

Bird's-Eye View Vision System for Vehicle Surrounding Monitoring*

Yu-Chih Liu, Kai-Ying Lin, and Yong-Sheng Chen**

Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

Abstract. Blind spots usually lead to difficulties for drivers to maneuver their vehicles in complicated environments, such as garages, parking spaces, or narrow alleys. This paper presents a vision system which can assist drivers by providing the panoramic image of vehicle surroundings in a bird's-eye view. In the proposed system, there are six fisheye cameras mounted around a vehicle so that their views cover the whole surrounding area. Parameters of these fisheye cameras were calibrated beforehand so that the captured images can be dewarped into perspective views for integration. Instead of error-prone stereo matching, overlapping regions of adjacent views are stitched together by aligning along a seam with dynamic programming method followed by propagating the deformation field of alignment with Wendland functions. In this way the six fisheye images can be integrated into a single, panoramic, and seamless one from a look-down viewpoint. Our experiments clearly demonstrate the effectiveness of the proposed image-stitching method for providing the bird's eye view vision for vehicle surrounding monitoring.

1 Introduction

The major aim of intelligent driving assistant systems focuses on collision avoidance. These systems achieve this goal by providing drivers with warning signals or visual information so that drivers can keep safe distances to obstacles. This kind of systems are very helpful, particularly when drivers want to maneuver their cars into garages, back into parking space, or drive in narrow alleys. Even though side-view and rear-view mirrors can help the driver to see areas behind, the ground areas all around the vehicle remains difficult to see. Recently, rear-view cameras become more and more popular because of their capability in reducing invisible areas behind the car. This motivates us to construct a vision system that can provide the driver with an image covering all the vehicle surroundings without blind spots.

Consider a bird's-eye view vision system that displays the image all around a vehicle from a look-down viewpoint. This kind of system can be easily constructed by placing a camera facing down on the top of the vehicle. However, this

* This work was supported in part by the Ministry of Economic Affairs, Taiwan, under Grant 96-EC-17-A-02-S1-032.

** Corresponding author.

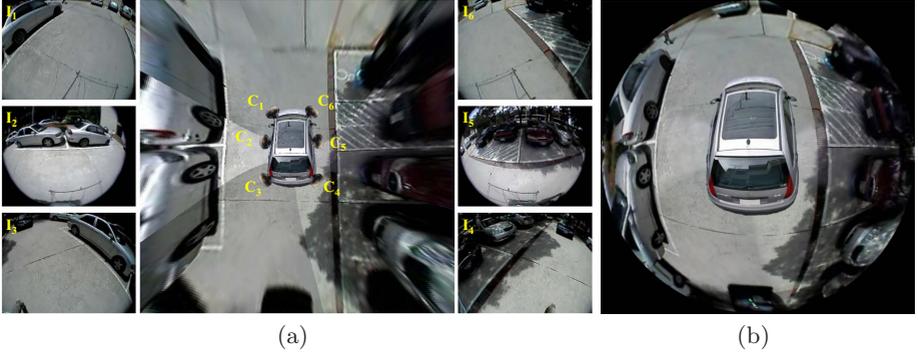


Fig. 1. The proposed vehicle surrounding monitoring system using multiple cameras. (a) Images I_1 , I_2 , ..., and I_6 captured from the corresponding cameras C_1 , C_2 , ..., and C_6 , respectively. (b) The bird's-eye view image synthesized from the six fisheye images.

approach is barely feasible because it is very dangerous to hang a camera high above a vehicle. The Matsushita company proposed an image synthesis display system [1] which uses several cameras around the vehicle and synthesizes the images captured from these cameras into a whole picture. Because they simply average the image pixel values in the overlapped regions, the ghost artifact is severe and the driver cannot know the actual location of the objects. Ehlgen and Pajdla mounted four omnidirectional cameras on a truck and find several subdivision ways for dividing the overlap regions to construct the bird's-eye view image [2]. No matter how to choose the subdivision way, there are some blind spots in the subdivision boundaries.

In this paper we propose a novel method of synthesizing a seamless bird's-eye view image of vehicle surroundings from six fisheye cameras mounted around the vehicle. As shown in Fig. 1(a), these six fisheye cameras cover the whole surrounding area and have some overlapped areas between adjacent cameras. Our goal is to stitch the six distorted images captured from the fisheye cameras and provide an image of vehicle surroundings from a bird's eye viewpoint, as shown in Fig. 1(b).

2 Vehicle Surrounding Monitoring System

Flowchart of the proposed vehicle surrounding monitoring system is depicted in Fig. 2. We first calibrate the fisheye cameras to obtain their intrinsic parameters. Then these parameters can be used for correcting distortion and transforming the fisheye images into perspective ones. These perspective images are registered to a ground plane by using planar homographies. The objects coplanar with the ground plane can be registered perfectly in this way. For 3-D objects, some advanced registration process is necessary. We select a seam between each pair of adjacent images according to the residual error obtained in previous calibration

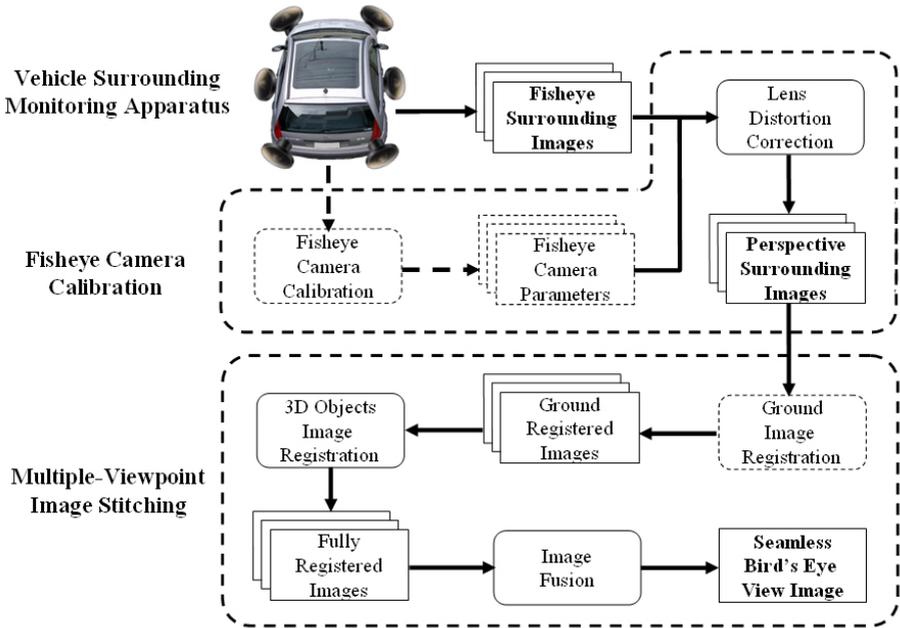


Fig. 2. Flowchart of the proposed vehicle surrounding monitoring system

step and register image points on this seam by using a dynamic programming approach. Then we use Wendland functions to smoothly propagate the registration results from the seams to other areas in the whole synthesized image. Finally, we use weighted blending to reduce brightness discrepancy and obtain the seamless bird's eye view image. In the following we presents these procedures in detail.

2.1 Fisheye Camera Calibration

Here we propose a simple and accurate method for calibrating fisheye cameras, which only requires the fisheye camera to observe a planar pattern in a few orientations. We use the Field of View (FOV) [3] as the model of radial distortion introduced by fisheye lenses based on the principle of fisheye projection. With the initial values of the radial parameters as well as other intrinsic parameters, we can dewarp a fisheye image into a perspective one for a virtual perspective camera. Zhang's method [4] is then used to calibrate the camera parameters of the virtual perspective camera and the residual error between the model-predicted and the measured positions of the feature points on the planar pattern can be regarded as an evaluation criterion of the accuracy of parameters in the FOV model. Levenberg-Marquardt method [5] is then applied to optimize the values of these parameters by minimizing the total residual error of all feature

points. Finally, we rectify the fisheye images into perspective ones by using the FOV model with the estimated parameter values.

2.2 Planar Image Alignment with Homography

Since the goal of this work is to create a whole bird’s eye view of the monitoring area, we need to stitch together the images captured by all the fisheye cameras around the vehicle, after the fisheye images are rectified into perspective ones with calibrated camera parameters. Suppose the partial scene image I_i is captured by the camera C_i . By specifying the feature points in a reference image I_r of the ground plane and their corresponding points in image I_i , we can establish a homography $\mathbf{H}_{r,i}$ relating the homogeneous coordinate \mathbf{p}_i of each point in I_r and the coordinate \mathbf{p}_r of its corresponding point in I_r :

$$\mathbf{p}_r = \mathbf{H}_{r,i}\mathbf{p}_i . \quad (1)$$

In this way all images can be transferred into the same coordinate system and then be registered together. It is good to stitch the images of the object coplanar with the ground plane, since the perspective projection are preserved for the objects located at the ground plane in world coordinate. On the other hand, it is problematic to stitch the images of the objects not coplanar with the ground plane, since they are not consistent with the homography obtained in the aspect of the ground plane. Therefore, advanced processing of registration is necessary to deal with this problem. Thereinafter, we call the objects not coplanar with the ground plane “the non-ground-level objects.”

2.3 Optimal Seam Selection

Once having registered all images on a planar surface, the next step is to decide which pixels should contribute to the final composite. We want to create a composite with as less artifacts as possible. Toward this goal, we select an optimal seam which lessens mis-registration and this involves how to decide the placement of seam to avoid the discontinuities along the seam and to provide a smooth transition between adjacent images.

In our application, the source images are captured by the fisheye camera and they have severe radial distortion. The accuracy of the camera parameters directly affects the image alignment result. Therefore, we use this property to select the seams. However, we can obtain the residual error only for the feature points in the camera calibration process. In order to obtain the pixel-by-pixel measurement which usually increase with its radial radius, we use a quadratic function E_i to model the distribution of the residual error over the whole image captured by camera C_i :

$$E_i(x, y) = a_i x^2 + b_i xy + c_i y^2 + d_i x + e_i y + f_i , \quad (2)$$

where a_i , b_i , c_i , d_i , e_i and f_i are the six coefficients of the quadratic function which can be estimated by using the least squares approach. We can formulate

the cost function according to the residual error for the optimal seam selection as follows:

$$C(s_{ij}) = \sum_{\mathbf{p}} E_{L(s_{ij}, \mathbf{p})}(\mathbf{H}_{L(s_{ij}, \mathbf{p})r} \mathbf{p}) \quad , \quad (3)$$

where \mathbf{p} is a pixel in the composite, s_{ij} is a seam in the overlapped region of registered images I_i and I_j which divides the whole image into part i and part j labelled by $L(s_{ij})$. The error map E_i is modelled by a strictly monotonic increasing function, hence we can obtain an optimal seam through the greedy approach. For this reason, the optimal seam selection can be reduced to the pixel selection problem. The labelling of the source image for the pixel \mathbf{p} can be decided by

$$L(s_{ij}, \mathbf{p}) = \arg \min_k E_k(\mathbf{H}_{kr} \mathbf{p}) \quad . \quad (4)$$

In this manner, we can create the image mosaic by choosing one point with least residual error from among the possible viewpoints for each pixel of the image mosaic. Once every pixel in the composite have decided which source images it comes from, the optimal seam can be determined. We can stitch images along these optimal seams to create a composite that has less mis-registration caused by the distortion.

2.4 Seam Registration Using Dynamic Image Warping

Since we use the optimal seam approach proposed in the previous section to create the composite, the blurring, ghosting, and the artifacts (for example, mis-registration and exposure differences) can only lie along the seam. For the smooth transition from one view to another, we only need to find the right correspondences on the seam. Dynamic time warping (DTW) is an efficient technique for solving this problem. Instead of aligning data sequence in the temporal domain, we register the seams by aligning pixel data of two seams in the spatial domain.

Suppose we have two image I_a and I_b to be stitched together on the seam S_{ab} . Since the seam S_{ab} lies in the overlapped region of I_a and I_b , we can obtain two series of pixel data A and B which come from I_a and I_b along S_{ab} :

$$A = a_1, a_2, \dots, a_i, \dots, a_n \quad (5)$$

$$B = b_1, b_2, \dots, b_j, \dots, b_n \quad , \quad (6)$$

where n is the length of S_{ab} . We use the proposed dynamic image warping (DIW) method, described in the following, to align two sequences. We construct an $n \times n$ matrix \mathbf{C} and each element $C(i, j)$ corresponds to the alignment between the nodes a_i and b_j . A warping path W , as shown in Fig. 3, is a set of matrix elements that describes a mapping between A and B :

$$W = w_1, w_2, \dots, w_k, \dots, w_m \quad n \leq m \leq 2n - 1 \quad , \quad (7)$$

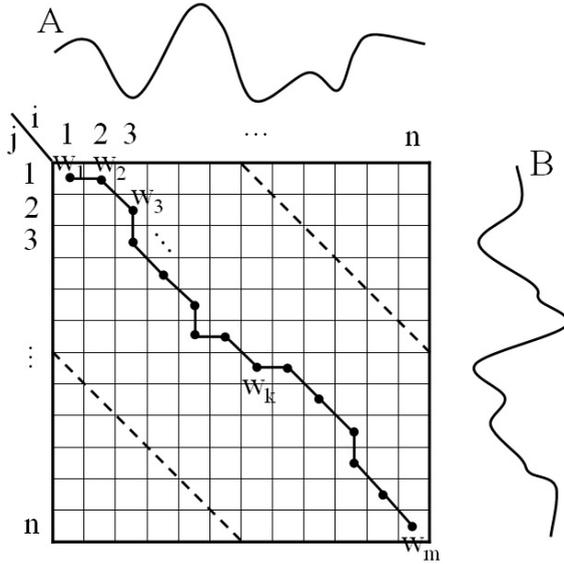


Fig. 3. Warping path of dynamic image warping. The warping path starts from $w_1 = (1, 1)$ and ends at $w_m = (n, n)$. The element (i, j) in the warping path describes that the node a_i should be aligned to the node b_j .

where $w_k = C(i, j)$ is the k -th element of W , and m is the number of elements of W . The matrix C is also the cost function for DIW defined as follows:

$$C(i, j) = D(a_i, b_j) + H(a_i, b_j) + \min \begin{cases} C(i, j-1) + p_v \\ C(i-1, j-1) \\ C(i-1, j) + p_h \end{cases}, \quad (8)$$

where p_h and p_v are the standard deviation of data sequence A and B and $D(a_i, b_j)$ is the difference measurement between data points a_i and b_j . This difference consists of two terms, the difference of absolute values and the difference of first derivative angles:

$$D(a_i, b_j) = d(a_i, b_j) + d'(a_i, b_j), \quad (9)$$

where

$$d(a_i, b_j) = (a_i - b_j)^2 \quad (10)$$

$$d'(a_i, b_j) = (\arctan(a'_i) - \arctan(b'_j))^2. \quad (11)$$

The idea of $d'(a_i, b_j)$ term inherit from the Derivative Dynamic Time Warping (DDTW) [6] proposed by Keogh et al. Besides, the seam registration result in current frame may differ from that obtained in the previous frame and results

in jittering effect. Therefore, we take the seam registration results of a sequence of images into consideration in $H(a_i, b_j)$ term:

$$H(a_i, b_j) = s ((i - h_b(f, j))^2 + (j - h_a(f, i))^2) \quad , \quad (12)$$

where $h_a(f, i)$ is the average index of the corresponding point in data series B to the j -th point in data series A in f previous frames, $h_b(f, j)$ is the average index of the corresponding point in data series A to the j -th point in data series B in f previous frames, and s is a scalar. As s increases, the warping path in current frame emphasize the history of the seam registration. Once the term $H(a_i, b_j)$ has added to the cost function, we can obtain a smooth registration result between consecutive frames.

2.5 Image Deformation Using Wendland ψ -Function

Once we have aligned the seams of the adjacent images, the next step is to propagate this alignment result to the rest of the images. This is a typical problem of the image deformation that involves how to produce an overall deformation field to the whole images and preserve the known correspondence as much as possible. The linear deformation method can interpolate smooth surfaces to maintain overall characteristics, but it is difficult to keep landmarks in position. The major approaches for image deformation are based on the radial basis functions (RBFs). Selection of the type of the RBFs depends on the tradeoff between the overall characteristics such as the smoothness and the locality of the transformation function.

In general, image deformation is accomplished by the transformation function $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. The interpolation transformation function $T(x)$ must map the landmark $\mathbf{p}_i = (p_{ix}, p_{iy}) \in \mathbb{R}^2$ in the source image to its landmark $\mathbf{q}_i = (q_{ix}, q_{iy}) \in \mathbb{R}^2$ in the target image:

$$T(\mathbf{p}_i) = \mathbf{q}_i, \quad i = 1, \dots, n \quad , \quad (13)$$

where n is the number of the landmarks. The transformation functions in two coordinates are calculated separately:

$$T(\mathbf{p}_i) = (t_x(\mathbf{p}_i), t_y(\mathbf{p}_i)) \quad (14)$$

$$= (q_{ix}, q_{iy}), \quad i = 1, \dots, n \quad , \quad (15)$$

where t_x and t_y are the transformation functions in x and y coordinate respectively. In radial basis function approach, the transformation consists of two terms:

$$t(\mathbf{x}) = R_s(\mathbf{x}) + L_s(\mathbf{x}) \quad , \quad (16)$$

where $R_s(\mathbf{x})$ is the non-linear term of the weighted RBFs, and $L_s(\mathbf{x})$ is the linear term containing m bases of polynomials with degrees up to d :

$$R_s(\mathbf{x}) = \sum_{i=1}^n \alpha_i R(\|\mathbf{x} - \mathbf{p}_i\|) \tag{17}$$

$$L_s(\mathbf{x}) = \sum_{j=1}^m \beta_j L_j(\mathbf{x}) \tag{18}$$

where α_i and β_j are coefficients, $R(\|\mathbf{x} - \mathbf{p}_i\|)$ is the radial basis function centered at landmark \mathbf{p}_i , and its value only depends on the Euclidean distance from \mathbf{x} to \mathbf{p}_i . In order to preserve the overall smoothness as much as possible, the coefficients α_i is typically subject to the constraint

$$\sum_{i=1}^n \alpha_i L_j(\mathbf{p}_i) = 0, \quad i = 1, \dots, m \ . \tag{19}$$

A linear combination of the coefficients $\alpha = [\alpha_1, \dots, \alpha_n]^T$ and $\beta = [\beta_1, \dots, \beta_m]^T$ can be derived from the above equations:

$$\begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{q} \\ \mathbf{0} \end{bmatrix} \ , \tag{20}$$

where \mathbf{K} is the $n \times n$ sub-matrix in which $K_{ij} = R(\|\mathbf{p}_i - \mathbf{p}_j\|)$, and \mathbf{P} is the $n \times m$ sub-matrix in which $P_{ij} = L_j(\mathbf{p}_i)$.

Fornet et al. [7] use ψ functions of Wendland [8] as RBFs for elastic registration:

$$\psi_{d,k}(r) = I^k (1 - r)_+^{[d/2]+k+1}(r) \tag{21}$$

with

$$\psi(r) = (1 - r)_+^v \tag{22}$$

$$= \begin{cases} (1 - r)^v, & 0 \leq r < 1 \\ 0, & r \geq 1 \end{cases} \tag{23}$$

$$I\psi(r) = \int_r^\infty t\psi(t)dt \quad r \geq 0 \ . \tag{24}$$

Because of the images to be deformed in our application are 2-D images, we prefer the RBFs can create the smooth deformation, and choose $\psi_{2,1}(r)$ as our RBFs:

$$\psi_{2,1}(r) = (1 - r)_+^4 (4r + 1) \ . \tag{25}$$

2.6 Image Fusion Using Weighted Blending

After deforming the images using Wendland ψ -function, all edges cross seams could be stitched together on the seams. But the seams may look obviously since the images being stitched were not taken with the same exposure condition. In order to compensate the exposure difference, we use the bias and gain model to adjust the global exposure [9]:

$$I'_i = \alpha_i I + \beta_i \ , \tag{26}$$

where β is the bias and α is the gain. The bias and gain for each image can be obtained in the least squares manner:

$$E_i = \sum_j \sum_{\mathbf{P}} [\alpha_i I_i(\mathbf{H}_{ir}\mathbf{p}) + \beta_i - I_j(\mathbf{H}_{jr}\mathbf{p})]^2, \quad (27)$$

where image I_j is the adjacent image of image I_i , and \mathbf{p} is the image point in the overlap of image I_i and image I_j . In this approach, the images can be adjusted into similar exposure. But the seams may still be visible, since the bias and gain model only compensate the global exposure difference. Hence, a further smoothing process is necessary to eliminate the seams.

We use an image fusion method based on the weighted blending to smooth the seams away. We take the residual error of camera calibration as the weight to blend the images sources of the same pixels in the final composite. Though the weights of both image sources are equal along the seam, the weighting function is not continuous between the overlapping region and the non-overlapping region.

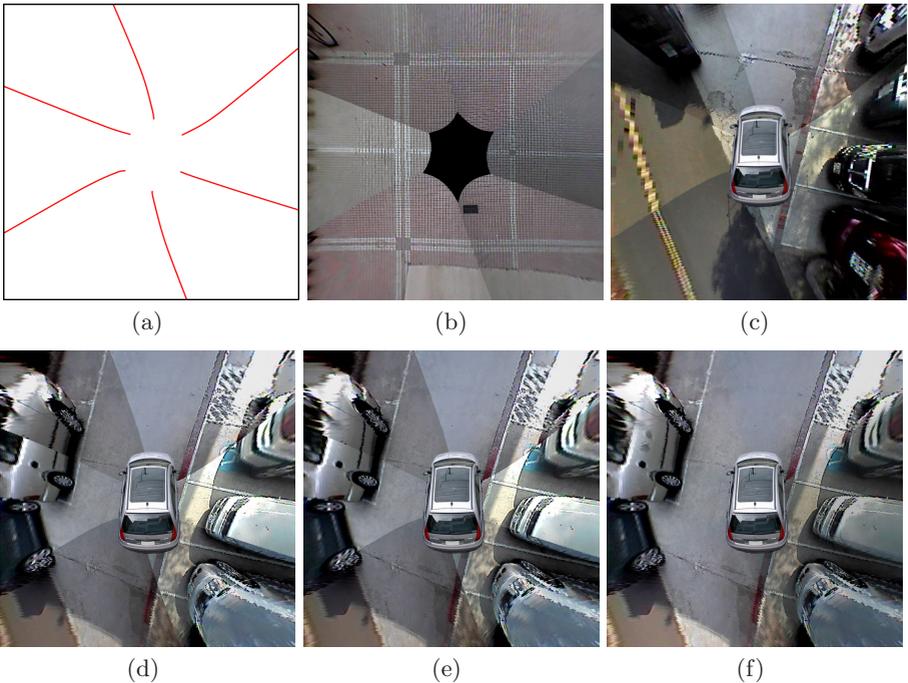


Fig. 4. The composite images from six fisheye cameras. (a) The optimal seams between adjacent images. (b) The composite image of the field for registration. (c) The composite image of ground-level objects. (d) The composite image of non-ground-level objects. (e) The composite image after the registration along the seam. (f) The composite image after exposure compensation and blending.

For this reason, we further take the minimum distance to image boundaries into weighting function. The proposed blending function is formulated as

$$I(p) = \frac{\sum_i E_i(\mathbf{H}_{ir}\mathbf{p})B(\mathbf{H}_{ir}\mathbf{p})I_i(\mathbf{H}_{ir}\mathbf{p})}{\sum_i E_i(\mathbf{H}_{ir}\mathbf{p})B(\mathbf{H}_{ir}\mathbf{p})} \quad (28)$$

$$B(u, v) = \left(1 - \left|\frac{2u}{width} - 1\right|\right) \left(1 - \left|\frac{2v}{height} - 1\right|\right), \quad (29)$$

where I is the image of the final composite, E_i is the calibration error function of the camera C_i , $width \times height$ is the resolution of images, and B is the weighting function with value 1 on the image center and 0 on the image boundary. We can reduce the seams in the integrated image by using this fusion method for intensity unification.

3 Experiment Results

3.1 Configuration of Image Acquisition

We mounted four fisheye cameras with 140-degree FOV on four corners of a vehicle and two fisheye cameras with 170-degree FOV on left and right side of the vehicle. In such a configuration, we can not only acquire images with higher resolution in the front and rear of a vehicle, but also minimize the blind spots using six fisheye cameras.

3.2 Results of Image Stitching

Which pixels contribute to the final composite are decided according to the residual error obtained in the calibration procedure. An optimal seam between adjacent images can thus be determined, as shown in Fig. 4(a). A composite image of the ground can also be created by stitching images along optimal seams,

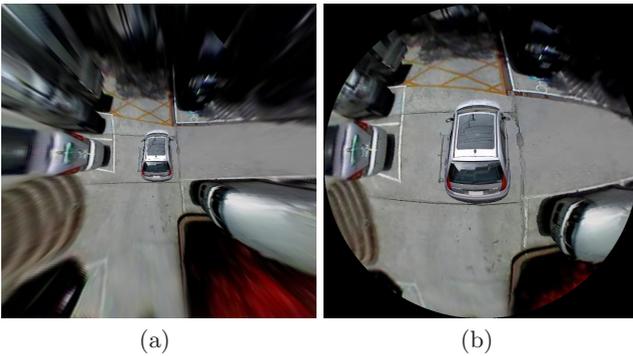
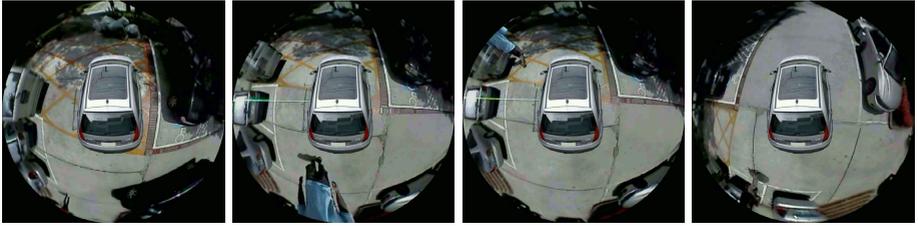


Fig. 5. Warped composites in fisheye view. (a) Original composite. (b) Virtual fisheye camera at a certain height.



(a)



(b)

Fig. 6. Image sequences of a car moving in reverse (a) and moving back into a parking space (b)

as shown in Fig. 4(b). As long as the cameras remain fixed, the perspective transformation from each rectified images to the bird's eye view and the optimal seam between each pair of adjacent images are invariant. Hence, the images of ground-level objects around the vehicle could be stitching into a bird's eye view, as shown in Fig. 4(c). But when a non-ground-level object exists in the surrounding scene, such as the car on the left side of the image shown in Fig. 4(d), and the image of this object cross the seam in the composite, a misalignment will occur on the seam. This misalignment is then removed by the DIW registration along the seam and the final composite image after the exposure compensation and weighted blending are shown in Figs. 4(e) and 4(f), respectively.

Some problems may occur in practice when all surrounding images are stitched in perspective presentation, such as the distortions of non-ground-level objects, low image resolution and amplified vibration in the farther surrounding area (see Fig. 5(a)). Warping the composite in a distortion level by placing the virtual fish-eye camera at certain height can cope with this problem, as shown in (Fig 5(b)). In Figs. 6(a) and 6(b), we show two sampled image sequences in which the vehicle is moving in reverse and moving back into a parking space, respectively.

4 Conclusion

In this paper, we have presented a driving assistant system which can provide the bird's eye view image of vehicle surrounding. The multiple fisheye cameras are mounted around the vehicle to capture images of the surroundings in all

orientations. All the fisheye images are rectified to the virtual perspective images by using the proposed calibration method and these rectified images are stitched into a seamless mosaic by using the proposed stitching method. The images coplanar with the ground are registered on a planar surface in a bird's eye view using perspective transformation, while the images of 3-D objects are aligned along the stitching seams using the dynamic image warping method. Finally, a seamless bird's eye view image is created by blending all registered images after exposure compensation.

With the support of this vehicle surrounding image, drivers can easily maneuver their vehicles by surveying the surrounding area of their vehicles on a single display. Beside, if a video recording system is mounted, it can also provide direct evidence when car accident happens.

References

1. Kato, K., et al.: Image synthesis display method and apparatus for vehicle camera. United States Patent 7,139,412 (2006)
2. Ehlgen, T., Pajdla, T.: Monitoring surrounding areas of truck-trailer combinations. In: Proceedings of the 5th International Conference on Computer Vision Systems (2007)
3. Devernay, F., Faugeras, O.: Straight lines have to be straight. *Machine Vision and Applications* 13(1), 14–24 (2001)
4. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11), 1330–1334 (2000)
5. More, J.: Levenberg–marquardt algorithm: Implementation and theory. In: Conference on numerical analysis, vol. 28 (1977)
6. Keogh, E.J., Pazzani, M.J.: Derivative dynamic time warping. In: SIAM International Conference on Data Mining (2001)
7. Fornefett, M., Rohr, K., Stiehl, H.S.: Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing* 19(1), 87–96 (2001)
8. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics* 4(1), 389–396 (1995)
9. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 674–679 (1981)