Chapter 14
Generating Reference Views of Traffic Intersection for Safe Driving Assistance

Jien Kato and Yu Wang

Abstract In this chapter, we address the problem of driving assistance along traffic intersections by providing drivers with additional visual information to expand their visual field. Our goal is to generate image stream of a virtual viewpoint which follows the host vehicle from a higher position, using images from multiple roadside cameras. Our approach is based on view morphing, but we extend it by integrating robust fundamental matrix estimation and sparse key point matching. This enables some tasks which previously rely on manual operation to be done automatically.

Keywords Driver visual field • Driving assistance • Image-based rendering (IBR) • Intersection assistance • Reference view • Vehicle blind spots

14.1 Introduction

Driving is becoming more and more stressful due to increasing traffic density. The situation seems to be more serious at intersections. According to the Annual Report 2007 from the National Police Agency of Japan, 46.3% of all traffic accidents in Japan occurred near intersections. In addition, a very large percentage of them happened either because of blind spots in the vehicles or inter-object occlusion due to traffic density at the intersections. These issues limit a driver's visual field. As a result, drivers find it more challenging to monitor their surroundings and forthcoming situations.

In the context of intersection assistance, Benmimnoun et al. presented a system [1] that utilizes intervehicle communication which updates the position measurements received from the onboard GPS and transmits all warning information to vehicles via

J. Kato (✉) • Y. Wang
Nagoya University, Nagoya 464-8603, Japan
e-mail: jien@is.nagoya-u.ac.jp; ywang@mv.ss.is.nagoya-u.ac.jp

J.H.L. Hansen et al. (eds.), *Digital Signal Processing for In-Vehicle Systems and Safety*,
DOI 10.1007/978-1-4419-9607-7_14,

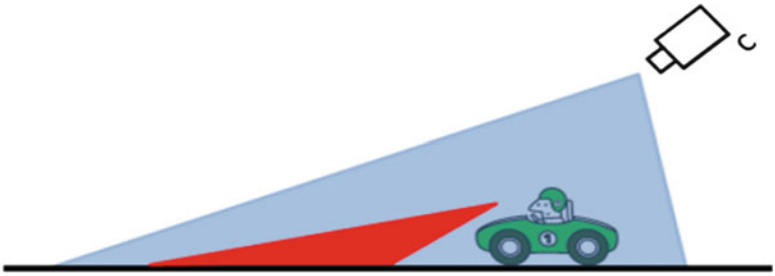

Fig. 14.1 Reference view

roadside-vehicle communication. With the use of a well-designed human–machine interface, their system could improve traffic safety to some extent by providing resulted warning signals to drivers. However, the final information which the driver receives is that of danger warning. Such information is helpful, but it is not easily accessible compared to what drivers directly obtain using their eyesight. Also, it is difficult and awkward to handle at times. In another work that pertains to image processing, Ichihara et al. extended their NaviView [2] to suit the environment at the intersections. But a simple affine transformation can only provide drivers with mirror image of views from a roadside camera. Though that system could extend the driver's visual field to the next intersection, it is still power-limited, and the information obtained is difficult to handle.

We believe that visual information is intuitive. It enhances the driver's ability in handling the surrounding situation. Note that a driver's visual field is limited by the vehicle's structure and inter-object occlusion. Broadening it will make it more efficient. With this in mind, we propose a method for generating a reference view (Fig. 14.1) which follows the vehicle's movement from a higher ground. The resulted view not only extends the driver's visual field but also provides information about the vehicle itself. This leads to the strengthening of robustness against forthcoming occlusions. Since this viewpoint is aligned with the vehicle's direction, then it has a direct relation with what the driver could see. Also, it is natural for the driver to handle such view as reference information.

To generate such kind of view, we expect to use roadside cameras located at the intersections. Nowadays, roadside cameras have been installed in places where traffic accidents occur frequently, especially at intersections. Using image data from those cameras will be cost-effective.

We choose image-based rendering (IBR) method to achieve our goal because it could provide a realistic novel view. Since the shapes in novel view have to be preserved, the IBR method based on implicit geometry, such as view morphing [3], needs to be adopted. The accuracy of these methods has increased in the last decades. But they have not been widely used in real applications due to their excessive dependence on manual operations and need for prior knowledge of scene geometry. In this work, we extend and apply view morphing in a real application by integrating robust fundamental matrix estimation and feature matching. Our method only requires a slight adjustment of existing camera settings to make it amenable for practical use.

14.2 Approach

We assume that plural cameras have already been set around the given intersection. Obviously, the more cameras there are, the better reference view could be generated. In our work, evaluation was done by positioning six cameras at uniform heights. The detailed arrangement is shown in Fig. 14.2 (left). Our method does not restrict the detail position of cameras technically. Such a symmetrical setting is used only for an easy explanation. Each camera and its clockwise neighboring camera form a pair which are denoted as C_{n0} and C_{n1}. Here, n is the number of the pair. Like most actual situations, cameras are not calibrated in advance.

Our onboard system is supposed to receive image streams that are generated by each roadside camera while approaching the intersection. The camera pair which produces an orientation closest to the host vehicle is selected. The two images are then prewarped to make their image planes become parallel without changing the optical center of the cameras. Afterwards, we produce a novel view by linearly interpolating the positions and color of the two prewarped images. The resulting image is parallel with the prewarped two images, and it is shape-preserved. The position of the perspective view is determined by the angle between the vehicle's direction and directions of two selected cameras. Then, the images are again warped to align with the host vehicle's direction. In this way, we generate a view for the virtual camera as C_s shown in Fig. 14.2. After a zooming stage via driver interaction, the final output of the system is the approximate view following the host vehicle's motion.

View morphing [3] is the inspiration for our method here. It could generate image from any viewpoint by linking two original cameras together. Note that the original method requires prior knowledge of the camera's projection matrices and excessive reliance on manual operation. Our team broadened it by integrating robust fundamental matrix estimation and sparse key point matching. The following paragraph further describes this method.

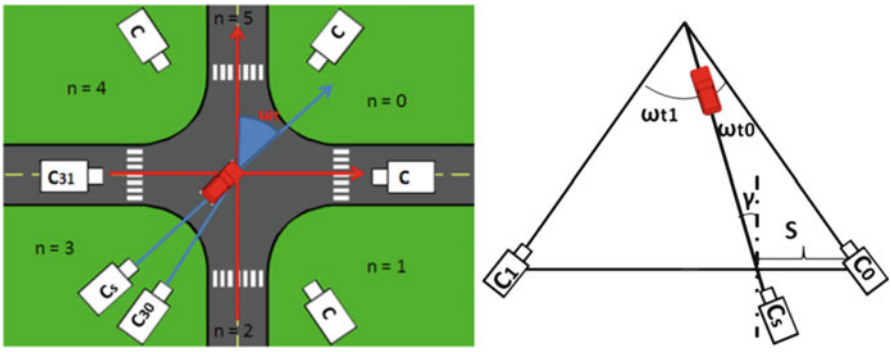

Fig. 14.2 Actual and virtual cameras

14.3 Actual and Virtual Cameras

As mentioned previously, the detailed position of roadside cameras is not restricted. We made it a precondition because the existing roadside cameras in many intersections are not set for our purposes. The number may not be enough, and the settings may not suit our needs. Adding one or two cameras or adjusting the existing settings will make it convenient to use. At the same time, a robust way is needed to combine the images in a direct way. Moreover, to be able to align it with the host vehicle's direction, the virtual camera's position and direction should also be determined via the online measurement of the host vehicle's motion.

14.3.1 Estimate Fundamental Matrix

For the camera pair n, its fundamental matrix F_n is invariable and only needs to be computed once. The first step we perform is to take two images I'_0 and I'_1 from two cameras C_{n0} and C_{n1} and establish correspondence between them. Since I'_0 and I'_1 are from disparate viewpoints, such feature matching across a wide baseline is an error-prone task.

To achieve good estimation of the F_n, we first use SIFT key point detector [4] to select a set of key points from each image. We choose SIFT key point because it is robust against image transformation, and with a descriptor associate with each key point, we can easily establish potential correspondence with high confidence. Then, we match the key points between the image pair by finding the nearest neighbor of their descriptors in Euclidean distance. Since it may still contain many outliers of the matching, we adapt RANSAC [5] to estimate the F_n. During each RANSAC loop, eight corresponding pairs are randomly selected, and associated fundamental matrix is estimated using eight-point algorithm [6]. The quality of the estimated fundamental matrix in each loop is assessed by counting the number of inliers. A match is treated as inlier when there is reprojection error under a threshold. After many iterations of RANSAC for each camera pair, we get consistent results of F_n.

14.3.2 Virtual View Point

Since our goal is to generate a view that dynamically follows the host vehicle, the virtual viewpoint's direction should be the same as that of the host vehicle. In this chapter, we assume that the online direction of the host vehicle is known as ω_t (Fig. 14.2 left), where t is the time index. Based on it, we take the camera pair C_{t0} and C_{t1}, which has the closest direction with ω_t, and compute the corresponding angles ω_{t0} and ω_{t1}. In order to produce the view of the virtual camera C_s, the s and the camera tilt γ (Fig. 14.2 right) are needed. The s determines the morphing rate when producing the intermediate

parallel view by interpolation, while the γ is needed when rotating the interpolated image to align the host vehicle's direction. We treat the position of $C_{t0} = 0$ and $C_{t1} = 1$; then, the s could be worked out approximately via $S = \omega_{t0}/(\omega_{t0} + \omega_{t1})$ while $\gamma = (\omega_{t1} - \omega_{t0})/2$.

14.4 Generating Reference View

In this section, we will introduce our method of generating the reference view. In each time step, one camera pair is selected, and the images I_0 and I_1 from C_0 and C_1 are used as source. Our method is an extension of view morphing [3]. We integrate a feature-matching procedure to avoid manual operations. Our method could be summarized as a four-step procedure as described in the following section.

14.4.1 Feature Correspondence

In order to produce a morph, the complete correspondence maps between each pixel of two source images should be specified. Previously, the user manually determines correspondences by specifying a sparse set of matching features. The remaining correspondences are then ascertained based on these matches by interpolation [7]. Excessive reliance on manual operation makes the process ambiguous and not easy to manipulate [8].

In our work, it is also necessary to obtain the correspondence maps to synthesize a shape-preserved novel view. To ensure the quality of the novel view, a sufficient number of matches and ample distribution of the images should be guaranteed. In this situation, again comes the issue of establishing correspondence between images across a wide baseline. Differences in estimation of the fundamental matrix arise. Here, the quality of potential matches is more important. At the same time, there is additional need for quantity and distribution concerning such matches.

We apply SIFT detector [4] and Harris Corner Detector [9] on both I_0 and I_1, and collect responses from the image pair. We again use the descriptor of SIFT key points to establish potential correspondence as we have done in Sect. 14.3.1. For each corner key point, we use the normalized cross-correlation criterion to find its best match [6]. The reason we use two detectors is that they have different properties. With a local descriptor, SIFT key point is efficient in establishing correspondence with high confidence. Beside SIFT, using Harris corner key point could ensure that sufficient shape-related correspondence can be found. We then collect the matches generated in this manner. In order to eliminate false matches, we further use the precomputed fundamental matrix to remove outliers by enforcing Epipolar constraint. This way, we obtained a set of sufficient correspondence with high confidence. These correspondences will then be used in the following view synthesis procedure.

14.4.2 Prewarping

In order to produce a shape-preserved morph, the two images should be rotated twice to align the image planes and scan lines. Then, the linear interpolation on the warped image could produce new perspective views as the camera moves along the line linking two cameras together. Therefore, in each time step, we need to perform projective transformations H_0 and H_1 on both I_0 and I_1.

We denote $R_{\theta_i}^{d_i}$ and R_{ϕ_i} $(i = 0, 1)$ are both 3 by 3 matrix. $R_{\theta_i}^{d_i}$ is a rotation of angle θ_i about axis d_i in depth, which makes the two image planes become parallel, while R_{ϕ_i} corresponds to an affine warping to align the scan lines. Given the fundamental matrix F of I_0 and I_1, the four matrixes could be determined by choosing a rotation axis d_0.

The first thing we do is to factorize the precomputed F with singular value decomposition and obtain two unit eigenvectors (epipoles) $\mathrm{e}_0 = \left[\mathrm{e}_0^{\mathrm{x}}, \mathrm{e}_0^{\mathrm{y}}, \mathrm{e}_0^{\mathrm{z}}\right]^{\mathrm{T}}$ and $e_1 = \left[e_1^x, e_1^y, e_1^z\right]^T$ of F and F^T respectively. We follow the recommended choice in [3] and select the rotation axis as$d_0 = [-e_0^y, e_0^x, 0]^T$. Then, we compute a vector $[x, y, z]^T = Fd_0$, and take $d_1 = [-y, x, 0]^T$. The angles of rotation in depth about d_i could be computed via

$$\theta_i = -\frac{\pi}{2} - tan^{-1}\left(\frac{d_i^y e_i^x - d_i^x e_i^y}{e_i^z}\right). \tag{14.1}$$

In this way, the two rotations in depth are determined.

Following the depth rotation is another affine warp R_{ϕ_i} to make the Epipolar lines parallel. After the first rotation, the new epipoles become $\left[\tilde{e}_i^x, \tilde{e}_i^y, 0\right]^T = R_{\theta_i}^{d_i} e_i$. Then, the angles of rotation ϕ_0 and ϕ_1 could be obtained via

$$\phi_1 = -tan^{-1}\left(\tilde{e}_i^y / \tilde{e}_i^x\right). \tag{14.2}$$

After each image has been rotated twice, the original fundamental matrix is formed:

$$\tilde{F} = R_{\phi_1} R_{\theta_1}^{d_1} F_n R_{-\theta_0}^{d_0} R_{-\theta_0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & b & c \end{bmatrix}. \tag{14.3}$$

To make sure F is in the form:

$$\left(H_1^{-1}\right)^T F H_0^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{14.4}$$

Another translation is then applied on I_1 as

$$T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -a & -c \\ 0 & 0 & b \end{bmatrix} \tag{14.5}$$

Now, two prewar transforms can be computed via $H_0 = R_{\phi_0} R_{\theta_0}^{d_0}$ and $H_1 = TR_{\phi_1} R_{\theta_1}^{d_1}$.

With the obtained H_0 and H_1, we perform the projective transformations on the two images I_0 and I_1, and obtain $\hat{I}_0$ and $\hat{I}_1$. In the previous step, we have obtained a set of feature matches. For the following interpolation step, we also perform the same projective transformation on their coordinates.

14.4.3 Image Interpolation

We have shown that in view morphing [3], the linear interpolation of parallel images is another parallel view. After the prewarping, both images $\hat{I}_0$ and $\hat{I}_1$ are capable for such kind of interpolation. In addition, during the transformation of a coordinate, the matching point's coordinates changed as well. The correspondence of the original images is preserved and represented as the new coordinate of the warped images. We then determine the maps of non-key points between two warped images using MATLAB 4 griddata method. This method could produce smooth surfaces for all pixels between $\hat{I}_0$ and $\hat{I}_1$ from a set of correspondence, namely two mapping function $T_0 : \hat{I}_0 \rightarrow \hat{I}_1$ and $T_1 : \hat{I}_1 \rightarrow \hat{I}_0$.

Using the morphing rate s, what we have estimated previously, and the two mapping functions in our hand, we then compute the displacement of each pixel $P_0 \in \hat{I}_0$ and $P_1 \in \hat{I}_1$via :

$$W_0(p_0, s) = (1 - s)p_0 + sT_0(p_0), \tag{14.6}$$

$$W_1(p_1, s) = (1 - s)T_1(p_1) + s(p_1). \tag{14.7}$$

Then, we integrate their colors by cross-dissolve procedure.

14.4.4 Postwarping and Zooming

After the interpolation, we have produced a novel view of the intersection. Since such view is parallel with the line linking two cameras together, we then perform a postwarp to make it align along the host vehicle's direction. The warping is a plane rotation of angle γ in depth. After postwarping, the driver may need to zoom to finally approximate or obtain the reference view he/she will use during decision making while driving through an intersection.

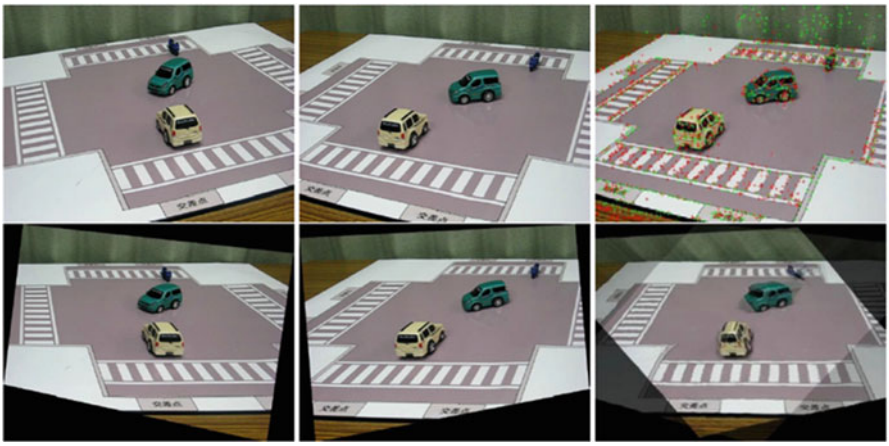

Fig. 14.3 Experimental results

14.5 Experimental Result

Our evaluation experiment is done by using an intersection model at 1:38 scale. We use six cameras with resolution 640 by 480. The camera setting is approximately the one shown in Fig. 14.2 (left). Remote toy cars and bikes are used to obtain test image sequences. Figure 14.3 (top left and middle) shows a pair of our sample input from the left and right cameras respectively.

First of all, the fundamental matrices are estimated in the way mentioned in Sect.14.3.1. The prewarping transformations H_0 and H_1 are then calculated based on F. We take a pair of cameras' images as examples as shown in Fig. 14.3 (top left and middle). By jointly using SIFT and Harris detectors, about two thousand key points were selected in each image, and the distribution is normalized as shown in Fig. 14.3 (top right, green: SIFT, red: Harris). Using the matching criterion in Sect. 14.4.1 and followed with a manually operated refining step, two hundred and ten features were finally selected as correspondence. We then make the projective transformations on the two images (Fig. 14.3 bottom left and middle) as well as the matching points' coordinates. Without automatic estimation of vehicle's direction, we then produce a reference image by manually assigned morphing rate s and the camera tilt angle γ. The resulting image is shown in Fig. 14.3 (bottom right). Even though the resulting image contains some ghost effect, it is evident that the proposed method works well.

14.6 Conclusion

In this chapter, we proposed a method to generate the reference view of traffic intersection for safe driving assistance. We adapted the view morphing approach and broadened it using robust fundamental matrix estimation and automatic feature

matching. This allows us to achieve the goal without any prior knowledge of scene geometry and excessive manual operation, which were crucial obstacles under the original model. Our experiment shows that our method works fine even for the images from large baseline disparate viewpoints.

During the processing, since the original images were resampled many times and occlusion exists, the resulted novel view contains some ghost effect. In order to solve these effects, we will optimize the raw output by introducing smoothness prior to the future work.

References

1. Benmimoun A, Chen J, Suzuki T (2007) Design and practical evaluation of an intersection assistant in real world tests. In: Proceedings of IEEE intelligent vehicles symposium Istanbul, pp 606–611
2. Ichihara E, Takao H, Ohta Y (1999) NaviVIew: bird's-eye view for drivers using roadside cameras. The transactions of the IEICE J82-D-II(10):1816–1825
3. Seitz SM, Dyer CR (1996) View morphing. In: ACM Proceedings of SIGGRAPH 96, New York, pp 21–30
4. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int'l J Computer Vision 60(2):91–110
5. Fischler MA, Bolles C (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm of the ACM 24(6):381–395
6. Ma Y, Soatto S, Kosecka J, Sastry S (2003) An invitation to 3-D vision: from images to geometric models. Springer, New York
7. Beier T, Neely S (1992) Feature-based image metamorphosis. In: ACM Proceedings of SIGGRAPH 92, Chicago, USA, pp 35–42
8. Shum H, Kang SB (2000) A review of image-based rendering techniques. IEEE/SPIE Visual Commun Image Process, pp 2–13
9. Harris C, Stephens MJ (1988) A combined corner and edge detector. In: Proceedings of Alvey vision conference, Manchester, England, pp 147–152