VEHICLE MOUNTED WIDE FOV STEREO FOR TRAFFIC AND PEDESTRIAN DETECTION

Tarak Gandhi and Mohan M. Trivedi

Computer Vision and Robotics Research Laboratory
University of California San Diego
La Jolla, CA 92093, USA
tgandhi@ucsd.edu,mtrivedi@ucsd.edu

ABSTRACT

This paper describes an approach for detecting objects in front of an automobile using wide field of view stereo with a pair of omni cameras. Several configurations are suggested for effective detection of vehicles and pedestrians. The omni cameras are calibrated using sets of parallel lines on a parking lot. The calibration is used to rectify the omni images. Stereo matching is performed on the rectified images to detect other vehicles and pedestrians. Experimental results show promise of detecting these objects on the road.

1. INTRODUCTION

In recent years, considerable research is being performed for developing intelligent vehicles having driver support systems to enhance safety. Computer vision techniques have been applied for detecting lanes, other vehicles and pedestrians to warn the driver of dangers such as lane departure and possible collision with other objects. However, an effective driver assistance system should be aware of the events taking place all around the car, and combine them with the inputs from the car dynamic sensors, as well as the events in the interior of the car. This paper describes a framework for using omni cameras in various configurations to perform detection of other vehicles and pedestrians using stereo analysis.

1.1. Related Research

Omni cameras with their panoramic field of view show a great potential in intelligent vehicle applications. In [1], an omni camera mounted inside the car was used to estimate driver's pose and generate the driver's view of surroundings using the same camera. In [2], feature-based methods detecting specific characteristics of vehicles, such as wheels were used to detect and track vehicles. In [3] the vehicle ego-motion was estimated and compensated, and features having residual motion were grouped to detect vehicles and generate a complete surround view showing the position and tracks of the vehicles.

It was observed that motion analysis gave effective results on the sides of the car, where independent motion of other vehicles give large motion disparity. However, for front objects, stereo based methods can be more effective since the stereo disparity between laterally placed cameras is usually greater than the motion disparity due to the longitudinal motion of a single camera.

Stereo cameras have been used by Bertozzi and Broggi [4] for lane and obstacle detection. They model the road as a planar surface and use inverse perspective transform to register the road plane between two images. The obstacles above the road would have residual disparity and are easily detected. For the case of curved roads, [5] create a V-disparity image based on clustering similar disparities on each image row. This approach was also used for omni cameras in [6] to monitor driver's blind spots around the vehicle.

2. FRAMEWORK FOR OMNI-STEREO ANALYSIS

The block diagram of the video analysis framework is shown in Figure 1. Camera calibration is performed off-line to determine the relationship between the world coordinates and pixel coordinates. Using the calibration information, the images are transformed to obtain virtual perspective views looking forward towards the road. This transformation, called rectification simplifies the stereo geometry making it easier to match corresponding features between the two images. Area-based correlation is then used to perform stereo matching between features. The result is a disparity map showing the displacement of features from one image to another. Based on the disparity map, the features are grouped into objects, and distance to the objects is computed.

This research was supported by a UC Discovery Program Digital Media Grant in association with industry partners.

0-7803-9134-9/05/$20.00 ©2005 IEEE

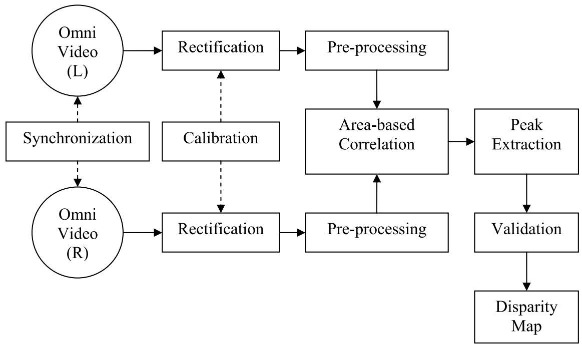

Fig. 1. Block diagram of stereo analysis from vehicle mounted omni cameras.

2.1. Camera Configurations

Due to the comparatively lower resolution of omni cameras, proper configuration is very important for obtaining good coverage, sensitivity, and foreground-background discrimination. Figure 2 shows various configurations that could be used for performing stereo using omni cameras. For example, cameras mounted on top of roof [3] can see vehicles at farther distance with better resolution. However, the disparity difference between vehicle and road is small, making it difficult to isolate the vehicles purely by stereo. Furthermore, such a configuration is not suitable for standard cars. Two omni cameras mounted near the side view mirrors can give stereo view of front, monocular views of sides, as well as views of the driver and passenger. Also, the disparity difference between vehicle and ground is larger than for a top-mounted camera making stereo discrimination easier. As a trade-off, the objects such as cars have smaller frontal area with window-mounted camera, reducing their image size and making it more difficult to detect vehicles that are farther away. Figure 2 (c) has cameras mounted in front of the car with narrower baseline. This configuration is more suitable for detecting nearby objects such as pedestrians. Similar configuration in the back of the car would be useful for monitoring blind spots behind the car effectively. [6] also uses cameras behind the car, but they are mounted coaxially so that the epipolar lines are along the radial lines instead of concentric circles.

2.2. Calibration and Rectification

In order to match the points between multiple cameras, and map them to the 3D space, it is necessary to calibrate the intrinsic and extrinsic parameters of the cameras. The intrinsic parameters relating the pixel and camera coordinates can be pre-computed before the cameras are installed using a setup in [6, 3]. The intrinsic parameters are used to convert the pixel coordinates to projective (2-D homogenous) coordinates representing the projection of the 3-D point on

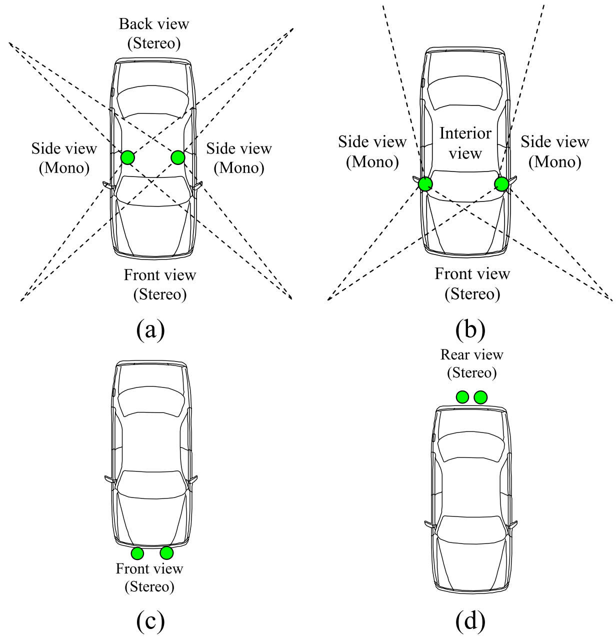

Fig. 2. Omni-camera configurations for stereo analysis with cameras mounted on (a) Top (b) Sides (c) Front (d) Rear of the automobile.

a virtual surface. Further processing can then be done in the projective coordinates.

The extrinsic parameters relate the world coordinate to each of the camera coordinates and need to be calibrated when the cameras are mounted on the vehicle. In particular, the rotation matrix of the camera is used to correct the pan, tilt, and roll angles in order to obtain rectified perspective views of the overlapping FOVs of the omni cameras. Calibration is performed by taking a car into a parking lot scene with a number of parallel lines in different directions. Note that the lines are mapped to curves in the omnidirectional image. A line passing through a number of points can be found by first converting the points to projective coordinates and then using singular value decomposition. If a number of parallel lines are obtained in this manner along the length of the car, the vanishing point of these lines can again be obtained using SVD. The columns of the rotation matrix converting between the world and camera projective coordinates are obtained from the vanishing points.

If the camera baseline is horizontal and perpendicular to the car's major axis, rectification can be performed by compensating the rotations of the cameras. Even if there is a slight difference between the camera heights and positions, the disparity caused by that parallax would be small for objects far from the cameras.

2.3. Stereo Analysis

In this work, area based correlation is used on rectified images for matching in order to obtain dense depth map. The implementation in [7] was used, which consists of the following steps:

1. Pre-processing: For effective stereo matching, band-pass filter is applied to the rectified images by convolving with a Laplacian of Gaussian filter.
2. Area correlation: Rectangular patches from one image are compared to neighboring patches from another image, displaced in horizontal direction using Sum of Absolute Differences (SAD). Coarse and fine level processing is used in order to fill the areas with low texture, and allow for a larger range of disparities.
3. Peak extraction: For each patch in first image, the corresponding patch in the second image with minimum value of SAD is extracted. The displacement between the patches is known as disparity of that pixel. A disparity image is formed by assigning a value proportional to the disparity to each pixel in the left image.
4. Post-filtering: This step cleans up the noise in the disparity image. An interest operator is used to reject uniform areas where the SAD does not have a sharp extremum and disparity computations therefore unreliable. Left-right check is used to eliminate errors due to depth discontinuities. Pixels with invalid disparity are given null value in the disparity image.

If the road is planar, the disparity in rectified images linearly increases with the row coordinates. Using the information from calibration module, the ground plane disparity can be compensated and a difference can be used to obtain regions corresponding to the objects. For non-planar roads, V-Disparity analysis proposed in [5, 6] would be useful. Further processing can be performed on these regions to identify vehicles and pedestrians.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The test bed car that is used for the project [8] focuses on creating a system capable of collecting large amounts of data from a variety of modular sensing systems and processing that data in order to be fed back to the human occupant. Sensor systems include rectilinear cameras, wide field-of-view camera systems, GPS and navigation systems, internal automobile vehicle state sensors, as well as other sensor systems useful for study in intelligent vehicles. The system contains an array of computers that serve for data collection as well as real-time processing of information.

A pair of omni cameras were mounted on the windows and the car was driven on the roads around the campus. Figure 3 (a)-(b) show the images captured from the cameras. Rectification was performed on these images as shown in Figure 3 (c)-(d). Figure 3 (e) shows the combined image from the two rectified images. It was noted that due to calibration errors, there was a vertical displacement of 3 pixels between the rectified image, which would cause error in stereo matching. Stereo disparities were obtained on the rectified images using the SRI SVS software. Figure 3 (f) shows the depth map. However, since the matching is performed along the horizontal scan lines, far-away objects such as the trees in background falsely get large disparities. This makes it necessary to have accurate calibration. In fact after manually compensating the vertical displacement as shown in Figure 3 (g), the resulting depth map shown in Figure 3 (h) is better. It is observed that the disparity calculation is much better, and the trees in background are correctly identified having nearly zero disparity. Figure 4 shows the result of stereo analysis on another image from the sequence. All the processing was performed on a Pentium IV yielding a frame rate of approximately 15 frames per second.

The experiment was repeated with the omni cameras mounted in front of the car in order to capture pedestrians. Figure 5 shows the result on an image sequence with pedestrians in front of the car.

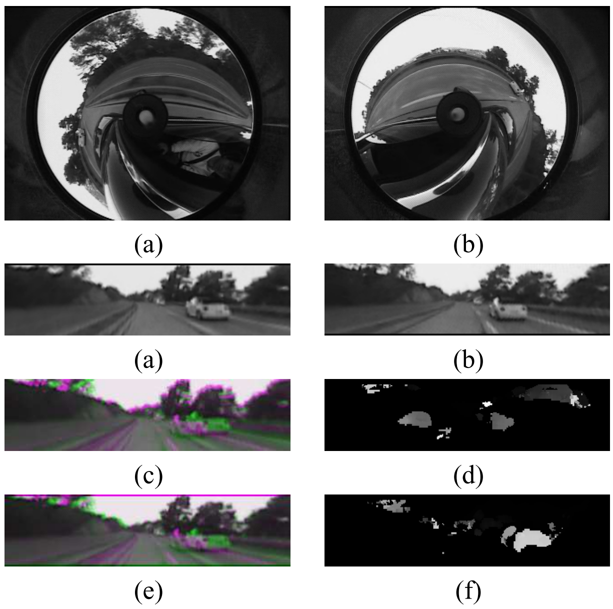

Fig. 3. Rectified perspective virtual views generated from the two omni cameras. The epipolar lines are along the rows of both the cameras. (c) Combined image. Note that due to calibration errors, there is some vertical displacement between features. (d) Depth map obtained using SRI SVS software. (e) Combined image after correcting the vertical disparity (f) Depth map obtained after correction.

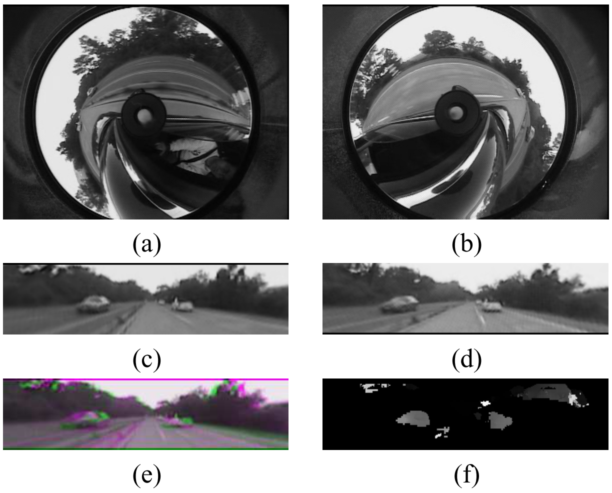

Fig. 4. Results of stereo disparity analysis on another image in the video sequence.

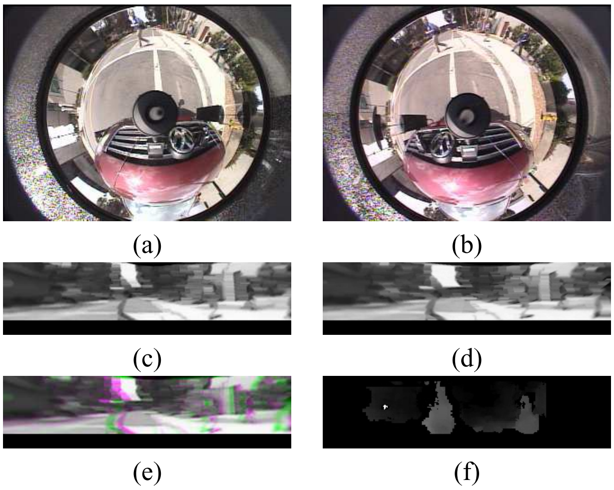

Fig. 5. Results of stereo disparity analysis for pedestrian detection.

4. SUMMARY

This paper described a framework for performing stereo analysis using a pair of vehicle mounted omnidirectional cameras in various configurations. Experimental results of detecting other vehicles and pedestrians were shown. It was observed that resolution of omni cameras is low for the front parts of the image. Due to this, segmentation and depth estimation becomes more challenging. However, use of only two sensors for analyzing front as well as side views make the approach attractive for designers. It was also noted that the stereo matching was somewhat sensitive to the accuracy and stability of the calibration. In order to compensate small changes in calibration due to camera vibration and drift, we plan to develop an incremental calibration by using the image motion of the car's own frame visible from the camera.

5. ACKNOWLEDGEMENT

We thank our colleagues from the CVRR Laboratory for their contributions and support.

6. REFERENCES

[1] Kohsia Huang, Mohan M. Trivedi, and Tarak Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in *IEEE Intelligent Vehicles Symposium*, June 2003, pp. 444–449.

[2] Ofer Achler and Mohan M. Trivedi, "Vehicle wheel detector using 2D filter banks," in *Proc. IEEE Intelligent Vehicles Symposium*, June 2004, pp. 25–30.

[3] Tarak Gandhi and Mohan M. Trivedi, "Parametric egomotion estimation for vehicle surround analysis using an omnidirectional camera," *Machine Vision and Applications*, vol. 16, no. 2, pp. 85–95, February 2005.

[4] Massimo Bertozzi and Alberto Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions On Image Processing,*, vol. 7, no. 1, pp. 62–81, January 1998.

[5] Raphael Labayrade, Didier Aubert, and Jean-Philippe Tarel, "Real time obstacle detection in stereovision on non flat road geometry through V-disparity representation," in *IEEE Intelligent Vehicles Symposium*, 2002, vol. II, pp. 646–651.

[6] Leanne Matuszyk, Alexander Zelinsky, Lars Nilsson, and Magnus Rilbe, "Stereo panoramic vision for monitoring vehicle blind-spots," in *Proc. IEEE Intelligent Vehicles Symposium*, June 2004, pp. 31–36.

[7] Kurt Konolige, "Small vision system: Hardware and implementation," in *Eighth International Symposium on Robotics Research*, October 1997, pp. 111–116, `http://www.ai.sri.com/~konolige/papers`.

[8] Joel McCall, Ofer Achler, and Mohan M. Trivedi, "Design of an instrumented vehicle testbed for developing human centered driver support system," in *IEEE Intelligent Vehicles Symposium*, June 2004, pp. 483–488.