Robot Body Occlusion Removal in Omnidirectional Video Using Color and Shape Information

Binbin Xu, Sarthak Pathak, Hiromitsu Fujii, Atsushi Yamashita, Hajime Asama
Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

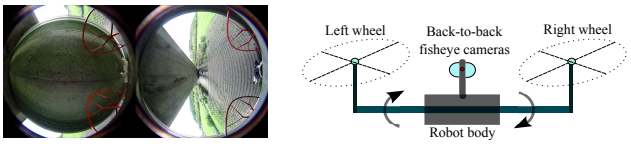
Omnidirectional cameras are widely used for robot inspection for their wild fields of view. However, the robot body will always be included in the view, causing occlusions. This paper deals with one such example of occlusion and proposes an inpainting based solution to remove it. Our method could generate a clean video automatically, without the need for a manually given mask. We propose an approximate shape fitting method combined with color information to generate the mask of robot body occlusion and followed by video inpainting. In experiments, the effectiveness of our proposed method is demonstrated by successfully removing robot body occlusions in omnidirectional videos.

1 Introduction

Omnidirectional videos captured by two back-to-back fisheye cameras, like Ricoh Theta cameras, have a wide variety of applications, such as robot surveillance and inspection. However, a common problem existing in these applications is that robot body itself always appears in the videos, which causes undesired regions as shown in Fig. 1(a). To solve this problem, it is necessary to (1) find robot occlusion regions and (2) do inpainting by finding appropriate background regions to replace occlusions in the omnidirectional view. Occlusion region in previous researches are either given manually [1] or found by frame alignment [2]. The latter is based on an accurately reconstructed 3D model, which is hard to obtain. Besides those methods, omnidirectional inpainting is done by projecting to perspective images and then inpainted in a frame-by-frame manner in [1] or done by finding camera position correspondence in a multi-camera system in [2]. These works are shape-dependent and thus hard to remove occlusions near discontinuous boundary regions. In this study, we aim at generating clean omnidirectional videos by finding undesired robot body regions automatically using prior knowledge about robot body's color and shape and then inpainting with the background scene by searching for the best patch over spatio-temporal dimension in video. Here we illustrate our method in a bridge inspection setting, where the robot body is shown in Fig. 1(b). It is composed of two wheels to help attach on the surface of infrastructure and two back-to-back fisheye cameras to capture a full 360 degree omnidirectional video. In this situation, the wheels always occlude the camera's view.

2 Robot body occlusion segmentation

In the proposed method, the first step is to find robot body occlusion region using color and shape information. Firstly, to utilize the color difference between robot body and background, we train color features with a Support Vector Machine (SVM) classifier. The parts containing the robot body are selected manually and exploited as positive samples for training. The background scene in the image frames are exploited as the negative samples. Here we use the HSV color space to extract pixel-level color feature, which is robust to the brightness component change. To eliminate the effect of lighting change, we discard the lightness value and only consider HS color space in this work. Then, we use a Gaussian kernel SVM to train a color detector based on the extracted color features.



(a) Robot body occlusion (b) Robot body model

Fig. 1 Example of occlusions caused by robot body when using a Ricoh Theta camera in inspection

Color segmentation could give a mask of robot body occlusion region. However, there will always be some noise due to misclassification. The extra mask not only causes increased computation time for inpainting, but also decreases quality of inpainting output. To deal with this problem, here we use shape fitting to get rid of most misclassification noise outside of robot body. The main occlusion in this case is caused by wheels, whose contours are circles and should be ellipses in sphere projection model. However, the wheels are only partly seen in each fisheye image and the stitching margin is discontinuous due to mapping distortion. Besides, distortion of wheels caused by external force leads to more distorted ellipses. Therefore, it is hard to use a least-square fitting algorithm to fit a distorted and discontinuous shape. Considering we only need an approximate shape contour to get rid of noise, we propose an iterative improvement to Fitzgibbon et al's method [3] with a better approximation for distorted and discontinuous ellipses without guessing accurate initial values and computationally expensive optimization. It consists of three steps: (i) initialization, (ii) refitting and (iii) propagation.

Initially, we do morphology operations such as dilation to binary mask images generated by color segmentation and split it into two parts. Each of them contains one wheel mask with some noise. Then we use Fitzgibbon et al's direct least-squares ellipse fitting method for each wheel [3]. It can give us an initial estimate, with the ellipse center (x_0, y_0), semi-major r_{max}, semi-minus axis radius r_{min} and orientation θ_0. Due to distortion and discontinuity, the resultant ellipse is not accurate at all (Fig. 2(c)) but we discover that the constructed ellipse still covers most part of wheels and less of noise than before. Using this phenomenon, we refit the ellipse based on the more reliable points which are inside the initially estimated ellipse to make more accurate approximation. We find that the refitted ellipse gives a smaller radius but a

more accurate center x_{inew}, y_{inew} and orientation θ_{inew}. The third step, propagation, encourages ellipse location to be shifted to a more accurate approximation and to embody more wheel points. In this step, we constructed a new ellipse by calculating weighted sum of parameters of previously fitted ellipse and the one fitted by points inside. That is to say, we constructed an ellipse, whose parameters are $x_i = kx_{i-1} + (1-k)x_{(i-1)new}$, $y_i = ky_{i-1} + (1-k)y_{(i-1)new}$, r_{max}, r_{min}, $\theta_i = p\theta_{i-1} + (1-p)\theta_{(i-1)new}$ in this step. The refitting and propagation steps are iterated several times to converge to a good solution. To deal with wheels with certain width, we simply expand the radius of ellipse to cover the whole ellipse at the final stage of the approach.

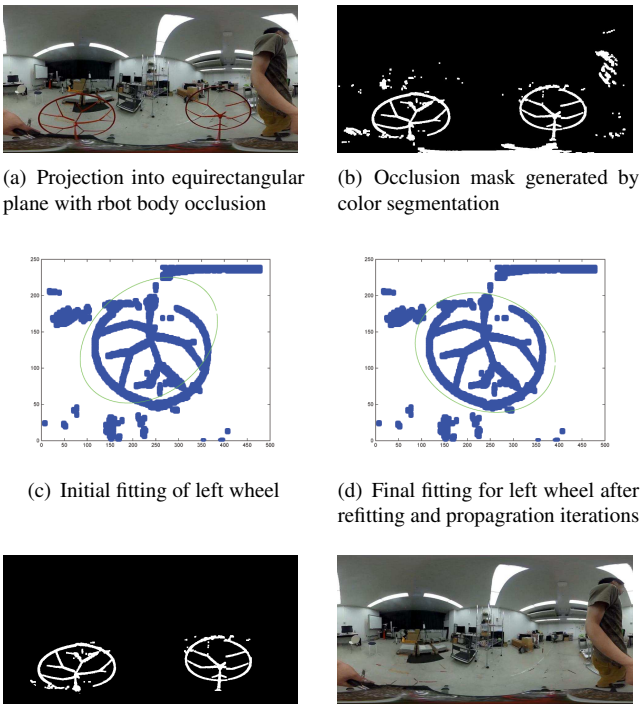
3 Omnidirectional video inpainting

Here we adopt a video inpainting method [4] to complete omnidirectional videos. The original video inpainting algorithm works only on videos captured by normal cameras, not fisheye cameras. To address this problem, we project and stitch them as in section 2 to generate an omnidirectional equirectangular video. Then we apply video inpainting algorithm on the equirectangular video. The main idea is to construct an image pyramid and use patchmatch algorithm to search for the nearest neighbors of occlusion patches in spatio-temporal extension by minimizing the similarity energy function. Then the video is reconstructed by choosing the most reliable color which minimizes the weight mean function in the spatio-temporal neighborhood.

4 Experiments and results

We demonstrated the effectiveness of our proposed method by completing missing regions in video sequences taken in laboratory and bridge inspection environment.

In this work, we used a Ricoh Theta m15 camera, which could capture a stitched omnidirectional video through two internally calibrated back-to-back fisheye cameras by an equirectangular projection. We collected training images in several environments and trained the color detector using SVM. To obtain the optimal parameters for SVM kernel, we used the cross validation method based on grid search for the discriminant function. In Fig. 2, we show the whole inpainting process on the 18th frame in one video. The originally captured omnidirectional image is shown in Fig. 2(a). Then we used our prebuilt color detector to generate an initial mask image (Fig.2(b)), which was then refined by using our approximate shape fitting algorithm. The noise in the initial mask image would unavoidably lead to blurred inpainting regions if it hadn't been removed. The initial shape fitting was inaccurate due to noise and distortion (Fig. 2(c)). In our work, we set k to 0.5 and p to be 0.3 in propagation step to find a more accurate approximation. After several iterations of refitting and propagation, the finally fitted shape was very close to the distorted wheel (Fig. 2(d)). The result after shape fitting is shown in Fig. 2(e). We then used the video inpainting method to this video with generated mask. The output can be seen in Fig. 2(f). We also evaluated our method in a real robot inspection environment with moving wheels. The robot body was also removed from video successfully.



(a) Projection into equirectangular plane with rbot body occlusion

(b) Occlusion mask generated by color segmentation

(c) Initial fitting of left wheel

(d) Final fitting for left wheel after refitting and propagration iterations

(e) Mask refined after shape fitting

(f) Output of inpainted video

Fig. 2 Experiment in laboratory environment

5 Conclusion

In this work, we presented an automatic robot body removal method for omnidirectional videos. It needs nothing but prior knowledge of robot body's color and shape information and works well in both laboratory and real inspection environment. In the future, we will apply our work to more kinds of robot situations.

ACKNOWLEDGMENT

This work was in part supported by the SIP Program (Cross-ministerial Strategic Innovation Promotion Program).

References

[1] Daniel Paredes, Paul Rodriguez, and Nicolas Ragot. Catadioptric omnidirectional image inpainting via a multi-scale approach and image unwrapping. In *2013 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, pages 67–72, 2013.

[2] Norihiko Kawai, Naoya Inoue, Tomokazu Sato, Fumio Okura, Yuta Nakashima, and Naokazu Yokoya. Background estimation for a single omnidirectional image sequence captured with a moving camera. *Information and Media Technologies*, 9(3):361–365, 2014.

[3] Andrew Fitzgibbon, Maurizio Pilu, and Robert B Fisher. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 1999.

[4] Alasdair Newson, Andrés Almansa, Matthieu Fradet, Yann Gousseau, and Patrick Pérez. Video inpainting of complex scenes. *SIAM Journal of Imaging Science*, 7(4):1993–2019, 2014.