from one viewpoint inside the sphere.

Conclusion

All the moving object removal approaches are evaluated and analysed. Best results are obtained combining both approaches: Mask R-CNN for object detection and MOG2 for shadow detection.

The main contributions of this Master thesis include: (1) data acquisition and semi-synthetic data generation; (2) the initialization method to apply the background modelling approaches on a sequence of few pictures; (3) improvement and evaluation of the moving foreground removal approaches. The experimental result shows that the ghosting artifacts and the foreground objects have been effectively removed (Fig. 2).

Fig. 2. Comparison of traditional panorama (top) and panorama with proposed method (bottom) (© Zhang Zheng).

References

Szeliski, R., 2006. Image Alignment and Stitching: A Tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1), pp.1-104.

Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction, *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference*, volume 2, pp 28-31.

Zivkovic, Z. and van der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction, *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773-780.

He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.