International Journal of Computational Intelligence in Control
Vol. 13 No. 1 (June, 2021)

Revolutionary Non-Blind Areas Intelligent Surveillance Systems with See-Through Technology

Yu-Chen Lin [1*], Cheng-Chuan Chang [1], Che-Tsung Lin [2]

[1] *Department of Automatic Control Engineering, Feng Chia University, Taichung 40724, Taiwan, R.O.C.*
[2] *Mechanical and Systems Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C.*
E-mail: *yuchlin@fcu.edu.tw*

Received : 17th September 2020 *Revised on 28th October 2020;* *Accepted: 06th November 2020*

Abstract：This paper proposes a revolutionary non-blind areas intelligent surveillance systems for parking lot management in underground environments. The feature based image stitching and translucent blending processes is used to let video streams captured by multiple cameras be fused into single output and then whose blind areas resulting from the occluding pillar could be easily seen through. Such system allows that anyone can immediately understand what is going on in the whole monitored area with a simple glance and without prior geometrical knowledge of the place because the blind area is translucent. The image stitching processes consists of four main parts: (1) feature correspondence detection and fundamental matrix estimation, (2) outlier filtering scheme based on the RANSAC and epipolar geometry, (3) computing the homography matrix between each pair of images, (4) projectively warping the image with this homography, and augment the front image with the non-overlapping part of the warped images. Finally, the process of translucent blending is applied to eliminate the blind areas of the augmented image, and then the corresponding cameras behind the occluding pillar provide pixels for translucentizing. We have implemented the preliminary system with six surveillance cameras at underground parking lot environments, and experiment results of real world video sequences have been performed to verify the proposed design.

Keywords: non-blind areas surveillance systems, parking lot, translucent blending, epipolar geometry, homography.

1. Introduction

Underground parking lot is usually a prime target for vandalism and theft. Roughly 80% of the criminal acts at campus, shopping malls and business offices occur in the parking lot [1, 2]. Therefore, the security for underground parking lot is always a serious issue that deserves special concern. Nowadays, security cameras can help deter thieves from breaking into cars, and can help aid law enforcement in identifying criminals who do rob your car or hurt you. One major usage of such devices is for the surveillance systems extensively utilized in parking lots, campus or apartments. Most existing systems require experienced people to monitor such area with prior knowledge of the geometrical relationship between each camera. That is why any resident/student belonging to a building/school could not understand the content of surveillance monitors with a simple glance.

In fact, in the era of IOT (Internet of Things) [3], anything connected to each other via internet is expected to exchange information. The UX (User Experience) of any software or apps dominate their popularity. An important factor of a successful service depends on if it is accessible. Remote monitoring became possible because smart devices, such as smart phone or iPad, are popular. Such hardware brings the potential of eligible people to remotely monitor the desired area. For example, parking space could be dangerous at nighttime and so one would like to see if there is any potential hazard before reaching there. An easily understandable monitoring image could be of great help because any one could assess if it is safe or proper to show up in the space being monitored.

A few previous studies have been attempted to use the see-through function for the practical application [4-6]. Kameda et al. [4] propose an outdoor see-through vision system allowing people to display live status of invisible areas hidden by some structure such as buildings, walls, etc. However, there is an inherently limitation, the 3D model of the monitored building, which is not available in most cases. A similar application is also proposed by Barnum et al [5]. They used 2D projective invariant to transfer image data between cameras to render views of dynamic occluded objects. However, their concept limited themselves in two-camera case, such as a traffic intersection see-through. In [6], they synthesize the hidden views with static and mobile surveillance cameras. The mobile device estimates its position by a real-time visual simultaneous localization and mapping (SLAM) system [7] utilizing stereo camera. However, such application is

not necessary because one would not benefit from holding such device just for seeing through some particular places in the building.

In this paper, we utilize feature matching, image stitching and translucent blending processes to see-through the blind areas resulting from the occluding pillars for underground parking lot; in addition, the idea of the panorama image is introduced to make monitoring a wide space, where a lot of cameras are installed. However, in practice, it is often possible to "extrapolate" or "guesstimate" such pixel correspondence relationships based on the correspondence relations of objects seen in both images. While such extrapolation may not be highly accurate and may fail occasionally, it is not necessary to derive such pixel mapping relations precisely, as the goal is to simply eliminate the blind area resulting from the front occluding pillars in the host image and provide a plausible and appealing visualization aid to the monitoring staff at surveillance systems.

The remainder of the paper is organized as follows. In Section II, we briefly introduce the system description and software block diagram. Section III presents the proposed algorithm dealing with image stitching and translucent blending scheme. In Section IV, experimental results are demonstrated. Finally, Section V contains concluding remarks.

2. System Description

The system consists of three cameras mounted on respective occluding pillars, and three stationary source cameras mounted on a host pillar to view three different viewpoints towards the occluded regions in the source view. The input of this systems is video streams from six cameras, and the output video stream can integrated into one panoramic image where the user can see right through each occluding pillar. Figure 1 shows the structure of the parking lot intelligent surveillance systems. Oftentimes occluding pillars lead to the blind spot problem in surveillance systems. Therefore, this paper presents a vision-based intelligent surveillance systems which allows any person or monitors to directly see-through the visual blind spot for monitoring, which to ensures enough time to react in a timely manner, so as to defuse the danger in time.

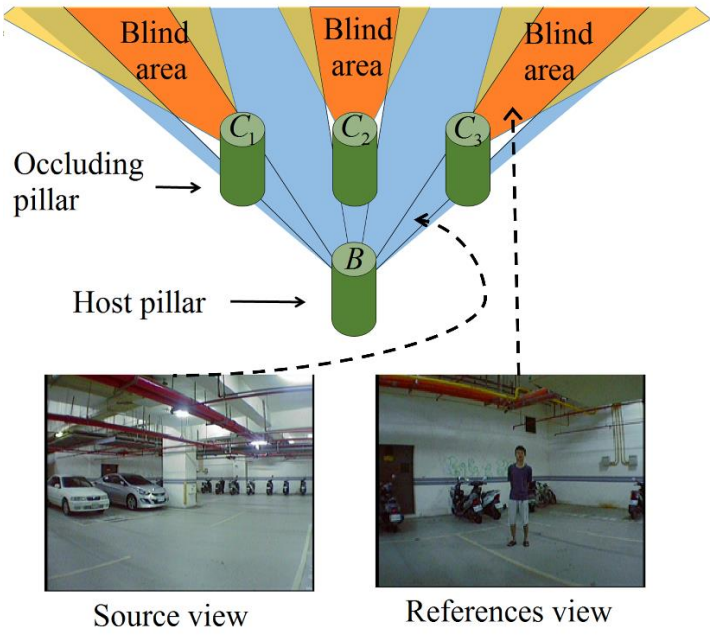


Fig. 1 Structure overview of parking lot intelligent surveillance systems

The corresponding flowchart of the proposed algorithms for non-blind areas intelligent surveillance systems is illustrated in Fig. 2.

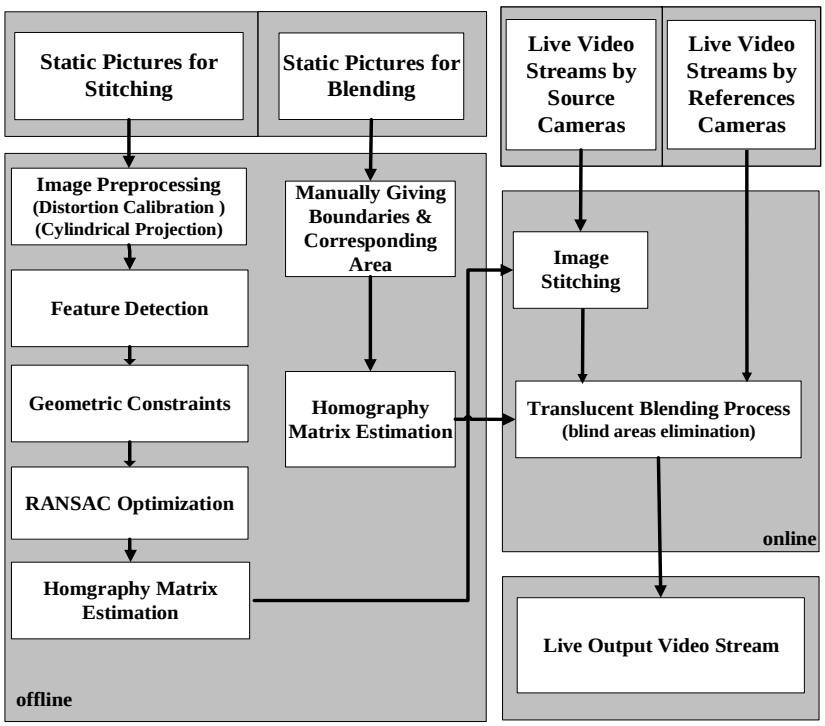


Fig. 2 Flow diagram of intelligent surveillance systems

The gist of the proposed algorithm is to use image registration methods to infer the relative pose among multiple cameras, stitch images possessing overlapped regions to each other, and blend images located in the blind areas of the stitched images. The

significant process makes use of feature correspondences extracted and identified in these images.

The image stitching part of the whole algorithm includes: (1) feature correspondence detection: detect feature in any two cylindrical images, match them to obtain feature correspondences and then estimate the fundamental matrix for describing the pixel relationship between images, (2) filter the outliers by RANSAC optimization and epipolar geometry, (3) estimate homography matrices between each pair of images, (4) stitching image pairs by their corresponding homography matrices. As to the blind areas which are expected to be translucentized, utilizing feature correspondence is not feasible because mostly objects seen by each pair of front-and-host cameras are not visually prominent enough to be automatically detected with any existing algorithms. Instead, manually assigning corresponding areas to derive the relative pose which is fixed between cameras is more feasible because such installation procedure only has to be executed once before the system is deployed. Besides, the blending boundaries are also artificially given. The reason is that blind area can be of any form of shapes which result in difficulties in being fully automated. The steps of image stitching process are described in more detail below.

As shown in Fig. 2, the goal of offline installation is to produce multiple homography matrices for image stitching and image blending. Only static images are required for this stage. Besides, since online deployment only moves pixels based on the obtained matrices, our proposed system is not computationally expensive and could be easily deployed on most of the existing hardware.

3. Technical Rationale

As mentioned before, the detail of our proposed algorithm for achieving see-through function is described as follows:

A. Cylindrical Projection

In order to stitch the multiple images together into a panoramic one, cylindrical image stitching is a much easier process than stitching images in rectangular coordinates. Therefore, each of undistorted image needs to be projected onto the cylindrical image [8]. Fig. 3(a) shows the relationship of rectangular image and cylindrical image. In order to model the displacement of each pixel (x, y) in the cylindrical image, the diagrams of the horizontal and vertical projection are defined as Fig. 3(b), and Fig. 3(c), respectively.

Figure 3 is wrapped on a cylinder with the appropriate radius that is defined the focal length f_c of each source camera, then expected length L of the cylindrical image is

$$L = 2\pi f_c \tag{1}$$

From Fig. 3(b) and 3(c), the image coordinates (x, y) can be easily determined to project coordinates on the cylinder (x', y') as

$$(x', y') = (f_c \sin\theta,\ f_c \tan\phi) \tag{2}$$

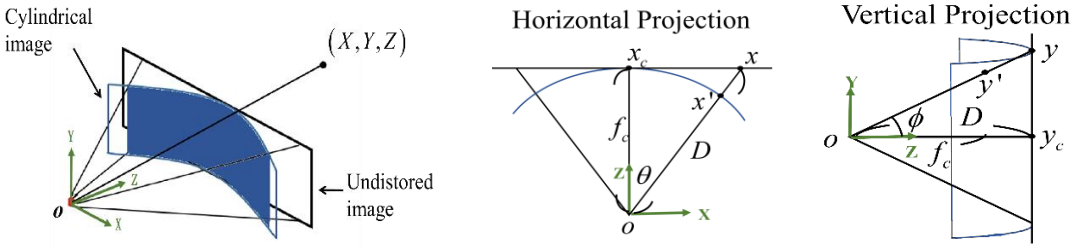
where

$$\theta = \tan^{-1}\left(x - x_c \big/ f_c\right),\ \phi = \tan^{-1}\left(y - y_c \big/ D\right) \tag{3}$$

(x_c, y_c) is defined from the center coordinate of the image; θ is the angle between f_c and $\overline{ox_c}$ on the XZ-plane; ϕ is the angle between D and $\overline{oy}$ on the YZ-plane; The length can be expressed as

$$D = \sqrt{f_c^{\,2} + (x - x_c)^2} \tag{4}$$

It is evident that the original undistorted image can be merged by a simple translation to project on the cylindrical image.



(a) (b) (c)

Fig. 3 Cylindrical projection

Notice that the focal length f_c directly affects how the image would appear as well as how well the subsequent image will be.

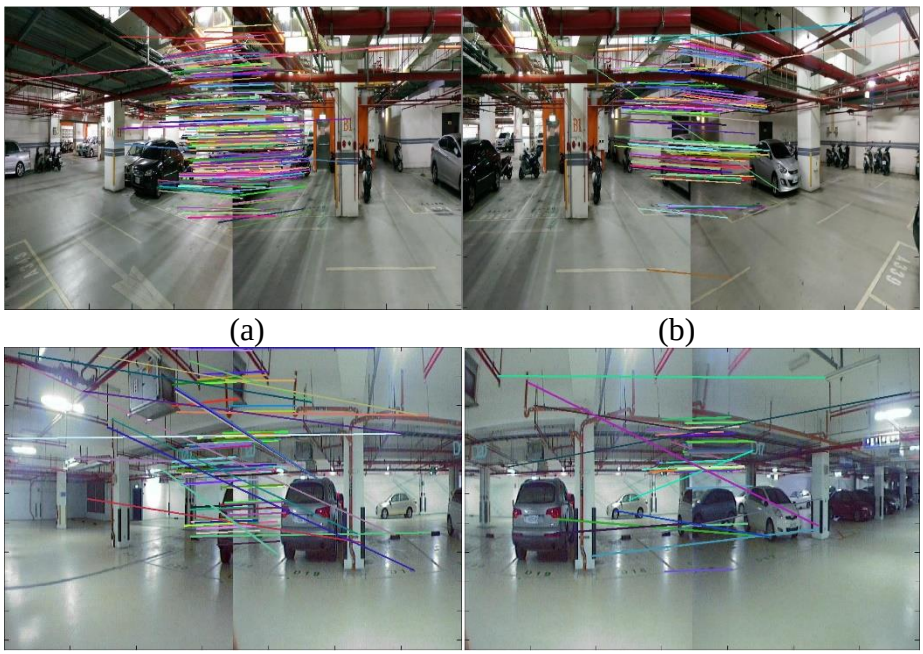
B. Feature Detection

In order to derive the relationship between two images which partially contain the same background, representative points could be extracted to locate the same objects and further derive if they were translated, scaled or rotated. Also, such features should

be detectable even with changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

SIFT (Scale-Invariant Feature Transform) scheme [9-11] is applied to detect and describe local features in images. It can robustly identify objects even among clutter and under partial occlusion because the SIFT feature descriptor is invariant to uniform scaling, orientation and partially invariant to affine distortion and illumination changes. Although SIFT is difficult to meet the requirement of real-time computation, in this paper, the feature detection only needs to be performed before the system is deployed. Therefore, SIFT, a powerful feature detection algorithm, is adopted even though it is not computationally efficient.

Matching SIFT feature descriptors in two images is a 2D search based on the similarity of the descriptors only. In Fig. 4 we show all the matches determined on two pairs of images of the front cameras. A visual inspection of the results could spot a few wrong matches. Such wrong matching results could be eliminated by using the information coming from geometrical constraint. That is to say, the locations of the feature correspondences in adjacent images are constrained by epipolar constraint. Fundamental matrix estimation provides such solution to better filter the outliers based on the translation vector and rotation matrix computation which is also based on feature correspondences.



(a) (b)

(c) (d)

Fig. 4 SIFT matching results

C. Refining false matches using geometric constraints

The false matches obtained in the initial matches may have an unanticipated geometry distribution in two images, and for a loose reject value their number may exist significantly more than the number of correct matches. To efficiently eliminate false matches, we utilize a search approach which employs epipolar geometry that is enclosed in a 3×3 matrix, of match pairs for removing false matches. Since the camera is uncalibrated (i.e., unknown intrinsic camera parameters); therefore, we need to apply the tentative point correspondences between two views to estimate a 3×3 singular matrix, called the fundamental matrix $F \in \mathbb{R}^{3\times3}$. The fundamental matrix is the algebraic representation of epipolar geometry. In other words, the relationship between two images of the same scene can indeed be described by a fundamental matrix that constrains where the projection of points from the scene can occur in both images. Considering the projection of a scene matching point into one of the images, the corresponding point in the other image is constrained to a line, reducing the search area, and allowing for the detection of false correspondences. Hence, the relation between corresponding image points which the fundamental matrix represents is referred to as epipolar constraint.

We now define the epipolar geometry between a pair of images on the basis of left-and-right configuration, as shown in Fig. 5. Let O and O' be a pair of pinhole cameras (i.e., optical centers) in 3D space. Let p and p' be the each pair of corresponding image points through cameras O and O' of a 3D point P in images $\prod$ and $\prod'$ respectively. The epipolar constraint [12] can be defined as

$$p'^{T} Fp = 0 \tag{5}$$

where F is the fundamental matrix [13]; p and p' the projected 2D feature locations in two images.

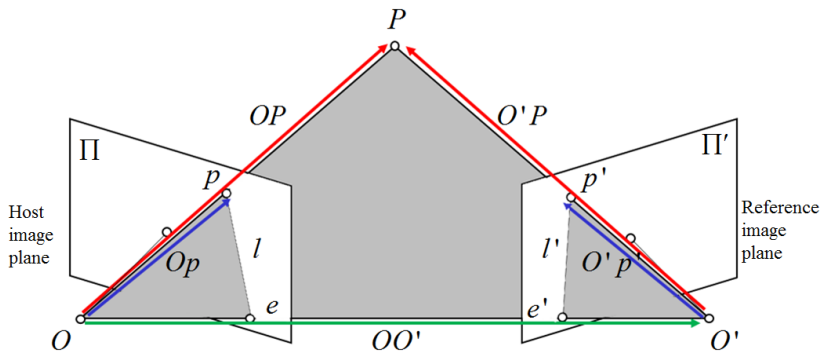


Fig. 5 Stereo configuration

Geometrically, it is easy to notice that points O, O', p, p' and P all lie on the same plane (the epipolar plane). Hence,

$$O'p' \cdot \left(O'O \times Op\right) = 0 \tag{6}$$

Denote the camera's intrinsic matrix as K and K' for the two cameras; we can convert p and p' from pixel coordinates to real-world coordinates as $K^{-1}p$ and $K'^{-1}p'$. If we denote the movement of the camera between the two consecutive frames as rotational matrix $R \in \mathbb{R}^{3\times3}$ and translation vector $T \in \mathbb{R}^{3\times1}$, then quite obviously $O'O = T$ and Op expressed in the primed frame is then

$$\begin{aligned} &Op = RK^{-1}p + T, \\ &\left(K'^{-1}p'\right)^{T}\left(T \times Op\right) = 0 \end{aligned} \tag{7}$$

Clearly, this can be equivalently expressed as

$$p'^{T}\left(K'^{-T}T \times RK^{-1}\right)p = 0 \tag{8}$$

Therefore, the fundamental matrix can be described as $F = K'^{-T}T \times RK^{-1}$. Each corresponding feature pair identified then provides one constraint on F. With enough such feature correspondences, we can solve for F, and hence to find the relative pose (translation vector T and rotation matrix R). The process can be based on either the Calibrated 5-Point Algorithm of Nister [14], the Normalized 8-Point Algorithm of Hartley [15], or the 7-Point Algorithm [16].

While the names of the inference algorithms refer to the minimum numbers of pairs of matched image features in two views that are needed for deducing the camera's motion parameters, in reality, we can often match significantly more features than just five or eight. Furthermore, matching results are necessarily imprecise due to noise and image quantization, and catastrophic failure in erroneous pairing assignments does happen occasionally. To improve the robustness in camera motion inference, we use a nonlinear selection and filtering strategy called RANSAC [17] to better condition the feature matching process by imposing the epipolar constraints. The essence of the RANSAC selection process is to repeatedly apply 5-point, 8-point or 7-point computation to a small subset of randomly selected "seed" correspondences, in hope that at least one set of seed correspondences were not corrupted by bad, outlier correspondences. Such outlier-free seeds will lead to a F that produces small residual errors in $\left|p'^{T}Fp\right|$ for most inlier correspondences. Hence, outlier correspondences are

identified and filtered out as those with unreasonably large residual errors. The best seed selection corresponds to the one that produces the minimum median $\left|p'^{T} Fp\right|$ residual error. One final run of the 8-point algorithm is then performed using all inlier correspondences in a least square sense. Then, in application is to compute a homography matrix between images using those extracted matches via RANSAC.

Figure 6 demonstrates that the proposed approach eliminates wrong matches by applying RANSAC which can reduce the number of wrongly matched correspondences which might potentially result in poor image registration result.

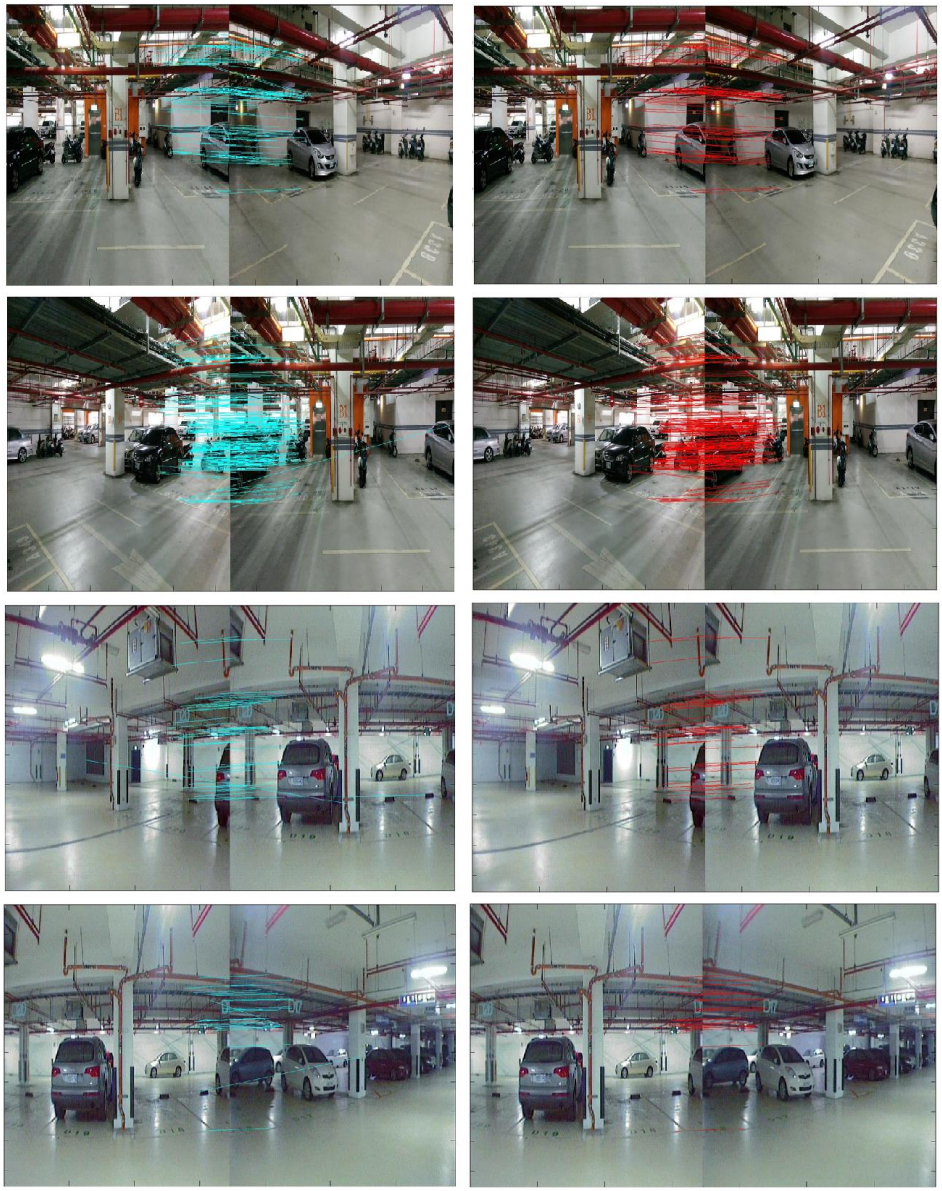


Fig. 6 SIFT matching results with RANSAC

D. Estimation of the homography matrix between each pair of images

Assuming that the inliers are obtained through the above-mentioned epipolar constraint, the geometrical relation between images to be stitched (for images adjacent to each other) could be described by homography [18] matrices estimated by the inliers between each pair of images. Since outliers are filtered by the epipolar geometry, the stitching effect is supposed to be promising. Even though motion parallax is inherently unavoidable due to the translation between cameras providing captured images, the side effect which is the artifact will not affect the judgment of users to instantly understand the monitored area.

As to the pixels to be translucent zed, the above-mentioned approach could not be guaranteed to perfectly work because the translation between front-and-host cameras varies according to the space configuration. If the spatial disparity between them is significant, no existing feature detection algorithm could work properly and stably. To design a fully deployable system which could function in arbitrary scenarios, the approach of this paper is to manually give a corresponding area, which could be described by four correspondence points, seen by both cameras.

The equation $x' = Hx$ indicates the position of a pixel before and after transformation. A 3x3 homography matrix could be determined if a set of four 2D to 2D feature correspondences ($x_i \leftrightarrow x_i'$) is given. By rearranging, the equation is $Ah = 0$ where h contains the 9 elements of. Each H feature correspondence gives rise to two independent equations. The dimension of A is therefore $2n \times 9$. If four feature correspondences are given, A has rank 8 and thus has a 1-demensional null-space which provides a solution for h, such a solution h can only be determined up to a non-zero scale factor. However, H is in general only determined up to scale, so the solution h gives the required H. A scale may be arbitrarily chose for h by a requirement on its norm such as $\|h\| = 1$. In our application, mostly the images to be stitched provide $n >> 4$. As to the translucentization part, only four sets of correspondences are needed and they could be manually given based on the object seen by both cameras in our application. The solution of h could be estimated by obtaining the Singular Value Decomposition (SVD) [19] of A and the unit singular vector corresponding to the smallest singular value is the solution h. Specifically, if $A = UDV^t$ with D diagonal with positive diagonal entries, arranged in descending order down the diagonal, then h is the last column of V.

E. Image stitching process

After a homography matrix is estimated for source image pair [20], image stitching could be performed. One of the images is transformed according to the corresponding homography matrix, and then this transformed image can be blended with the other input image when keeping the coordinate constraints of both images [21].

The purpose of the stitching is to produce a resulting image where no transition can be seen between the original source images. In general, a simple linear method for average weighted blending in the overlapping regions is applied. As shown in Fig. 7, let the I_A and I_B be two input images; l_1 and l_3 are the edge of respective images. Suppose that composite image I_S is created from cylindrical aligned images I_A and I_B, then

$$I_S = \alpha I_A + (1-\alpha) I_B \tag{9}$$

where $\alpha = {p_x^s - l_2}/{l_3 - l_2}$ and (p_x^s, p_y^s) is the coordinates of the composite stitch image I_S.

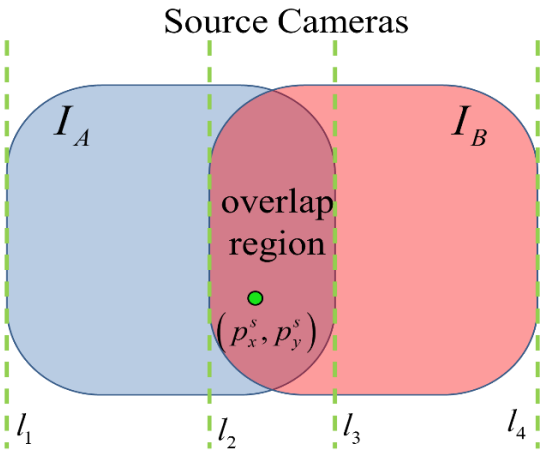


Fig. 7 Image stitching illustration

F. Translucent blending process for visualizing blind areas elimination

After stitching process is done, we apply the base for rectangular color blending process with the corresponding pixels seen by the front cameras to see-through the blind areas from occluding pillar. The range of the blind area is defined by the size of rectangular, i.e. visualized region, as shown in Fig. 8. The translucent blending weight β is adjusted to use more color from the front image I_F (reference view) close to the center of the rectangular, more color from the host image I_H (source view) away from

the center of rectangular and toward the edge of the rectangular of confusion. The mixture is controlled by a transparency parameter λ that makes the occluding pillar more or less apparent inside the region of confusion.

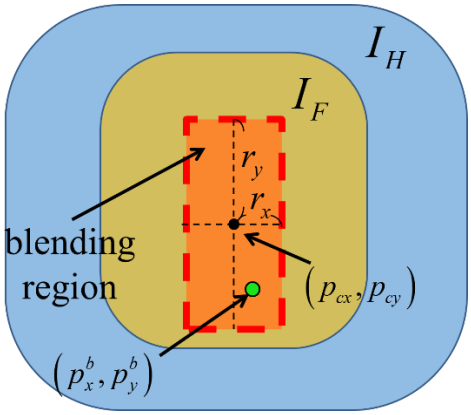


Fig. 8 Diagram of the image blending between the host camera and front camera

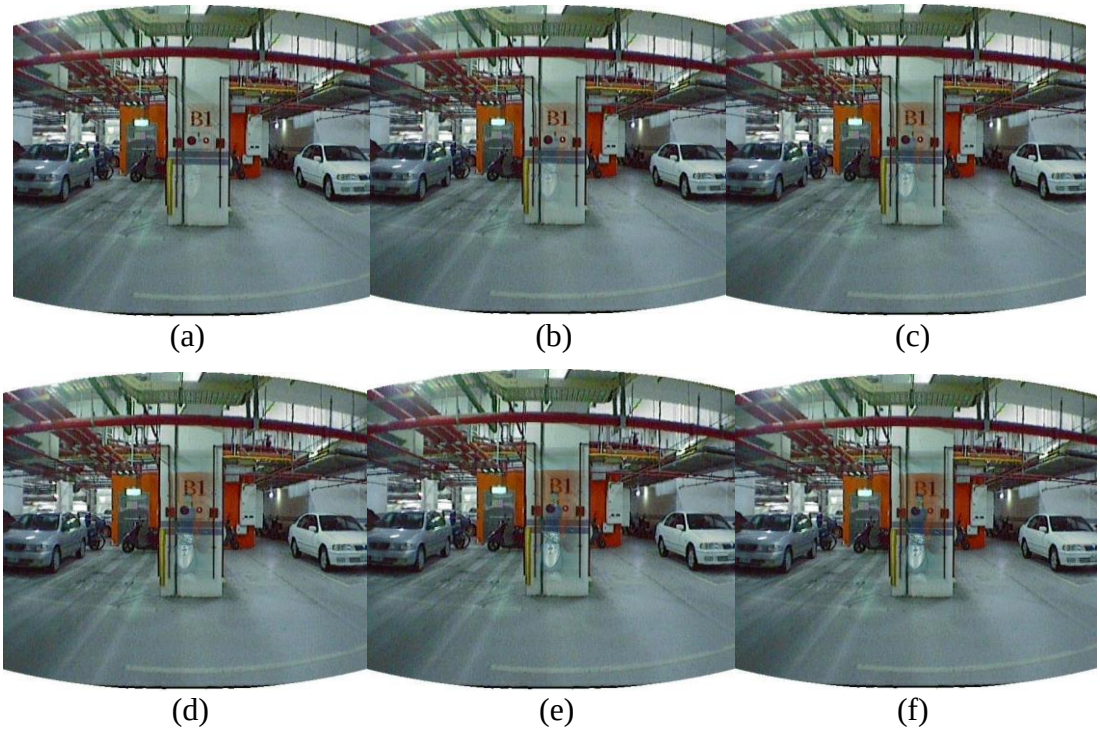
(p_x^b, p_y^b) is the coordinates of the composite blend image I_B; r_x and r_y are the distance between the center and edge of rectangular, respectively. Now a blending method is given by

$$I_B = \beta I_H + (1-\beta) I_F \tag{10}$$

where β is a translucent blending weight that can determined from intrinsic distance of the defined visualizing blind region.

$$\begin{aligned} &\beta = \frac{\psi}{\sqrt{r_x^2 + r_y^2}} \mathrm{e}^{\lambda(\sqrt{r_x^2+r_y^2}-1)}, \\ &\text{subject to } \left\{ \left| p_x^b - p_{cx} \right| \le r_x \cap \left| p_y^b - p_{cy} \right| \le r_y \right\} \end{aligned} \tag{11}$$

and (p_{cx}, p_{cy}) is the center of the rectangular area and (p_x^b, p_y^b) is the coordinates of the composite stitch image I_B. ψ is $\min\left(r_x - \left|p_x^b - p_{cx}\right|, r_y - \left|p_y^b - p_{cy}\right|\right)$, and λ is the parameter of transparency that can decide the degree of transparency. Then the corresponding cameras behind the occluding pillar provide pixels for translucentizing. Figure 9 displays the degree of see-through for different transparency parameter. As can be seen that with increased transparency, the occluding pillar become less visible.



(a) (b) (c)

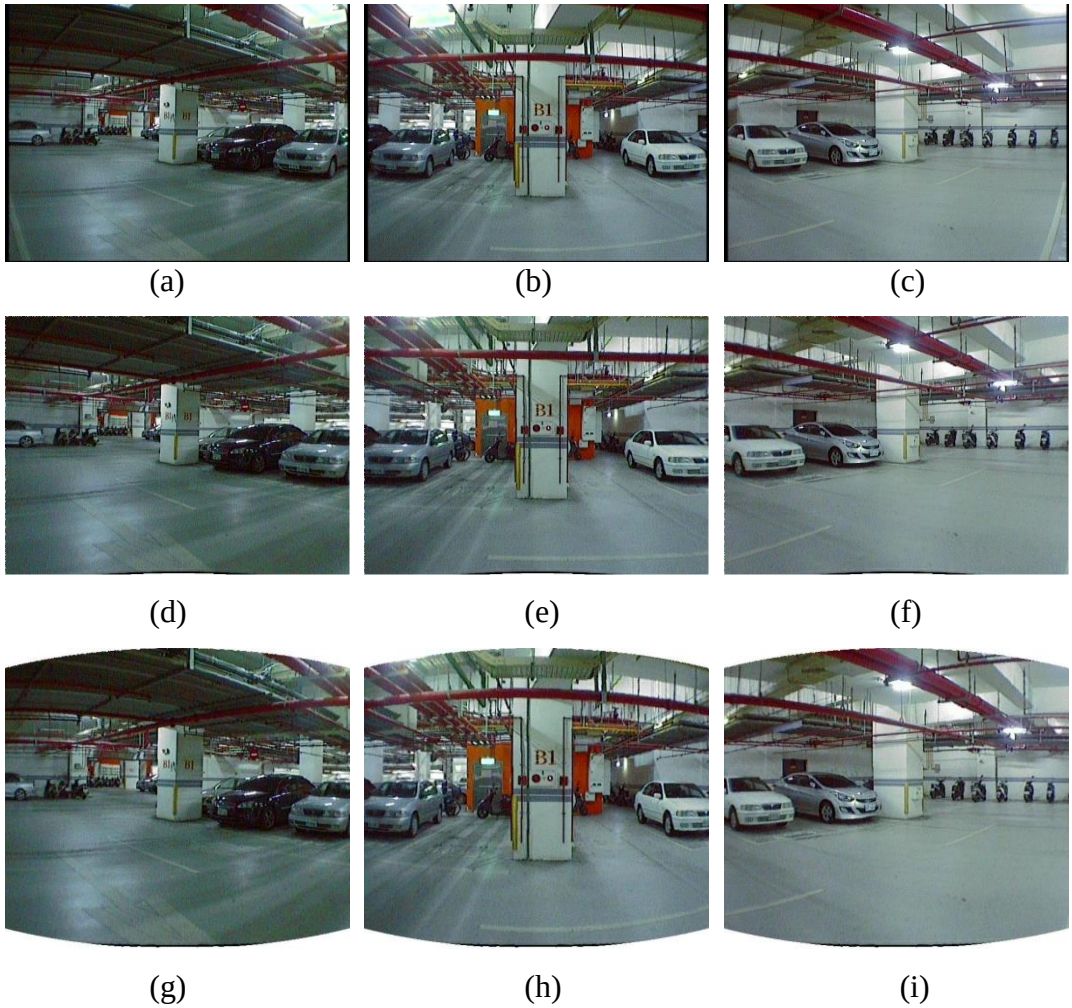
(d) (e) (f)

Fig. 9 Compared with different transparency. (a) $\lambda = 0$; (b) $\lambda = 1$; (c) $\lambda = 2$; (d) $\lambda = 3$; (e) $\lambda = 4$; (f) $\lambda = 5$

4. Experimental Results

According our proposed intelligent surveillance systems configuration of Fig. 1, we have installed multiple cameras inside parking lot to prove the feasibility of our system. A video capturing system was also used in order to store the capture video streams for later analysis. The input image sequences from the vision sensors are grabbed with the 720*480 resolution and 8 bits in each color channel (RGB).

First, the process of image preprocessing is best understood by following examples as shown in Fig. 10. This process includes the image distortion and cylindrical projection, and the quality of preprocessing directly affects the subsequent results. Figure 10(a)-10(c) are the original source view from the host cameras, respectively. Figure 10(d)-10(f) generates an undistorted copy of each original image. It then converted these undistorted images to cylindrical form by inverse warping the cylindrical coordinates as illustrated in Fig. 10(g)-10(i).



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 10 Image preprocessing, ((a),(b),(c) are the original image from source cameras, (d),(e),(f) are the undistorted images, (g),(h),(i) are the cylindrical images)

Based on the SIFT-based matching, homography matrix estimation, RANSAC optimization and epipolar geometry, the program could then be run iteratively to expand the field of view of the resulting image to integrate all of the cylindrical source images, as shown in Fig. 11.

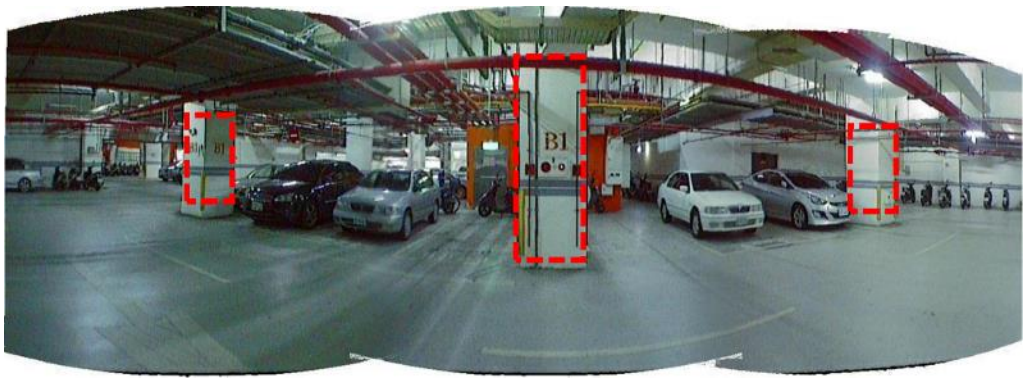


Fig. 11 Stitching output frame

In Fig. 11, we can easily observe that the front larger monitor blind areas (red rectangle regions) were caused by each occluding pillar at parking lot. Therefore, we then apply a translucent blending scheme to eliminate the visually blind area. As seen in Fig. 12, the blind area from each occluding pillar becomes slightly visible, i.e. the blind area is translucent.

Fig. 12 See-through the blind areas

Figure 12 shows that by using our proposed image fusion algorithm, it is possible to make the front camera (reference view) "semi-transparent" in the host video frame (source view); hence, it provides a much better visual monitor management and improve the safety of parking lot in underground environments.

5. Conclusion

We propose a feasible computer-vision-based parking lot intelligent surveillance systems which utilizes image stitching and translucent blending processes to integrate multiple images and see-through the blind areas from the occluding pillar. The benefit is that not only the viewable range is augmented but also the whole field could be viewed in a single image. As long as the monitored areas are overlapped, they could then be intelligently fused into one image. Any person including monitors could easily understand what is going on in the monitored areas with a simple glance. In the future, the see-through intelligent surveillance system could be furthered by introducing TOF cameras because since the 3D structure of the monitored area could be reconstructed, the 3D relations between images could be obtained to form 3D models of the monitored areas. The parallax resulting from the spatial disparity between cameras could be then completely eliminated.

Acknowledgment

This research was sponsored by Ministry of Science and Technology, Taiwan,

R.O.C., under the contract No.: 103-2218-E-035-008.

References

[1] J. E. Eck, R. V. Clark, and R. T. Guerette, "Risky facilities:crime concentration in homogeneous sets of establishments and facilities," *Crime Prevention Studies*, vol. 21, pp. 225-264, 2007.

[2] WakeUpWalMart.com, "*Crime and Wal-Mart - 'Is Wal-Mart Safe?'* An Analysis of Official Police Incidents at Wal-Mart Stores," Washington, DC, USA, 2006.

[3] W. T. Sung, A. H. Chen, D. C. Huang, and Y. H. Ju, "Multisensors realtime data fusion optimization for IOT systems," *IEEE Int. Conf. Systems, Man and Cybernetics*, San Diego, California, USA, pp. 2299-2304, 2014.

[4] Y. Kameda, T. Takemasa, and Y. Ohta, "Outdoor see-through visionutilizing surveillance cameras," *3rd IEEE and ACM Int. Symposium on Mixed and Augmented Reality*, Arlington, VA, USA, pp. 151-160, 2004.

[5] P. C. Barnum, Y. Sheikh, A. Datta, and T. Kanade, "Dynamic seethroughs: synthesizing hidden views of moving objects," *8th IEEE Int. Symposium on Mixed and Augmented Reality*, Orlando, FL, USA, pp. 111-114, 2009.

[6] C. Mei, E. Sommerlade, G. Sibley, P. M. Newman, and I. D. Reid, "Hidden view synthesis using real-time visual SLAM for simplifying video surveillance analysis, " *IEEE Int.Conf. Robotics and Automation*, Shanghai, China, pp. 4240-4245, 2011.

[7] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052-1067, 2007.

[8] B. S. Kim, S. H. Lee, and N. I. Cho, "Real-time panorama canvas of natural images," *IEEE Trans. Consumer Electronics*, vol. 57, no. 4, pp. 1961-1968, 2011.

[9] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91-100, 2004.

[10] Z. Hua, Y. Li, and J. Li, "Image stitch algorithm based on SIFT and MVSC," *IEEE Int. Conf. Fuzzy Systems and Knowledge Discovery,* Shandong, China, pp. 2628-2632, 2010.

[11] P. Xu, L. Zhang, K. Yang, and H. Yao, "Nested-SIFT for efficient image matching and retriecal," *IEEE MultiMedia*, vol. 30, no. 3, pp. 34-46, 2013.

[12] O. Chum, T. Werner, and L. Matas, "Two-view geometry estimation unaffected by a dominant plane," *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Washington, DC, USA, pp.772-779, 2005.

[13] R. Hartley, and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, MA, 2003.

[14] D. Nister, "An efficient solution to the five-point relative poseproblem," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756-770, 2004.

[15] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580-593, 1997.

[16] T. Hong and K. Nick, "Estimation of the fundamental matrix based on complex wavelets," *Int. Conf. Networking and Information Tech.*, Manila, Philippine, pp.350-354, 2010.

[17] M. Fischler, and R. Bolles, "RANdom sample sonsensus: A paradigmfor model fitting with applications to image analysis and automatedcartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, 1981.

[18] T. Liyoshi, and W. Mitsuhashi, "Homography-based image mosaicing for automatically removing pratial foreground object," *IEEE Int. Conf. Signal Proc. and Communication Syst.*, Gold Coast, Australia, pp. 1-9, 2008.

[19] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *J. Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, vol. 2, no. 2, pp. 205-224, 1965.

[20] S. Seo, S. Jeong and S. K. Lee, "Efficient homography estimation method for panorama," *IEEE Int. Conf. Frontiers of Computer Vision*, Incheon, Korea, pp. 209-212, Jan, 2013.

[21] Y. Xiong and K. Pulli, "Sequential image stitching for mobile panoramas," *Int. Conf. Information, Communications and Signal Proc.*, Macau, China, pp. 1-5, 2009.