CCC 2018

Proceedings of the Creative Construction Conference (2018)
Edited by: Miroslaw J. Skibniewski & Miklos Hajdu
DOI 10.3311/CCC2018-067

Creative Construction Conference 2018, CCC 2018, 30 June - 3 July 2018, Ljubljana, Slovenia

Generating a visual map of the crane workspace using top-view cameras for assisting operation

Yu Wang[a], Hiromasa Suzuki[a], Yutaka Ohtake[a], Takayuki Kosaka[b], Shinji Noguchi[b]

[a]*Department of Precision Eng., The University of Tokyo, Tokyo 113-8656, Japan*
[b]*TADANO LTD., Takamatsu Kagawa 761-0301, Japan*

Abstract

All terrain cranes often work in construction sites. Blind spots, limited information and high mental workload are problems encountered by crane operators. A top-view camera mounted on the boom head offers a valuable perspective on the workspace that can help eliminate blind spots and provide the basis for assisting operation. In this study, a visual 2D map of a crane workspace is generated from images captured by a top-view camera. Various types of information can be overlaid on this visual to assist the operator, such as recording the operation and projecting the boom head's expected path through the workspace. Herein, the process of generating a visual map by stitching and locating the boom head trajectory in that visual map is described. Preliminary proof-of-concept tests show that a precise map and projected trajectories can be generated via image-processing techniques that discriminate foreground objects from the scene below the crane. These results show a way to help the operator make more precise operation easily and reduce the operator's mental workload.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2018.

Keywords: all-terrain crane, top-view camera, optical flow, image stitching;

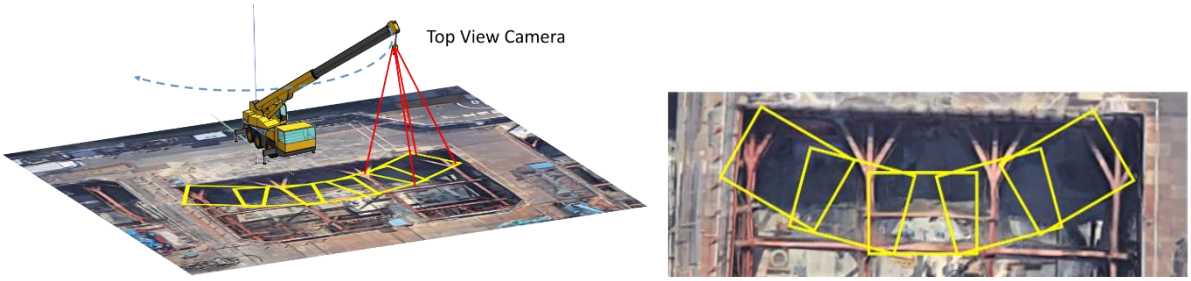
1. Introduction

All-terrain cranes are widely used in construction, transportation and other industries due to their good mobility and capacity [1]. Fig. 1 shows an all-terrain crane in operation. A load is suspended from the boom and transported.

The operators of all-terrain cranes face a constantly changing workspace. The chief dangers to crane operators are a congested working environment, neglect of hidden dangers, and a lack of information for decision making especially for the operators with little working experience. Oscillation of the suspended load can also present a challenge to crane operators. Many studies have addressed methods for reducing oscillation with the control theory for more precise lifting [2-4] and have mostly focused on structural features of the crane itself. Lifting path planning is also a promising way to improve the precision and efficiency of crane operation [5]. Safety and precision can be improved with a careful consideration of the workspace. Thus, accurate 3D information about the environment is required. Precise scanning of the environment and data processing

Fig.1 A crane in operation. A load is suspended to the boom and transported. A top view camera is also suspended at the top of the boom to survey the working environment.

Corresponding author: Yu WANG email: kylinwang0123@gmail.com



(a) Top-view camera surveying workspace (b) Stitching to generate a workspace map

Fig. 2. (a) A top view camera suspended from the boom head continuously capturing images shown in yellow rectangles of the workspace. (b) Image-stitching process to stitch these images to produce a wide-range image of the workspace.

tend to be time consuming and cost a lot. One lab, for example used data from crane sensors to roughly examine the working environment for path planning [5].

In this study, the crane operator's limited visibility and insufficient information about the workspace are the primary concerns. As shown in Fig. 2. (a), a top-view camera is mounted on the boom head that moves over the workspace along with the boom. Bird's-eye view images can be captured using the top-view camera. With several images captured from the top-view camera, a wide range of the workspace can be represented by stitching and rendering these images, as shown in Fig. 2. (b). The stitched top-view camera image can provide a rich range of information to the operator. Herein, the stitched wide-range image is referred to as the workspace map. A variety of assistance applications of the workspace map can be considered. An optimal path to transfer a load can be displayed on the workspace map to aid the operator. Information, such as the position of the boom head and a 2D projection of the lifting path, can also be included in the workspace map, along with other representations that researchers may devise in the future.

2. Foreground detection and mask generation

To generate a clear overall workspace map, the workspace must be imaged beforehand. This pre-shooting process requires the following conditions, which are diagrammed in Fig. 3.

- The top-view camera is located at the top of the boom at a sufficient height to cover a wide area of the workspace.
- The optical axis of camera should be pointed vertically at the ground.
- While taking the images, the top-view camera rotates with the boom's rotation only. If an extension of the boom is necessary, the height of the top-view camera should be kept as constant as possible to make images captured having a close scale.
- Images captured with the top-view camera include background and foreground objects. The background is the ground and objects resting on it. The foreground includes objects that are hung to the boom and move along with it, such as a hook, a wire, and a swinging load. During the pre-shooting process, the foreground should be removed as much as possible as they must not show up in the workspace map as ghosts. For this reason, during the pre-shooting, winding up the cable and detaching the load are recommended. But it is impossible to exclude them completely. We need some means to detect the foreground.

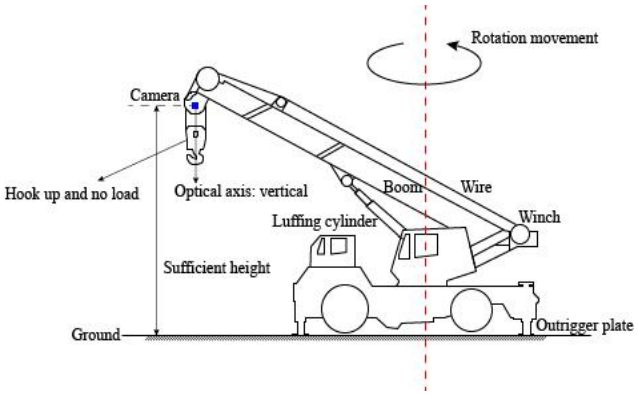
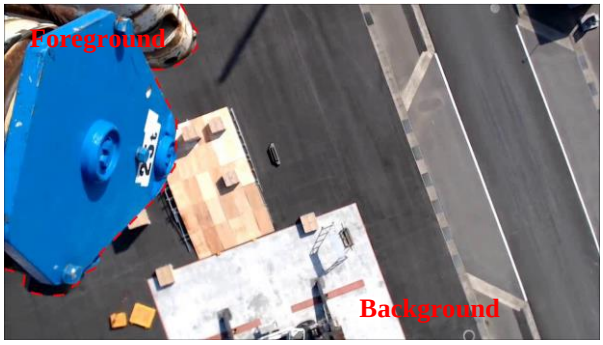


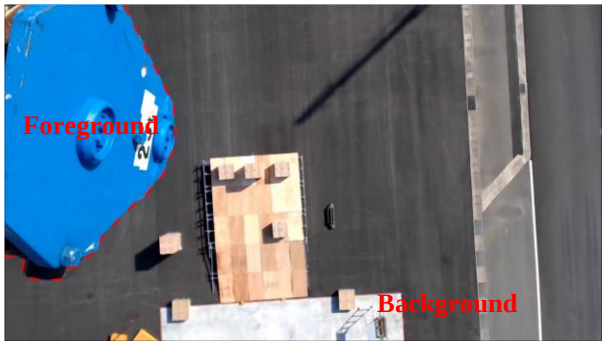
Fig. 3. Conditions of pre-shooting of the background images

Under these assumptions, the images captured with the top-view camera can be composed using panoramic stitching [6]. Fig. 4 shows typical images captured by the top-view camera under the conditions mentioned above. These images are stitched together to generate a workspace map that covers the whole background over which the crane has gone through. However, even under the conditions mentioned above, these images still often contain such foreground objects as a boom head and a hook. Simply stitching these images will cause ghosts of these foreground objects to appear in the final stitched workspace map. For example, in Fig. 4., stitching the workspace map directly from images (a) and (b) captured by the top-view camera yields a ghost in the resulting image (c).

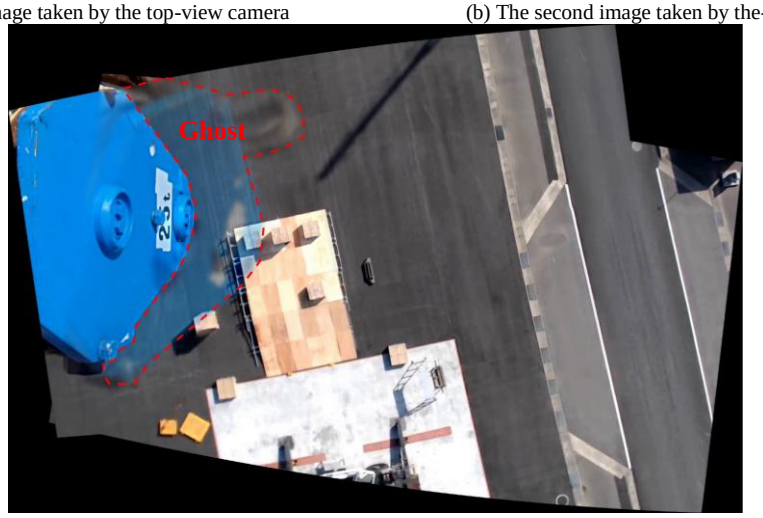
To generate a clean workspace map, it is better to generate a mask to eliminate the foreground objects and prevent ghosts appearing in the reference map.



(a) The first image taken by the top-view camera



(b) The second image taken by the-top view camera



(c) Stitching result

Fig. 4. (a)(b) Typical top view camera images contain foreground object (a hook) enclosed with dotted red lines the background. (c) Stitched image directly using images (a) and (b) has ghost of the hook in the region enclosed by red lines.

2.1. Motion segmentation with moving camera

We should remove the foreground objects by masking them with a mask covering them precisely. It is necessary to discriminate the background and foreground objects in the images. One method to distinguish them is by the difference of their motion. This problem is known as motion segmentation with a moving camera [7].

Serajeh has proposed a method for this problem based on epipolar geometry and dense optical flow [8]. That method is intended to extract moving objects from images captured with a hand-held moving camera. This paper considers the addressing of this problem using such a structure from motion (SFM) technique. First, with RANSAC algorithm, the epipolar geometry between two images is estimated to calculate the fundamental matrix. Second, the dense optical flow is calculated to find the corresponding point in the second image for every pixel in the first image. Then the corresponding points in the second image that keep a significant distance to the epipolar lines are detected as moving objects in the scene. This process is applicable to a wide range of cases.

Under some conditions, however, the SFM technique will not yield satisfactory results. One extreme condition is if the image planes of two camera position while capturing is parallel to each other. The epipolar lines on both images will be parallel. If the foreground objects are moving on these epipolar lines, all the foreground objects will be on epipolar lines, making the measurement of epipolar distances impossible. This extreme condition happens rarely because it requires the two image planes parallel to each other and the foreground objects moving on the epipolar lines. Another example is that in which the movement of a moving object is complicated with both translational and rotational movements. In this case, some points on the moving object may lie on epipolar lines of the second image while the rest do not. Then, part of the moving objects can be detected in the first image. Unfortunately, this condition always happens for the images captured by the top-view camera of crane.

2.2. Proposed method

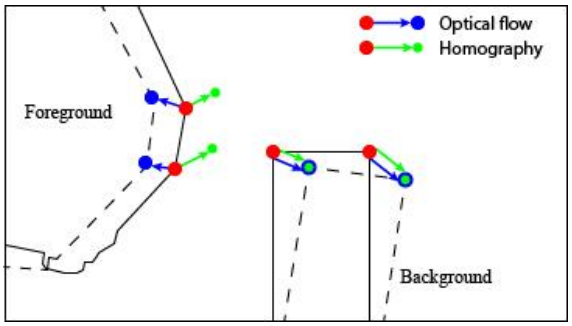
From one image to another image captured with the top-view camera, foreground objects will always show a complicated movement because of the crane hook's oscillation. Because of this complicated movement, only parts of the foreground object will lie on epipolar lines from the perspective of the second image, so the foreground cannot be detected in full.

One of this article's main contributions is to make up for this shortcoming in epipolar geometry by separating foreground and background objects based on a combination of dense optical flow and homography. This method computes the relatively subtle trajectory of the background as represented in homography and compares that motion with the dense optical flow between the images.

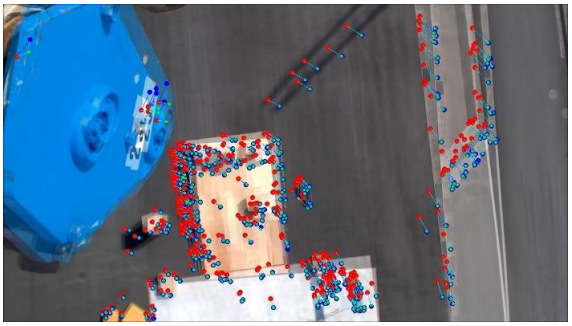
Homography represents a linear transformation between two images. The optical flow of the background should be consistent with the homography. However, the optical flow of foreground objects will not be consistent with the homography between the images.

To illustrate the difference of foreground and background optical flows' matching with homography, a simple diagram of the proposed method appears in Fig. 5(a). Both optical flow and homography are representations of pixels' movement from one image to another. In Fig. 5(a), optical flow is represented as the movement from red points to blue points. And homography is the movement from red points to green points. For background, the movements of red points to blue points and red points to green points are the same, i.e., the optical flow is consistent with the homography. However, for the foreground, movements of red points to blue points and red points to green points are not the same, i.e., the optical flow is not consistent with the homography. If and only if pixels of the foreground have the same movement with the background, detection of these pixels will fail.

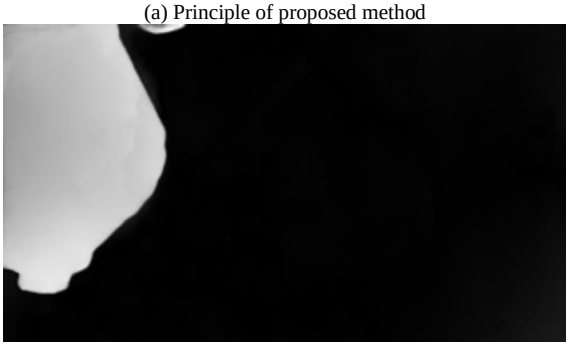
Fig. 5. (b) shows a test on images captured with the top-view camera of the proposed method. Fig. 5. (b) shows a test on images captured with the top-view camera of the proposed method. As can be seen, the foreground is a blue hook. The red points are SIFT (scale invariant feature transformation) features [9]. The homography estimated from the matched features of the two images is represented as the movement from red points to green points. The dense optical flow computed with flownet2 [10] returns the pixel movements from the red points to blue points. Just as mentioned above, in the background, these two movements match. On the other hand, in the foreground, the homography and optical flow of foreground objects are not consistent. The result of foreground-object detection is shown in Fig. 5. (c), and comes from identifying the points for which optical flow and homography do not match. Fig. 5. (d) overlaps the binarized foreground mask in Fig. 5. (c) onto the image that will be used in stitching.



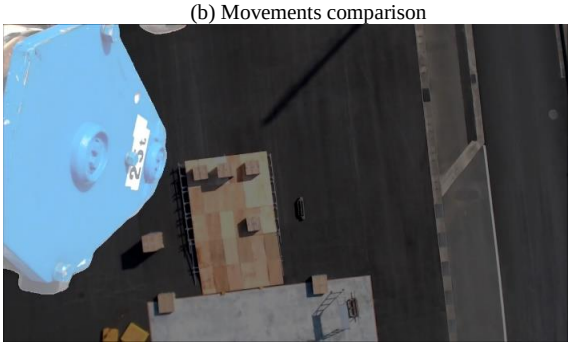
(a) Principle of proposed method



(b) Movements comparison



(c) Detected foreground mask of our method



(d) Comparison of binarized mask and image

Fig. 5. (a) A principle diagram of the proposed method to detect foreground by examine of the distance from the blue point to the green point. (b) A test result of the method on two images captured by the top-view camera showing sparse optical flow and homography. (c) Detection of foreground by comparing the difference between the dense optical flow and homography. (d) Comparison of binarized mask with the image.

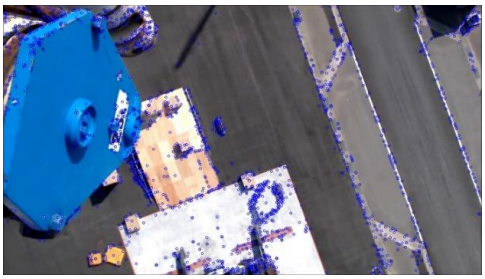
3. Workspace map generation with image stitching

The image-stitching problem is well understood. Image alignment and stitching through feature-based matching to estimate a homography are the most important steps. Fig. 6 shows the process of stitching two images into one panoramic image.

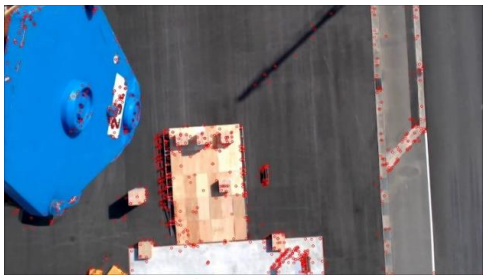
The first step is to find robust features such as SIFT features [9], KAZE features [11] in the two images. Here, SIFT features are chosen because their good scale invariance, rotation invariance and illumination invariance. As shown in Fig. 6. (a) and (b), SIFT features are extracted from each image. The second step is the estimation of the homography by matching the SIFT features detected the first step. To estimate the homography with the matched results robustly, the RANSAC (random sample consensus) algorithm is implemented [12], which robustly identifies inliers of the matched features and estimates the homography with high precision. Fig. 6. (c) shows the result of RANSAC inliers.

After the estimation of homography H from the matched features between the two images, the homography H is used to warp the first image with projective geometry. Coordinates in the image to be warped are represented as $P_i(x_i, y_i, 1)$. The corresponding point in the second image is $P_i'(x_i', y_i', 1)$. P_i' can be easily obtained with equation $w_i P_i' = H P_i^T$, where w_i is a scale parameter and H is a 3×3 matrix representing the homography. So all pixels in the first image find their corresponding point in the second image. Fig. 6 (d) shows the result after aligning the first image to the second image.

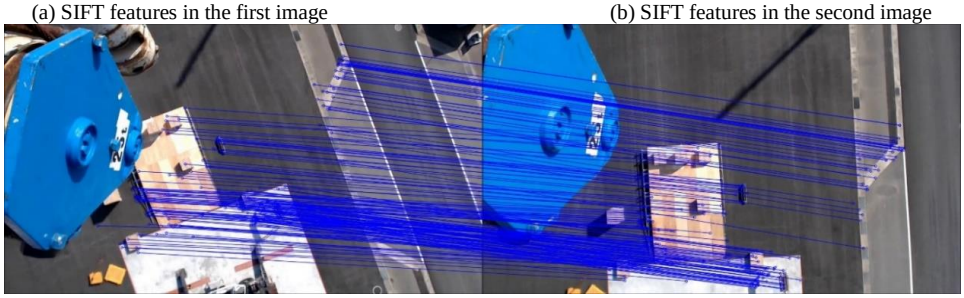
Once the images are aligned, they simply need to be blended together. Multiband blending is used for this process because of its good performance on many examples of image stitching [13]. Fig. 6. (e) shows the blended result of Fig 5. (c) using masks detected using the method proposed in section 2. To make the foreground mask more reliable, morphological dilation is applied [14].



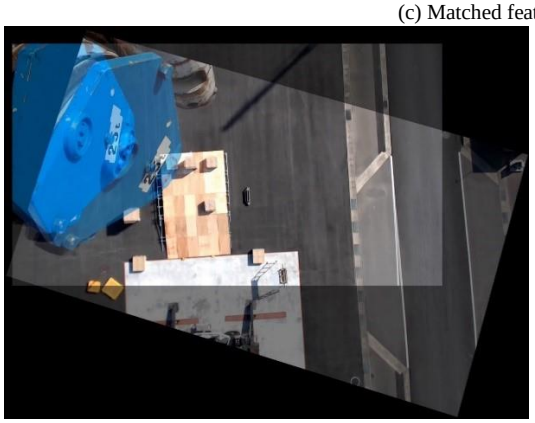
(a) SIFT features in the first image



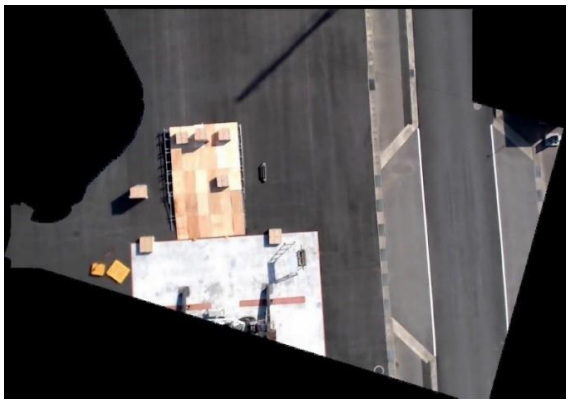
(b) SIFT features in the second image



(c) Matched features of RANSAC inliers



(d) Aligned result according to homography



(e) Stitching result

Fig. 6. (a)(b) SIFT features are detected from both images. (c) RANSAC inliers are extracted from the matched SIFT features. Homography is estimated with these RANSAC inliers. (d)With the applying of the homography estimated in (c), the first image is warped and aligned with the second image. (e) The result of stitching by applying multiband blending on (d).

4. Experiments with application of path location display

The methods described above were tested with prototype software to prove that the concept is feasible. The input is a video V_{pre} recorded in the pre-shooting under the conditions described in section 2. V_{pre} is a sequence of images from which blurred frames are excluded with a simple filter. Then several key frames for V_{pre} are selected automatically by considering the overlap ratio. A reasonable overlap ratio could ensure that there are enough features existing in the overlapped region. This could ensure the homography between tow images being estimated successfully. On the other hand, the overlap ratio can ensure that stitching one image to another with a significant non-overlapped area. The selected key frames are noted $k_0, k_1, \cdots, k_N$. The generation of the workspace map is made by the following two steps.

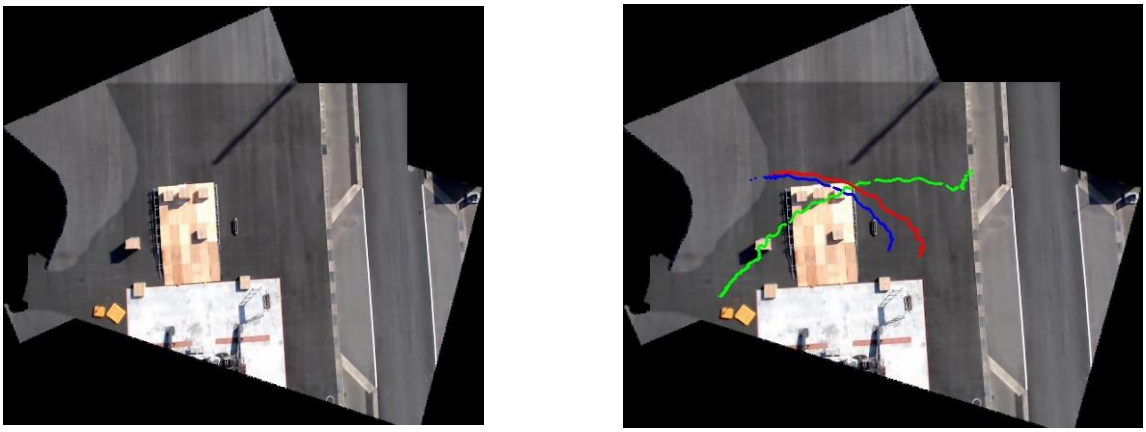
The first step is to compute a mask for each key frame. For each of the other key frames k_i, a foreground mask should be detected by computing the homography vectors $\boldsymbol{v}_{\mathrm{homo}}(\boldsymbol{p})$ and optical flow vectors $\boldsymbol{v}_{\mathrm{opt}}(\boldsymbol{p})$ for all the pixels $\boldsymbol{p}$ of k_i [10]. For this purpose, one support frame s_i was chosen for computing the homography and optical flow. The support frame s_i were chosen from a set of 20 frames near k_i in V_{pre} one by one to generate an optimal mask. This selection was done manually for the experiment in this article but can be automated in the future. Only three key frames were chosen from V_{pre} for these preliminary tests, which was not difficult to perform manually. From k_i to s_i, for all

the pixels $\boldsymbol{p}$, $\boldsymbol{v}_{\text{homo}}(\boldsymbol{p})$ is computed by the homography. The homography H is a mapping from the key frame k_i to s_i, that is, all the pixels $\boldsymbol{p}$ of k_i is mapped to their corresponding locations $H(\boldsymbol{p})$ on s_{i}. Thus $\boldsymbol{v}_{\text{homo}}(\boldsymbol{p})$ can be computed by $\boldsymbol{v}_{\text{homo}}(\boldsymbol{p}) = \boldsymbol{p} - H(\boldsymbol{p})$. For $\boldsymbol{v}_{\text{opt}}(\boldsymbol{p})$, it can be directly estimated with flownet2[10]. Then the difference D between $\boldsymbol{v}_{\text{homo}}(\boldsymbol{p})$ and $\boldsymbol{v}_{\text{opt}}(\boldsymbol{p})$ for all the pixels $\boldsymbol{p}$ of k_i can be computed with $D = \left\|\boldsymbol{v}_{\text{homo}}(\boldsymbol{p}) - \boldsymbol{v}_{\text{opt}}(\boldsymbol{p})\right\|$. The foreground pixels $\boldsymbol{p}_{\text{f}}$ and the background pixels $\boldsymbol{p}_{\text{b}}$ vary a lot in the value of D. Thus, by filtering with a threshold, the foreground pixels $\boldsymbol{p}_{\text{f}}$ can be picked out as the foreground mask m_i with a binarization process.

The second step is to stitch the key frames to form the workspace map. First the base key frame k_{base} is chosen, and is usually the image lying near the center of the workspace. For each key frame k_i, its homography H_i to k_{base} is computed through the matched feature points in the region they share. The homography is a mapping transforming key frame k_i to k_{base} to form the aligned images. With multiband blending, these aligned images can then form the stitched workspace map [13]. However, the workspace may be too large for k_{base} and k_i to share a common region. The manual selection of the key frames from V_{pre} is constrained in that the near key frames must have a reasonable overlapping ration for the homography between two key frames can be calculated. If this condition holds, the homography $H_{i,i+1}$ ($H_{i,i-1}$) between two key frames k_i and k_{i+1} (k_i and k_{i-1}) can be computed successively. Then H_i can be recursively defined as $H_i = H_{i,i+1}H_{i+1}$ for $i < base$ ($H_i = H_{i,i-1}H_{i-1}$ for $i > base$). This computation process should begin with the base key frame and extend out to both sides. These conditions ensure that all the information needed to warp and place the selected key frames in the correct positions can be calculated. A workspace map W can be generated by stitching k_i masked with m_i with respect to k_{base}. Fig. 7. (a) shows an example of W with three key frames chosen from V_{pre}.

One goal of this study is to utilize the workspace map W for displaying some information to assist the operator. Here an application to overlay the path of the boom on W as a simple and useful piece of information is proposed to assist the operator. Another video V was recorded under the crane's ordinary working conditions to test the process for overlaying the boom positions onto the workspace map W. The path of the boom head T can be identified in this input video V and overlaid on the workspace map W. For each frame $f_i \in V$, we computed the homography from f_i to W. Assuming the center of the top-view camera is always just below the boom head, the boom head position T_i on W is represented by the center position of f_i. By plotting T_i for all $f_i \in V$, the path of the boom head can be overlaid on W.

Fig. 7. (b) shows the results of displaying the boom head position on the stitched workspace map. Three videos were used. The first video was recorded under the constrained conditions described section 2 with only a blue hook as the foreground in frames. This video was used to generate the workspace map. The second and third videos are recordings of crane's ordinary working operations of moving an object with a hook. As shown in (b), the three paths consisting of many locations are clearly shown. Some short gaps appear in the blue and green paths due to the blurry frames that were removed from the sample videos. The software designed in this study successfully produced basic workspace maps.



(a) Stitching workspace map with three frames (b) Result of displaying boom head's path on workspace map

Fig. 7. (a) The workspace map is generated by stitching with three key frames from the pre-shot video. (b) Three clear paths are formed by locating the image's position on the workspace map.

5. Conclusion

The proposed prototype software with proposed methods can create a clear workspace map from videos recorded from the boom head of a crane. No ghosts were apparent in the maps generated and the locations of objects in the videos are clearly represented in the generated workspace map. All the location points formed a clear moving path in the workspace map.

More work is required before this technology is ready for commercial use. First, the generation of masks for key frames is not automatic in the process of this article. An automated process should balance camera motion and precise optical flow calculations. The general optical flow calculation method used above cannot give precise result when two frames vary greatly. Second, the hook's position represented in the workspace map could be helpful to operators, and the workspace map method can be extended to include this information. Third, some conditions may require a more-developed warping method for the image-stitching process. Homography used alone cannot deal with very complicated cases such as a workspace that includes tall buildings. Finally, research is needed into additional information that can be displayed in the workspace map, serving purposes beyond than representing the boom's path.

References

[1] Ren, W., Wu, Z., & Zhang, L. (2016). Real-time planning of a lifting scheme in mobile crane mounted controllers. Canadian Journal of Civil Engineering, 43(6), 542-552.

[2] Wu, T. S., Karkoub, M., Yu, W. S., Chen, C. T., Her, M. G., & Wu, K. W. (2016). Anti-sway tracking control of tower cranes with delayed uncertainty using a robust adaptive fuzzy control. Fuzzy Sets and Systems, 290, 118-137.

[3] Burul, I., Kolonić, F., & Matuško, J. (2010, May). The control system design of a gantry crane based on H∞ control theory. In MIPRO, 2010 Proceedings of the 33rd International Convention (pp. 183-188). IEEE.

[4] Asad, S., Salahat, M., Zalata, M. A., Alia, M., & Al Rawashdeh, A. (2011). Design of fuzzy PD-controlled overhead crane system with anti-swing compensation. Journal of Engineering and Computer Innovations, 2(3), 51-58.

[5] Reddy, H. R., & Varghese, K. (2002). Automated path planning for mobile crane lifts. Computer-Aided Civil and Infrastructure Engineering, 17(6), 439-448.

[6] Brown, M., & Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. International journal of computer vision, 74(1), 59-73.

[7] Namdev, R. K., Kundu, A., Krishna, K. M., & Jawahar, C. V. (2012, May). Motion segmentation of multiple objects from a freely moving monocular camera. In Robotics and Automation (ICRA), 2012 IEEE International Conference on (pp. 4092-4099). IEEE.

[8] Serajeh, R., Mousavinia, A., & Safaei, F. (2017, May). Motion segmentation with hand held cameras using structure from motion. In Electrical Engineering (ICEE), 2017 Iranian Conference on (pp. 1569-1573). IEEE.

[9] D. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the International Conference on Computer Vision ICCV, Corfu, 1999, pp. 1150–1157

[10] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., & Brox, T. (2017, July). Flownet 2.0: Evolution of optical flow estimation with deep networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Vol. 2).

[11] Alcantarilla, P. F., Bartoli, A., & Davison, A. J. (2012, October). KAZE features. In European Conference on Computer Vision (pp. 214-227). Springer, Berlin, Heidelberg.

[12] Fischler, M. A., & Bolles, R. C. (1987). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In Readings in computer vision (pp. 726-740).

[13] Burt, P. J., & Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. ACM Transactions on Graphics (TOG), 2(4), 217-236.

[14] Serra, J. (1983). Image analysis and mathematical morphology. Academic Press, Inc..