

Parking Space Vacancy Monitoring

Catherine Wah
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093
cwah@cs.ucsd.edu

Abstract

Current methods of monitoring vacancies in parking lots have achieved varying levels of success, but most do not adequately deal with instances with severe vehicular occlusion. We propose a stereo-vision based system that seeks to address this issue. In this project, multiple cameras are used to monitor the vacancy status of the P502 parking spaces on UCSD campus.

1. Introduction

The availability of parking spaces on UCSD campus is a significant concern, and looking through crowded lots for available spots is both frustrating and time-consuming. It would save time for the driver to be notified when spots are available, rather than have to search for them himself or herself.

Current monitoring systems surveyed in [9] include sensor networks such as inductive loop detectors, magnetic sensors, or weigh-in-motion sensors which are embedded in the pavement. However, the cost of individual sensors becomes a limiting factor for large parking spots with many spaces to monitor.

In comparison to sensor-based parking space detection, which requires the installation and maintenance of networks of sensors, vision-based systems are more cost-effective as well as non-intrusive. However, current vision-based methods of vacancy detection do not adequately deal with occlusion, particularly vehicle-on-vehicle, and they assume the camera to be positioned such that it has a view of non-overlapping cars [10]. This project seeks to address that issue by applying stereo vision. By using stereo pairs of images, we can recover metric information from the fixed length of the baseline.

Stereo vision algorithms have been used in parking applications, including automatic parking. Current parking assistant systems [6] use a stereo camera setup to locate spaces and determine their pose relative to the vehicle.



Figure 1: An example of severe vehicular occlusion in the scale model.

Fabián [2] presents a method that takes occlusion into account in evaluating the vacancy status of a parking space, extracting regions of interest in a 3D model of the lot. However, his algorithm does not take into account instances of severe vehicular occlusion (Figure 1).

In this paper, we present a method for monitoring vacancies in parking lots using a stereo camera system to create a 3D reconstruction of the scene, which enables us to determine the vacancy status of a particular parking space under vehicular occlusion. Additionally, we compare results for 3D reconstruction using uncalibrated versus calibrated cameras.

2. Approach

The vacancy statuses of parking spaces for the P502 parking lot on UCSD campus (Figure 2) will be monitored using photos taken from the roof of CalIT2 with multiple cameras. To this end, this system must be able to identify vacancies while differentiating between spaces for different permit holders (faculty versus students).

Ideally, we would be able to provide an exact count of the number of available spaces, but most likely we will have to

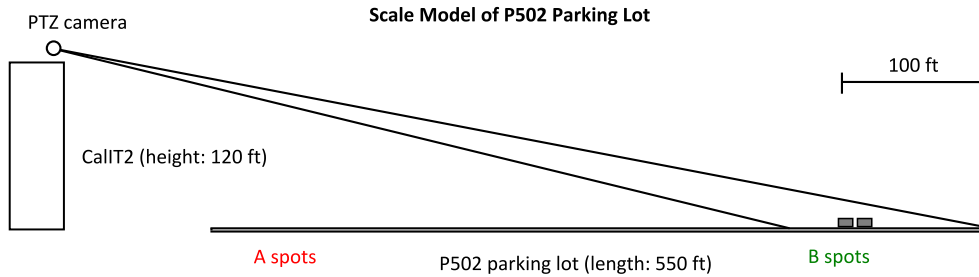


Figure 3: We recreated a 1:64 scale model of the P502 parking lot with approximate dimensions.



Figure 2: A view of the P502 parking lot from the roof of CallIT2.

appeal to a statistical notion of vacancy, as certain spots may be too heavily occluded by trees, other vehicles, etc. to be monitored with very high accuracy. This information will ultimately be integrated with a status dissemination tool, where drivers will be able to query the parking lot status via mobile phone.

Our proposed system involves 5 steps: (1) image capture, (2) feature extraction and point matching, (3) 3D reconstruction, (4) vacancy identification, and (5) driver notification. In the following sections we discuss each step in depth.

2.1. Image capture

Roof-mounted pairs of pan-tilt-zoom (PTZ) cameras will monitor the P502 parking lot, periodically performing raster scans of the lot at a specified zoom level for sufficient image resolution. Necessary considerations for scanning include image overlap, scan frequency, and scan time.

To facilitate testing in the lab environment, we simulated parking lot situations with a 1 : 64 scale model of the P502 parking lot (Figure 3) using toy cars of various models.

2.2. Feature extraction and point matching

For extracting features from the parking lot images, we use Harris and Stephens' [4] corner detection method to automatically find corners in pairs of images. Harris corner detection finds interest points based on changes in gradient direction, calculated from the sum of square differences.

In stereo reconstruction, one wishes to be able to perform robust feature matching, or automatically find correspondences between interest points in a pair of images, given spurious features and noise in the pointsets. The extracted corners are then used for RANSAC-based matching. Random Sample Consensus (RANSAC) [3] is an iterative method for robustly fitting a model in the presence of outliers, leaving us with the inlying matches.

First, one finds putative matches of interest points by searching for points of maximal correlation within windows surrounding each point. RANSAC is then used to fit a model with the largest number of inliers, discarding spurious correspondences [5].

2.3. 3D reconstruction

The next step after feature extraction is to perform reconstruction of the scene and camera structure by applying epipolar geometry [8], which describes how one can geometrically relate 3D points to their projections on the 2D images in stereo vision.

To briefly review, in the case of two cameras, the projection of each camera's focal point onto the other camera's image plane is the epipole, and the line segment connecting the two focal points is the baseline. An epipolar line is the intersection of a plane containing the baseline with the image plane. The fundamental matrix F captures this geometric information algebraically, describing the relations between corresponding points in stereo images for uncalibrated cameras. The counterpart for calibrated cameras is the essential matrix E .

In performing reconstruction, one can choose whether to first calibrate the cameras or not. Uncalibrated versus calibrated reconstruction each have their advantages and disad-

vantages, which we discuss in the following two sections.

2.3.1 Uncalibrated reconstruction

Uncalibrated reconstruction is powerful in that the structure of a scene can be computed from point correspondences alone, which makes it attractive for our purposes as one does not need to supply the system with additional information about the camera or scene. This lack of information can however reduce the accuracy of the reconstruction. The steps for uncalibrated reconstruction are as follows [8]:

1. Compute the fundamental matrix using RANSAC
2. Compute the camera projection matrices
3. Perform triangulation to recover the projective structure
4. Direct upgrade from projective to Euclidean structure

Reconstruction from two uncalibrated views yields only the projective structure, and in order to directly recover the Euclidean structure, we can use five ground-truth correspondences in general position (no four points are coplanar) to solve for the homography matrix that can take the 2D points to 3D. These five points are in units separate from the 2D images and the interest points, which capture pixel location.

2.3.2 Calibrated reconstruction

In calibrated reconstruction, we first perform camera calibration to obtain the extrinsic and intrinsic parameters of the camera. The extrinsic parameters are rotation and translation, and the intrinsic parameters include focal length of the camera, skew, and the principal point. As calibration provides us with this additional information to apply to the 3D reconstruction, the results are potentially more accurate than that of uncalibrated reconstruction. Nevertheless, small errors in calibration propagate and can still cause significant distortions in the 3D reconstruction [6].

The essential matrix

$$E = \hat{T}R \in \mathcal{R}^{3 \times 3} \quad (1)$$

is a compact representation of camera pose $g = (R, T)$. The relation between sets of points in two images is captured by the essential matrix and represented by the epipolar constraint

$$x_2^T E x_1 = 0, \quad (2)$$

which describes how a line in camera 2's image plane corresponds to a point in camera 1's image plane, and vice versa.

A disadvantage of calibration is the necessity of a calibration rig, an object in the scene with known geometry.



Figure 4: Harris corner detection results on a stereo pair of images.

For our purposes, since the dimensions of the parking lot are known and measurable, as well as the position of man-made landmarks in the scene such as lampposts, calibration is feasible without the necessity of placing an external object in the scene.

2.4. Vacancy identification

From the reconstruction of the scene structure, we can recover the 3D world coordinates of the interest points. If we assume each parking space to be represented by a rectangular prism, then we can determine the vacancy status of a particular spot by counting the number of interest points found in a certain prism or parking space. For example, the larger the number of features found within a given 3D parking space, then the more likely that that space is not vacant.

To determine yes/no vacancy, we can set a threshold for the maximum number of features that can be found in a space to be still be considered vacant. However, as there will most likely be spuriously reconstructed points, it is probably more appropriate to generate a probability of a given spot being vacant, based on the number of found interest points in the 3D region.

2.5. Driver notification

The final step in our approach is notifying drivers via mobile phone of the vacancy status of the parking lot. This will be done by SMS and voice-activated dialing using VoiceXML. The vacancy status of the parking lot will be updated periodically, depending on the frequency of the raster scans of the lot.

3. Data and results

Testing data was acquired using a pair of Sony SNC-RZ30N PTZ cameras that were mounted on tripods at a suitable height as designated by the chosen scale. The model parking lot itself was drawn on paper and placed on a table. As we are only concerned with daytime lighting conditions, we used the default light settings and collected pairs of images at 480x640 resolution.



(a)



(b)

Figure 5: (a) Putative matches; (b) Inlying matches consistent with the fundamental matrix, with the left image of a stereo pair overlaid on the right image.

In Figure 4, we have a stereo pair of sample images with the results of running Harris corner detection, using Kovese’s MATLAB toolbox [7].

From these interest points we find correspondences, for which we experimented with both wide and small baselines. This distinction refers to images in which the cameras that take each image are separated by a large translation versus a small one. We discovered that small baselines of 1.5 cm in the scale model (which translates to roughly 1 meter to scale) were insufficient for 3D reconstruction, and wider baselines, for instance around 8 cm, were more appropriate and produced sufficient observable change between the images.

We were unable to obtain an accurate uncalibrated reconstruction with the generated correspondences, which suggests that the Harris Corner Detector yielded too many spurious features that corrupt the calculation of the fundamental matrix. This can be seen in the inlying matches found using RANSAC (Figure 5), which contain features that obviously are incorrectly matched. For this reason, we chose to calibrate the pair of cameras and reconstruct using the

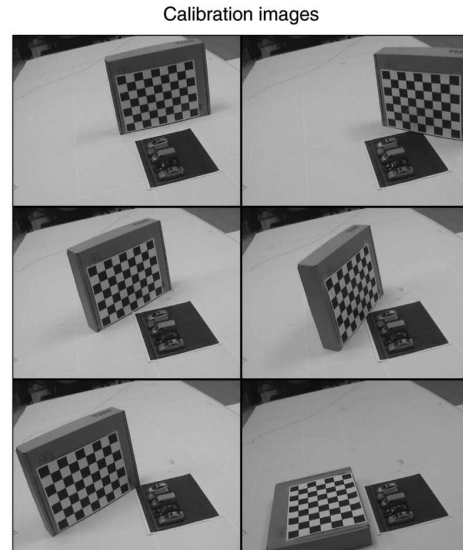


Figure 6: Various positions of the calibration object for stereo calibration (right camera only).

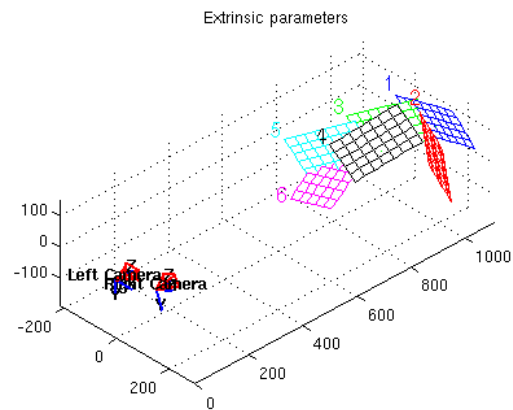


Figure 7: Extrinsic parameters of the stereo rig. Axes are in millimeters.

camera parameters.

We use Bouguet’s MATLAB toolbox [1] to perform stereo calibration and compute the extrinsic and intrinsic camera parameters. A checkerboard pattern is used as a calibration object, and we take pairs of images of the calibration object in six different positions (Figure 6). The spatial configuration of the cameras and checkerboard positions are shown in Figure 7. Before attempting calibrated reconstruction from automatic correspondences, we tested the accuracy of reconstruction with manually labeled correspondences, applying stereo triangulation to compute the 3D location of interest points given left and right image pro-

jections. The 3D reconstructed point cloud accurately represented variations in depth in the scene.

Since we know that calibrated reconstruction is accurate for manual correspondences, we investigated methods of automatically finding matches. One way is to perform image rectification, by projecting the images onto a common image and making the epipolar lines match the horizontal scan lines. For the rectified calibration images, we can create a disparity map that displays the disparity value at each point in one image. While this can be effective, it is sufficient to solve the problem mathematically by taking advantage of the epipolar constraint. We plan to continue this line of research in our future work.

4. Conclusion

This paper presents a technique for identifying vacant spaces in parking lots under severe vehicular occlusion. Current systems for parking space vacancy monitoring do not support adequate means of monitoring vacancies of this nature. Future work can involve applying more advanced feature extraction algorithms, for example SIFT, to generate better features for matching.

Also, we plan to compute the calibrated reconstruction for automatic correspondences by making use of the epipolar constraint. Lastly, it is worth investigating how camera calibration can be applied to images taken with varying focal lengths, so instead of recalibrating for different zoom levels, we can calibrate once and transform the camera parameters as necessary.

5. Acknowledgments

Thanks to Serge and the Winter 2009 class of CSE 190a for their valuable feedback and advice during the course of the quarter, and to CalIT2 for their funding of this project.

References

- [1] J.-Y. Bouguet. Camera calibration toolbox for matlab, 2001. Available from: <<http://www.vision.caltech.edu/bouguetj/calibdoc/index.html>>.
- [2] T. Fabián. An algorithm for parking lot occupation detection. In *7th Computer Information Systems and Industrial Management Applications*, pages 165–170, June 2008.
- [3] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
- [4] C. Harris and M. Stephens. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, pages 147–151, Manchester, UK, 1988.
- [5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2nd edition, 2003.
- [6] N. Kaempchen, U. Franke, and R. Ott. Stereo vision based pose estimation of parking lots using 3d vehicle models. In *Proceedings of IEEE Intelligent Vehicle Symposium*, pages 459–464, 2006.
- [7] P. D. Kovesi. MATLAB and Octave functions for computer vision and image processing. School of Computer Science & Software Engineering, The University of Western Australia. Available from: <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>.
- [8] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Springer Verlag, 2003.
- [9] L. Mimbela and L. Klein. A summary of vehicle detection and surveillance technologies used in intelligent transport systems. Technical report, The Vehicle Detector Clearinghouse, Las Cruces, NM, August 2007. Available from: <<http://www.nmsu.edu/~traffic/Publications/Trafficmonitor/vdst.pdf>>.
- [10] N. True. Vacant parking space detection in static images. University of California, San Diego, 2007.