Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013

Panoramic Stitching for Driver Assistance and Applications to Motion Saliency-Based Risk Analysis

Alfredo Ramirez, Eshed Ohn-Bar and Mohan Trivedi, *Fellow, IEEE*

***Abstract*— Cameras can provide the capability for surround vehicle analysis and intuitive visualization for driver support in critical situations. As salient objects can be distributed anywhere around the vehicle, a wide view of the surround has many benefits for studying driver behavior in a holistic manner, as well as developing effective driver assistance systems. In this work, we are concerned with the hardware setup and calibration needed to provide such stitched views. In particular, views with large translation differences, many moving objects, and large changes in brightness are handled. We show the qualitative effects of calibration scene and camera orientation on common stitching algorithms. Finally, we analyze the stitched view for salient objects to detect critical events.**

I. Introduction

Surround awareness is critical for driving safety. According to the 2006 Traffic Safety Facts report conducted by the US National Highway Traffic Safety Administration [1] 50.7% of the accidents occur at the front of the vehicle, 11.8% on the sides, and 24.5% in the rear. For monitoring dangerous situations and improving the driving experience, different sensors have been incorporated, such as cameras monitoring lanes and assistance systems through camera, radar, or lidar. Examples of such camera systems include back-up cameras and 360-degree birds-eye view systems, which are very useful for assisting the driver in parking and maneuvering around tight spaces. However, they are not effective at large distances, and are thus impractical outside of driveway or parking lot scenarios. These systems can be implemented using fish-eye lenses[2], or with omni-directional cameras[3], [4], both of which provide a very large field of view at the cost of large distortion. To avoid such lens distortion, in this paper, we use multiple wide-angle rectilinear cameras and stitch them together using common stitching techniques to create a large field of view video of the rear of a vehicle. We qualitatively evaluate the wide-view images created with this camera setup and discuss what affects the performance of the stitching algorithms. 170-degree wide view images were able to be created with a large amount of visible ghosting and stretching near the edges. We found that the scene used to calibrate the system as well as the angular separation of the cameras greatly affect the final outcome of the stitched view.

Compared to looking in the mirrors, our visualization provides a consistent representation of the surround with a higher coverage of the scene. Furthermore, such a framework can be used to better analyze the scene for vision algorithms, as they can potentially provide longer trajectories of objects and additional scene geometry and lane cues. In particular, we show the usability using an overtaking vehicle detection system that visualizes risk to the driver. Furthermore, the proposed system could be combined with a heads-up display[5] in order to further reduce the driver's need to shift gaze off the road.

Figure 1: The view is produced from three wide-angle cameras. Images are shown for highway and urban settings. Moving objects, large displacement between the cameras, and brightness changes produce a difficult stitching problem.

The outline of the paper is as follows. First, we discuss the stitching techniques needed to produce the wide view, such as image registration and homography estimation. Next, we describe the calibration process needed for the camera system. The stitched views are then presented and evaluated. Finally, we discuss the camera system in the context of motion saliency analysis.

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego (UCSD), La Jolla, CA, 92092 USA (e-mail: alr034@eng.ucsd.edu; eohnbar@ucsd.edu; mtrivedi@ucsd.edu).

978-1-4799-2914-613/$31.00 ©2013 IEEE

II. Related Research Studies

Extensive work has been done in the field of image stitching. For image registration and alignment we follow selected portions of the approach used in [6], [7] using SIFT features to estimate a homography between the different cameras. Generally, homography estimation only works well for stitching images when there is only a rotation between the cameras, or when the scene in the images is far away from the cameras. Neither of these qualifications are true for our camera setup; there is a large translation between the cameras, and there are often pedestrians or other vehicles near by in driving scenarios. The authors of [8] follow a similar stitching process as in this paper. However, we would like to evaluate how well the off-the-shelf panorama techniques work under these conditions and how we can minimize any resulting distortions by finding the optimal camera setup and registration scene.

III. Stitching

The large-view panorama rendered from the vehicle-mounted cameras was formed using well-proven[6] image stitching techniques. During calibration, SIFT features are extracted from each view, and are matched to features in the other overlapping views using fast approximate nearest neighbor (FLANN). The matched features are used as inputs to random sampling consensus (RANSAC) in order to remove any mismatched points, and to estimate the homography between overlapping views. Since the cameras are in fixed positions, calibrating offline once will be sufficient for far away objects to appear undistorted. Online, the images are warped into the center cameras frame of reference using the homographies. Warping often results in errors in the overlapping regions, such as intensity differences and other artifacts due to parallax effects and misregistration. Feathering [9] is used on the final images to reduce the distortions in the overlap. The distance map, $D_i(x, y)$, for each view $i \in \{1, 2, 3\}$, is calculated and used to find the weights, $W_i(x, y)$ for each image.

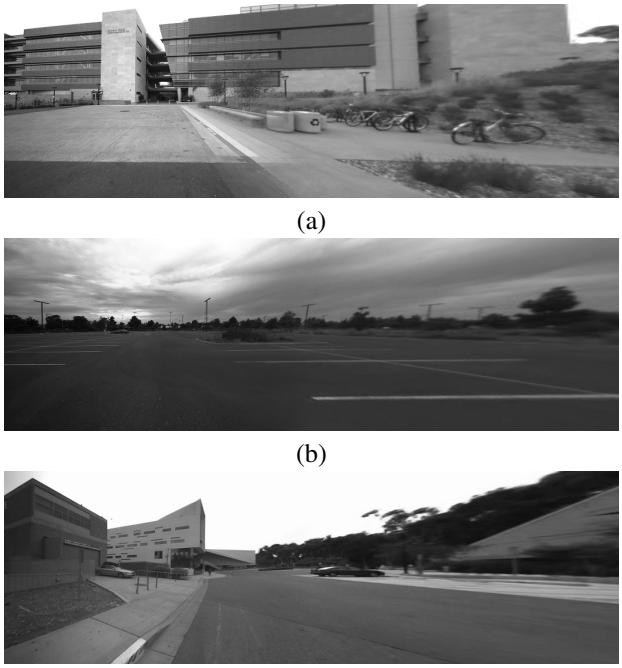
$$W_i(x, y) = \frac{D_i(x, y)}{\sum_i D_i(x, y)} \tag{1}$$

The final panorama is formed by adding all of the weighted images together.

$$I_{stitched}(x, y) = \sum_i I_i(x, y) W_i(x, y) \tag{2}$$

IV. Experimental Setup

The test-bed used for our experiments is comprised of a VW Passat mounted with 3 Point Grey Flea3 cameras fitted with Theia SY125M lenses. One camera is mounted in the center rear of the vehicle, and two cameras on either side pointing out. The outer cameras are placed with a lateral separation of 62 cm from the center camera, and a longitudinal separation of 30 cm from the center camera, with all cameras are set to 0-degree pitch.



(a)

(b)

(c)

Figure 2: The panorama results from calibrating the camera system using different scenes with the same separation and a rotation of 40 degrees. (a) The facade of planar building. (b) An empty parking lot. (c) An empty street with buildings on either side. Note the more severe stretching of the building (c) and the ghosting in the light poles and parking lanes in (b).

V. Results

Calibration of the camera setup entails capturing video of a scene and calculating the homography for the two outer cameras with respect to the center camera. Testing was done to determine the optimal settings for the calibration that balanced field of view of the panorama and visual distortion. First, three separate scenes were tested for calibration: an empty parking lot with many trees and structures on the horizon, an empty street with buildings and trees relatively close by, and the facade of a distinct, planar building from far away. These particular scenes were picked for testing due to being fairly flat and static. Video footage with the outer cameras rotating away from the center camera at 10 degree intervals was taken at these locations. To get an initial visual feedback using different camera configurations, a quick calibration was performed in real time; 250 features from the different views were extracted and matched, RANSAC was run for 1000 iterations, and a stitched image was rendered. The calibration scenes were evaluated qualitatively, based on the field of view that could be achieved with the stitched image during this experiment and the amount of distortion in the image. Next, a more through calibration was performed in order to determine the optimal angle of separation. Using footage captured while rotating the cameras by 10 degree intervals, 650 SIFT features are extracted and matched, and

978-1-4799-2914-613/$31.00 ©2013 IEEE

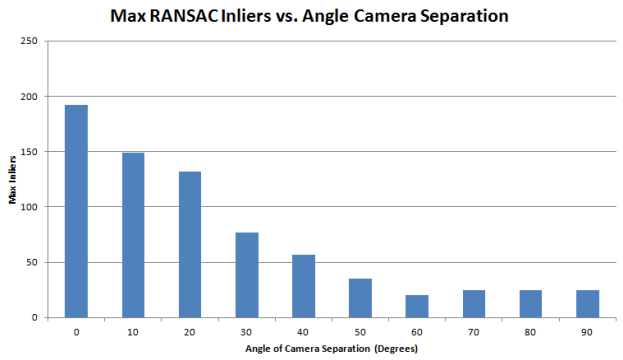
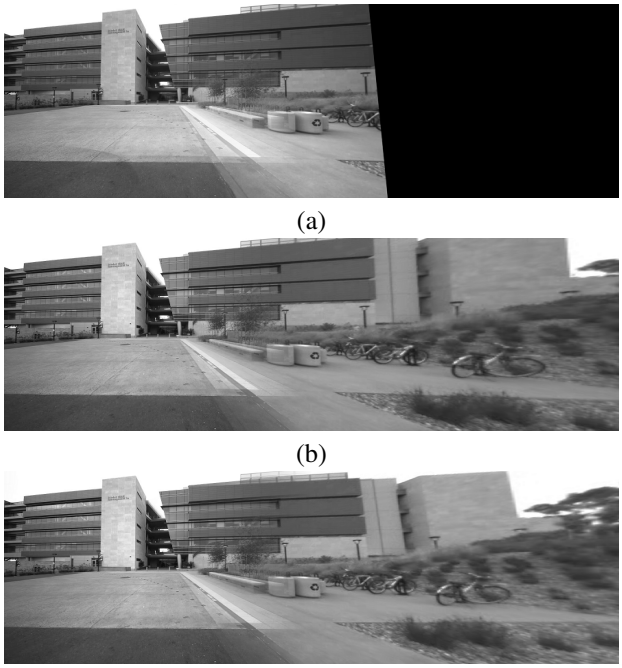


Fig. 3: The maximum number of RANSAC inliers produced during the homography estimation of the building scene for different camera separation angles. Note the large drop in max inliers at the 50 degree mark and after the 50 degree mark. These drops correspond to increased distortion in the final stitched panorama.

a homography for each angle is estimated using RANSAC running 2 million iterations.

After the experimentation, we found that the building scene was the best calibration setting for our purpose. With the building scene, a homography could be estimated with up to 50 degrees of camera separation, while the parking lot could not go past 40 degrees. As can be seen in Figure 2a, there are fewer ghosting artifacts and less stretching towards the edges in the building scene when compared to the other scenes. The face of the building has many distinct corners, which were easily identified and matched with SIFT and FLANN, and the planar structure facilitated the homography estimation. The parking lot scene had many mismatched points, which was very detrimental to the homography estimation; the large amount of sky and pavement is difficult to describe with SIFT, and the parking space markings and light posts were often mismatched because they all appear too similar. These problems limited the field of view for which a reliable homography could be estimated. The results (Figure 2b) show significant ghosting in the light posts and the painted lines, and the projection distortion is apparent in the stretched out and leaning posts. The street scene had problems similar to the parking lot (a large amount of sky and pavement in the scene), which hindered its performance. As can be seen in Figure 2c, there is a heavily stretched building on the right edge of the image, indicative of large projection distortion in the street scene.

We also found that, for our camera configuration, a 40-degree separation between the cameras is optimal. Figure 3 shows a plot of the maximum inlier count for each separation interval. As expected, the max inlier count decreases as the angle increases, due to the decrease in overlap between the views. As can be seen in the graph, there is a sudden drop in inliers in the 30 degree range, and another drop after 50 degrees. Figure 4 shows samples of stitched images at several angles. Figure 4a shows a panorama with 10 degrees of camera separation. This image has a very limited field



(a)

(b)

(c)

Fig. 4: Panorama results from different angles of camera separation. (a) 10 degrees. (b) 40 degrees. (c) 50 degrees. Note the increased field of view in (b) and (c) compared to (a), and the ghosting in the roof of the building in (c). 40 degrees was determined to be the optimal angle due to the large fov with less distortion than 50 degrees.

of view, and displays ghosting for some nearby objects (the trash bins and the fire lane), but far away objects appear undistorted. Figure 4b shows a sample at 40 degrees. Note that some of the nearby objects are still ghosted, but the building remains undistorted. Also note the greatly increased field of view compared to the 10 degree sample: most of the building is now in view, and another bicycle can be seen near the camera. At 50 degrees, shown in Figure 4c, the field of view increases even more (a large tree in now in view near the edge of the picture). However, there is now ghosting in the background objects as well, seen in the roof of the building near the center of the image. At angles larger than 50 degrees, no usable homography could be estimated (i.e. images warped with these homographies produced gibberish), which corresponds with the drop seen in the plot in Figure 4. We saw that below a certain percentage of max inliers to matches (5% for our setup), a homography could no longer accurately estimated. From this experimentation, we concluded that 40 degrees of camera seperation would be optimum for our setup because we can still produced a very wide angle view (170 degree panorama with three cameras) without the distortion at large distances that are introduced at 50 degrees.

A sample image with the optimum settings (40 degree separation, calibrated with the building scene) can be seen in Figure 5. The figure shows severe ghosting in the lanes

978-1-4799-2914-613/$31.00 ©2013 IEEE

Fig. 5: Three individual views and the resulting panorama for a residential scene.

and the nearby vehicles, as well as stretching near the edges of the image. However, despite these problems, much information can still be gathered that can be useful to a driver: there is a clear view of the lane adjacent to the ego-vehicle and the relative position of surrounding vehicles can be easily discerned.

VI. Motion Saliency for Critical Situation Analysis

With the rich stitched view at hand, we would like to analyze it for salient objects that may provide risk so that we may warn the driver. To that end, we analyze overtaking vehicle scenarios. As the stitched view contains many vehicles in different orientations and changing aspect ratio, motion can be used as a reliable cue to detect moving objects in the proximity of the vehicle. Furthermore, we would like a saliency measure that extends beyond the single frame. Therefore, in order to exemplify the usefulness of the stitched view, we turned to motion saliency from long term analysis from point trajectories in the scene[10]. Such an approach provides several advantages. First, the method handles ego-motion since trajectories are groups with relative motion to each other. Second, it is more reliable in tracking as our agents of interest may be stationary over some period of time, and then move relatively to the ego-vehicle. Using trajectories increases the chance of avoiding clustering such objects as background. This method is computationally intensive and not yet suitable for real-time, but recent efforts were made to push such methods to real-time [11].

As a case study, we detect overtaking vehicles in the scene

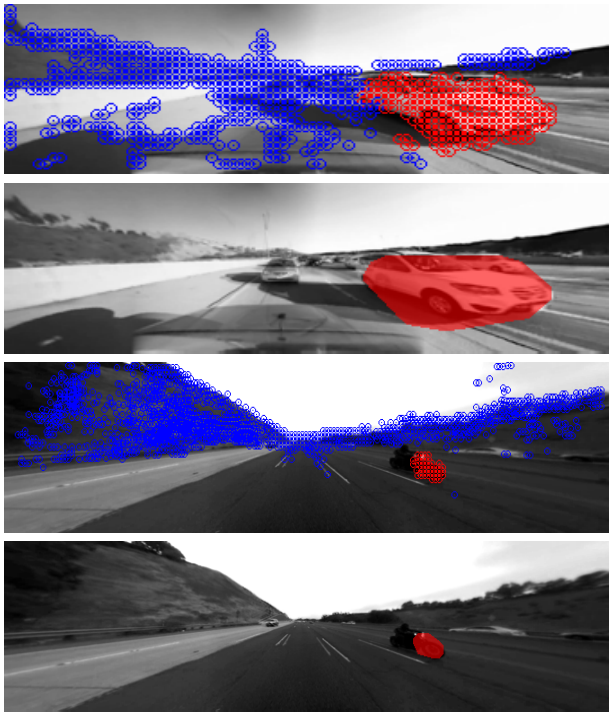


Fig. 6: Motion-saliency in the panoramic view. Long term point trajectories are clustered,and the convex hull is calculated to produce a blob, which is detected as an overtaking vehicle using a trained SVM classifier.

978-1-4799-2914-613/$31.00 ©2013 IEEE

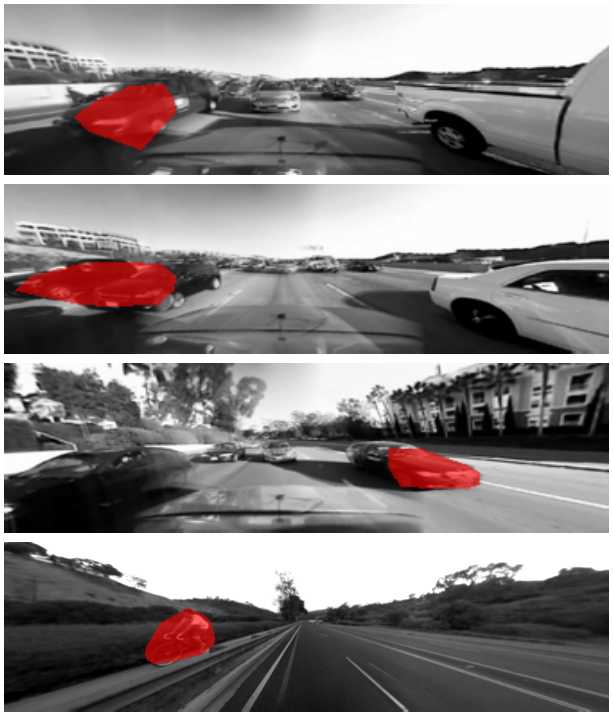


Fig. 7: Visualizing risk potential of vehicles in dense freeway and urban settings. In red is the convex hull of the dense trajectories in a cluster classified as an overtaking vehicle.

based on their flow portfolio. To that end, we train a linear SVM to detect whether a vehicle is overtaking or not. The flow vectors are first summed into flow attributes as described in [12] and are inputted to the SVM. A data set of 60 vehicle overtaking instances in urban and freeway under light dense traffic is was constructed. The detector learns to identify patterns of motion relative to the ego-vehicle that imply an overtake is occurring. Sample outputs of the algorithm can be seen in Figures 6 and 7.

VII. Concluding Remarks and Future Work

We showed that with non-ideal settings (eg. large camera displacement), well known techniques for image registration and stitching can be used to provide a wide angle view of the rear of a vehicle. However, such view is plagued by ghosting artifacts and edge distortion. We discussed the scene dependence of the calibration and the effect camera separation has on the final rendering. Under the wider field of view and significant parallax error, we showed the feasibility of saliency analysis. In the future, we would like reduce the effects of ghosting from the stitched views through SVD analysis, as in [13]. We would also like to leverage the stitched view to incorporate additional scene information for better risk assessment. For instance, we could analyze motion patterns of the surroundings, similar to in [14] and [15].

VIII. Acknowledgment

We acknowledge support of the UC Discovery Program and associated industry partners. We also thank our UCSD LISA colleagues who helped in a variety of important ways in our research studies.

References

[1] National Highway Traffic Safety Administration, "Traffic safety facts 2010," Washington, DC, Tech. Rep. DOT HS 811 659, 2010.
[2] Altera Corp., "Generating panoramic views by stitching multiple fisheye images," San Jose, CA, Tech. Rep. WP-01107-1.0, 2009.
[3] K. S. Huang, M. M. Trivedi, and T. Gandhi, "Drivers view and vehicle surround estimation using omnidirectional video stream," in *IEEE Intell. Veh. Symp.*, 2003.
[4] T. Gandhi and M. M. Trivedi, "Vehicle surround capture: Survey of techniques and a novel omni-video-based approach for dynamic panoramic surround maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 293–308, Sep. 2006.
[5] K. Huang, S. Cheng, and M. Trivedi, "A novel active heads-up display for driver assistance," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2009.
[6] M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," *Interional Journal of Computer Vision*, vol. 74, pp. 59–73, Aug. 2007.
[7] R. Szeliski, "Image alignment and stitching: a tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, 2006.
[8] J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, . Gonzlez, and J. Almazn, "Surrounding view for enhancing safety on vehicles," in *IEEE Intell. Veh. Symp.*, 2012.
[9] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2011.
[10] P. Ochs and T. Brox, "Higher order motion models and spectral clustering," in *Computer Vision and Pattern Recognition*, 2012.
[11] T. Senst, R. H. Evangelio, I. Keller, and T. Sikora, "Clustering motion for real-time optical flow based tracking," in *Advanced Video and Signal-Based Surveillance*, 2012.
[12] E. Ohn-Bar, S. Sivaraman, and M. M. Trivedi, "Partially occluded vehicle recognition and tracking in 3D," in *IEEE Intell. Veh. Symp.*, 2013.
[13] A. Srikantha, D. Sidibe, and F. Meriaudeau, "An SVD-based approach for ghost detection and removal in high dynamic range images," in *International Conference on Pattern Recognition*, 2012.
[14] A. Geiger and B. Kitt, "ObjectFlow: A descriptor for classifying traffic motion," in *IEEE Intell. Veh. Symp.*, June 2010.
[15] B. Morris and M. Trivedi, "Unsupervised learning of motion patterns of rear surrounding vehicles," in *IEEE International Conference on Vehicular Electronics and Safety*, 2009.