Scale-Space Features in 1D Omnidirectional Images

Amy J. Briggs[1], Carrick Detweiler[1], Peter C. Mullen[2], and Daniel Scharstein[1]

[1] Department of Computer Science, Middlebury College, VT, USA
briggs, cdetweil, schar@middlebury.edu
[2] Woodbine Institute, Seattle, WA, USA

Abstract. We define a family of interest operators for extracting features from one-dimensional omnidirectional images, and explore the utility of such features for navigation and localization of a mobile robot equipped with an omnidirectional camera. A 1D circular image, formed by averaging the scanlines of a cylindrical panorama, provides a compact representation of the robot's surroundings. Feature detection proceeds by applying local interest operators in the scale space of the image. The work is inspired by the recent success of similar operators developed for 2D images. The advantages of using features in omnidirectional 1D images are fast processing times and low storage requirements, which allows a dense sampling of views. We present experimental results on real images that demonstrate that our features are insensitive to noise, illumination variations, and changes in camera orientation. We also demonstrate that most features remain stable over changes in viewpoint and in the presence of some occlusion, thus allowing reliable tracking of features through sequences of frames.

1 Introduction

Scale-invariant interest operators and feature detectors have received much recent interest in the computer vision and robotics literature [3, 9, 12, 13, 19]. These methods work by computing the *scale space* of an image [6, 23, 24], and finding the extrema of simple operators. Each such potential interest point (which identifies a location and a scale) is then augmented with a descriptor that encodes the image patch around this point in a manner that is invariant to a local (rigid, affine, or perspective) transformation of the patch, for example using modes or histograms. The advantage of such invariant descriptors is that they can be stored in a database, and subsequently be retrieved to identify scene locations observed from different directions and distances, with applications in object recognition, image retrieval, tracking, and robot localization.

The success of these methods inspires the work in this paper, where we apply similar ideas to extract stable features from the scale space of one-dimensional panoramic images. Instead of creating a database of distinct, globally invariant features, however, the focus of our work is on features that are only locally invariant to changes in scale and viewing angle. That is, we are interested in extracting a large number of features in each frame, most of which will remain stable over small changes in viewpoint. This allows tracking of features through sequences of frames as the viewpoint changes. Robot localization can then be obtained by matching groups of features between the current frame and stored features of reference frames.

Figure 1 illustrates our experimental setup. It shows the robot equipped with an omnidirectional camera, a sample panoramic view, and the evolution of a 1D circular image over time as the robot travels.

Using one-dimensional images is appealing due to low storage requirements and fast processing times, which enable dense sampling and real-time analysis of views. The reduced dimensionality also aids greatly in image matching, since fewer parameters need to be estimated. However, there are also several factors that make it difficult to extract stable, globally invariant features from 1D omnidirectional images:

1. A single scanline does not carry very much information, and distinct features that can be matched reliably and uniquely over wide ranges of views are rare. A unique descriptor would have to span many pixels, thus requiring a certain minimum feature size, which in turn increases the chance of occlusion. A 1D view of a typical indoor environment with little texture, such as a corridor, may not contain any unique features.
2. For global invariance to viewpoints, the imaged scene has to lie in the plane traversed by the camera, i.e., a single horizontal slice. While such images can be obtained with specialized sensors such as “strip cameras” [15], or simply by extracting a single scanline from a view taken with an omnidirectional camera, it requires that the robot travels on a planar surface, which limits the applicability to indoor environments. Furthermore, it is difficult to precisely maintain the camera’s orientation due to vibrations of the moving platform [28].
3. Finally, achieving scale invariance by finding extrema in the scale space only works for planar projection. For circular or cylindrical projection, a scale change in the scene no longer corresponds to a uniform scale change in the image, unless the object subtends only a very small viewing angle.

For all of the above reasons, we adopt a different approach. First, we forego global uniqueness of features in favor of a large number of simple features. Unique matching will still be possible in many cases by considering groups of features [3]. The appeal of

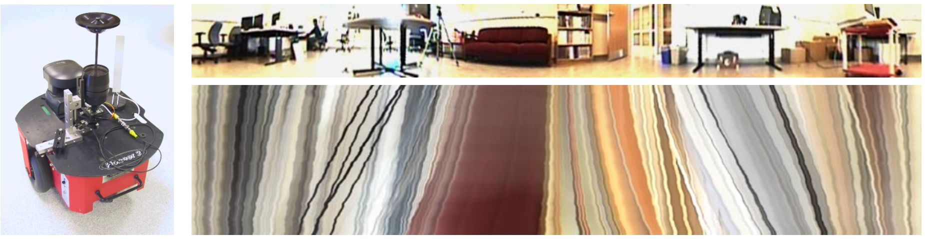

Fig. 1. Our robot with omnidirectional camera, a sample panoramic view, and the epipolar plane image (EPI). The EPI is a “stack” of one-dimensional images over time as the robot drives towards the bookshelf in the center of the image. Each such image is formed by averaging 50 scanlines in the center of the panoramic view. For pure translation of the camera, scene points traverse tangent curves in the EPI.

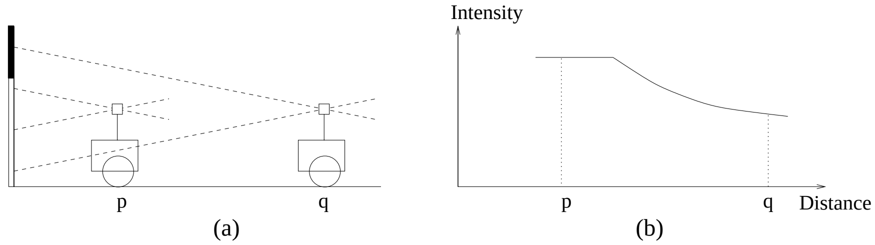

Fig. 2. (a) Averaging of scanlines increases robustness to vibration, but causes different parts of the scene to be imaged depending on the distance. (b) A sample distance-dependent intensity profile for the scenario shown in (a). The drop in intensities is caused by the black object coming into view. Note that intensities cannot change abruptly.

a scale-space approach is that interest points correspond to scene features of all sizes, ranging from small details such as a chair leg to large features, such as an entire wall of a room. Second, instead of using a single scanline, we compute the average over a range of scanlines, which is much less sensitive to small changes in camera orientation due to vibrations or uneven floor. Averaging scanlines, however, also introduces distance-dependent intensities changes, since the backprojection of a pixel into the scene now subtends a positive vertical angle (see Figure 2a). Thus, we trade distance-invariant intensities for robustness. Note, however, that intensities change smoothly with distance (see Figure 2b), which in turn causes smooth changes in the scale space. Finally, the lack of exact proportionality between object size and image size due to circular (rather than planar) projection is also not a problem for similar reasons: since the relationship is still smooth (proportional to the tangent of the object size), scale changes still correspond to smooth changes in the image.

2 Related Work

There has been considerable work in detecting invariant features in 2D images, including Lowe's SIFT detector [9, 10], which uses extrema in scale space for automatic scale selection, and the invariant interest points by Mikolajczyk and Schmid [12, 13]. Such features have numerous applications, including object recognition and image retrieval, as well as robot localization and navigation [18, 19]. Scale-space extrema have also been used for shape representation [4] and focus-of-attention [7]. A recent comparison of local image descriptors can be found in [14].

Analyzing an evolving scanline of a translating planar (not panoramic) camera is the classic epipolar-plane image (EPI) analysis approach [1]. This idea has been applied to panoramic views taken from a moving platform by Zhu et al. [28], with the application of dense matching and 3D reconstruction. Camera vibrations are compensated for using explicit image stabilization. In contrast, our work focuses on the tracking of interest points over long sequences and employs averaging of scanlines to obtain robustness to vibrations.

Much recent robotics work has focused on simultaneous localization and map building (SLAM) [5, 19, 22], which is related to the two-dimensional structure-from-motion problem [21]. Approaches to robot localization and map building from omnidirectional views include those of Zheng and Tsuji [27], and Yagi et al. [25, 26], who use vertical edges as features. Matsumoto et al. [11] present a method for robot navigation from memorized omnidirectional views, and Pajdla and Hlaváč [16] use the image phase of a panoramic view for robot localization. Panoramic views have also been used for motion estimation [20].

The work presented here extends our prior work on visually guided robot navigation. In previous work we have presented new path planning algorithms for navigation among artificial landmarks [2, 17]. The long-term goal of our research is to move to navigation using natural landmarks, and the features described here are an important component of extending our approach in this direction.

3 Scale-space analysis

The key idea of our method is to compute the scale space $S(\phi, \sigma)$ of each omnidirectional image $I(\phi)$ for a range of scales σ, and to detect locally scale-invariant interest points or “keypoints” in this space. The scale space is defined as the convolution of the image with a Gaussian $G(\phi, \sigma)$ over a range of scales σ:

$$S(\phi, \sigma) = I(\phi) * G(\phi, \sigma), \qquad \text{with} \quad G(\phi, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\phi^2/(2\sigma^2)}.$$

This convolution is slightly unusual due to the fact that the omnidirectional image I “wraps around”, i.e., $I(\phi + 2\pi) = I(\phi)$. In particular, there are no border effects, and the infinite integral of the convolution can be replaced with a finite integral with a convolution kernel G' that represents an infinite sum of Gaussians spaced 2π apart:

$$\begin{aligned} I(\phi) * G(\phi, \sigma) &= \int_{-\infty}^{\infty} I(\xi) G(\phi - \xi, \sigma) d\xi \\ &= \int_{0}^{2\pi} I(\xi) G'(\phi - \xi, \sigma) d\xi, \end{aligned}$$

where

$$G'(\phi, \sigma) = \sum_{n=-\infty}^{\infty} G(\phi + 2\pi n, \sigma).$$

While there is no closed-form solution for the infinite sum, it can easily be approximated by a finite sum since the tails of G quickly approach zero.

Similar to the related two-dimensional scale space approaches, we represent the scale space using a logarithmic scale for σ, so that neighboring values of σ in the discrete representation of S are a constant factor k apart. Theoretical and empirical motivation for this representation can be found in [10, 12]. For the results shown in this paper, we use $k = 2^{1/3}$, i.e., 3 samples per octave (doubling of σ). Figure 3a shows a gray-level representation of the convolution kernel G', and Figure 4b shows a sample scale space image.

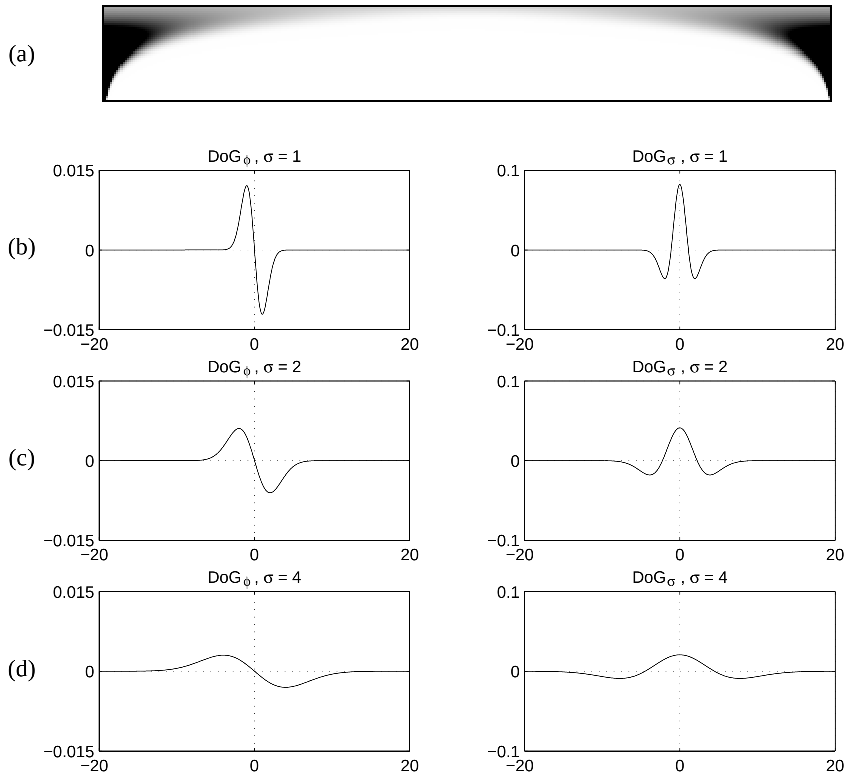

Fig. 3. (a) Gray-level plot of the circular convolution kernel G'. The horizontal axis is the image position $\phi = 0 \ldots 2\pi$; the vertical axis is the scale σ on a logarithmic scale, ranging over nine octaves: $\sigma = 2^{-1} \ldots 2^8$. (b–d) Effective difference of Gaussian (DoG) kernels resulting from computing column (ϕ) differences (left), and row (σ) differences (right), for smoothing scales $\sigma = 1, 2, 4$. To obtain comparable responses at each smoothing scale, the kernels on the left have been multiplied by σ.

After the scale space has been computed, we apply difference operators both vertically (between neighboring smoothing scales σ), and horizontally (between neighboring image locations ϕ). That is, we convolve the discrete version of $S(\phi, \sigma)$ with kernels $[-1, 1]^T$ and $[-1, 1]$, resulting in the difference scale spaces D_σ and D_ϕ, respectively. Differencing the scale space between neighboring values of σ (which differ by a constant factor k) is also done in the two-dimensional scale-space approaches [10]. It is equivalent to convolving the original image with a difference-of-Gaussian (DoG) operator ("Mexican hat operator"), as shown in the right-hand column of Figure 3b–d. This provides an approximation of the Laplacian of the image, which is an attractive 2D invariant due to its rotational symmetry. In our one-dimensional case, however, we assume a fixed orientation, and can thus also utilize the directional derivatives by subtracting horizontally neighboring image locations. Unlike in the vertical case, however, we have to multiply the differences by the corresponding value of σ to compensate for

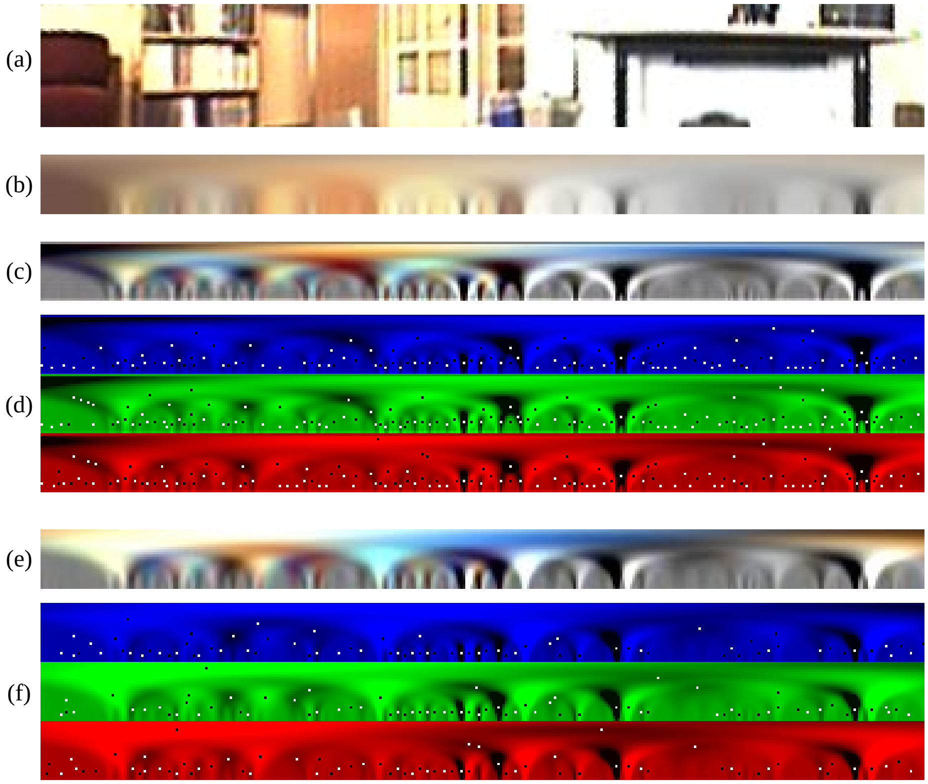

Fig. 4. (a) Part of the circular panorama shown in Figure 1. (b) The scale space S of the average of all 50 scanlines in (a), obtained by "circular convolution" with a kernel G' like the one shown in Figure 3a. (c) Differences of rows D_σ. (d) The three color bands of (c), with marked minima and maxima; these are candidates for stable features. (e) Differences of columns D_ϕ. (f) The three color bands of (e), with marked minima and maxima.

the decreasing height of the Gaussian. The resulting equivalent DoG operator is shown in the left-hand column of Figure 3b–d.

Finally, interest point selection proceeds by finding the minima and maxima of each of the three color bands of D_σ and D_ϕ. To do this, we consider all 3×3 neighborhoods in the six images, and check whether the center value is an extremum. We obtain subpixel estimates of both location ϕ and scale σ of all extrema by fitting a quadratic surface to the 3×3 neighborhood. This also provides estimates of the local curvature. The entire process of scale space computation, differencing, and interest point selection is illustrated in Figure 4.

The appeal of finding extrema in scale space is that it provides automatic estimates of both position and scale of features [8]. Intuitively, at each image location, the DoG kernel (Figure 3) that best matches the underlying image intensities is selected. We now investigate the stability and robustness of these features by tracking them through various image sequences.

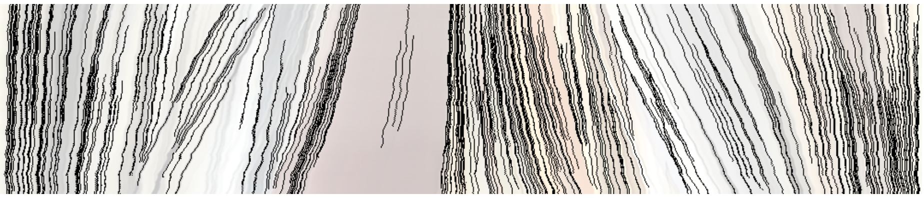

Fig. 5. Feature tracks overlaid on the EPI shown in Figure 1. Only feature tracks of length 100 or greater are shown.

4 Evaluating the robustness of features

A very good impression of the stability of the features can be obtained by observing a movie of images like those in Figures 4d and 4f. From viewing sequences of the 6 color bands with marked maxima and minima, it is immediately obvious that some extrema are very stable and persist for a long time, while other extrema are caused by noise and are very unstable.

4.1 Tracking features over frames

Given an image sequence with closely spaced views, a good way of measuring the robustness and stability of features is to track features from frame to frame, and to record the track length for each feature. Given a feature in the current frame, we search the next frame within a small neighborhood of the feature's current (ϕ, σ) location. We only consider neighboring scales σ (i.e. we use a vertical search radius of ± 1), since the scale of a feature cannot change very quickly. In the horizontal (ϕ) direction, we use a search radius of 2–4 pixels, depending on the motion present in the sequence. A feature is considered tracked only if there is an unambiguous match within the search window, both searching forward and backward (from the next to the current frame).

The presence of many robust features can be measured by counting the number of features whose tracks have a certain minimum length. This is demonstrated in Figure 5, where we show feature tracks that span at least 100 frames, overlaid on the EPI from Figure 1. It can be seen that despite significant scale changes and some occlusion, many features can be tracked reliably. Furthermore, the feature tracks closely follow the scene structure, which is promising if the features are to be used to estimate the robot's motion.

Instead of tracking a feature through many frames, it would clearly be useful to compute an indicator for a feature's stability from its own properties in a single frame, such as the size and shape of the scale-space surface in the vicinity of the extremum. We have performed many experiments investigating the correlation between a feature's track length and its properties, using a variety of image sequences. However, the only properties that show a slight correlation are the absolute value of extrema, and the curvature as given by the σ^2 and ϕ^2 terms of the quadratic surface fit. This is shown in Figure 6. These measures can be used to exclude a small number of unstable features with very short track lengths (using a threshold close to 0). A higher threshold, however,

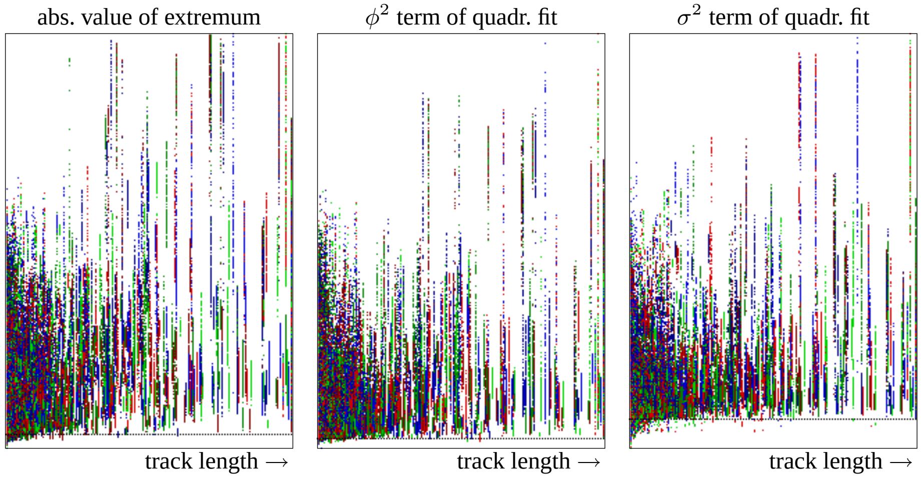

Fig. 6. Scatter plots of feature properties vs. track length for tracked features in D_σ of the EPI shown in Figure 1. The horizontal axis shows track length; the vertical axis shows the absolute value of the extremum (left), the horizontal curvature scaled by σ (middle), and the vertical curvature (right). The zero line is shown dashed. It can be seen that there is only a slight correlation between track length and each observed value. Plots for features in D_ϕ look similar.

would also exclude many features with very long track lengths. Thus, these measures do not enable the prediction of which features can be reliably tracked over many frames, and our remaining analysis of feature stability is therefore solely based on track length.

4.2 Robustness to varying orientation

To demonstrate the effectiveness of averaging scanlines in order to increase robustness to orientation changes, we have taken an image sequence while significantly changing the robot's tilt and yaw angles. Figure 7 compares the tracked features in an EPI generated from a single scanline, and one formed by averaging 50 scanlines. The improved performance by averaging scanlines is obvious.

4.3 Robustness to varying intensities

To evaluate robustness with respect to lighting conditions, we obtained an image sequence from the stationary robot while turning on and off the lights and opening and closing the window shades. The resulting sequence contains many frames in which most of the room is in almost complete darkness. Nevertheless, as shown in Figure 8, many features can be tracked through the entire sequence.

4.4 Effects of occlusion

In the final experiment reported here, we took an image sequence from the stationary robot while a person walked around it, thus temporarily occluding parts of the scene.

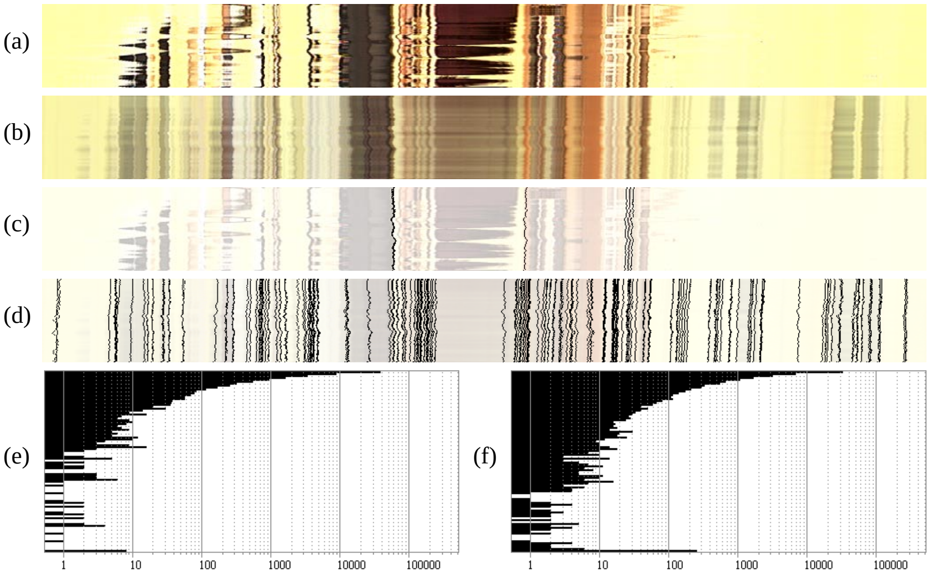

Fig. 7. Testing robustness to orientation changes using a “bumpy” sequence. (a) Single-scanline EPI. (b) EPI formed by averaging 50 scanlines. (c) Only 8 features can be tracked through all 94 frames of (a). (d) 253 features can be tracked through (b). (e,f) Histograms of track lengths (track length increases from top to bottom).

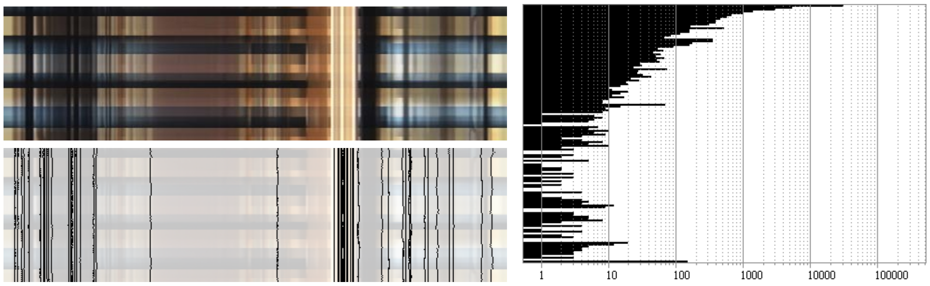

Fig. 8. Left top: Part of an EPI from an image sequence with changing lighting. Left bottom: Feature tracks that span the entire sequence. Right: The corresponding track length histogram. Note that there are many long tracks, despite the significant illumination changes.

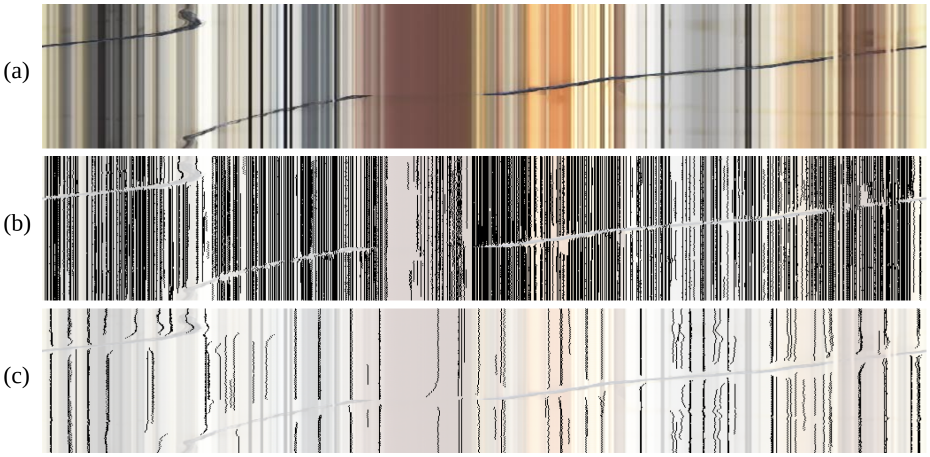

Fig. 9. (a) An EPI of a person walking around the robot. (b) Feature tracks spanning at least 20 frames, corresponding to features with smaller scales ($\sigma < 8$). (c) Feature tracks spanning at least 20 frames, corresponding to features with larger scales ($\sigma \geq 8$).

The resulting EPI and two plots of feature tracks with different scale ranges are shown in Figure 9. It can be seen that the features with smaller scales ($\sigma < 8$), which comprise the vast majority of all features, are unaffected by the neighboring occlusion. This is not the case for features with larger scales ($\sigma \geq 8$), whose tracks end many frames away from the actual occlusion event. This is not surprising, given the larger footprint of the smoothing kernels for higher values of σ, and has to be taken into account in feature matching algorithms. The stability of the majority of features, however, is a good prerequisite for employing robust matching techniques such as RANSAC.

5 Conclusion

In this paper we have defined interest operators for extracting stable features from the scale space of one-dimensional omnidirectional images. Interest points are selected by finding the minima and maxima of each of the three color bands of D_σ and D_ϕ, yielding many features in each of the 12 types. Our experimental results show that a large fraction of these features are stable, in the sense that they can be tracked through long sequences of frames, despite the presence of significant vibration, noise, lighting change, and nearby occlusion. An attractive property of scale-space extrema is their automatic scale selection, which avoids the sensitivity to parameter setting of traditional feature detectors.

The advantages of our approach are the reduced time and space complexities in dealing with one-dimensional images, which enables a dense sampling of views and therefore robust tracking of features through sequences of frames as the viewpoint changes.

We are currently working on an effective algorithm for matching features between more distant views. This can then be used during mobile robot navigation for course correction and localization by comparing feature locations in the current frame to those in stored reference frames. The long-term goal of our work is to provide robust alternatives to the complexities of full 2D image analysis in the context of vision-guided robot navigation.

Acknowledgements

The authors are grateful for the assistance of Vladimir Gurari in the implementation of early feature tracking strategies.

Support for this work was provided in part by the National Science Foundation under grants IIS-0118892, CAREER grant 9984485, POWRE grant EIA-9806108, and by Middlebury College.

References

1. R. C. Bolles and H. H. Baker. Epipolar-plane image analysis: A technique for analyzing motion sequences. Technical Report 377, AI Center, SRI International, February 1986.
2. A. Briggs, D. Scharstein, and S. Abbott. Reliable mobile robot navigation from unreliable visual cues. In Donald, Lynch, and Rus, editors, *Algorithmic and Computational Robotics: New Directions, A. K. Peters*, pages 349–362, 2001.
3. M. Brown and D. G. Lowe. Invariant features from interest point groups. In *Proceedings of the British Machine Vision Conference, Cardiff, Wales*, pages 656–665, September 2002.
4. J. L. Crowley and A. C. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):156–170, 1984.
5. M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localisation and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17(3):229–241, 2001.
6. T. Lindeberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):234–254, March 1990.
7. T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.
8. T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.
9. D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision, Corfu, Greece*, pages 1150–1157, September 1999.
10. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. To appear.
11. Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue. Visual navigation using omnidirectional view sequence. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'99)*, pages 317–322, 1999.
12. K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Proceedings of the International Conference on Computer Vision, Vancouver, Canada*, pages 525–531, July 2001.
13. K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proceedings of the European Conference on Computer Vision, Copenhagen*, volume 4, pages 700–714, May 2002.

14. K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison*, volume II, pages 257–263, June 2003.
15. S. K. Nayar and A. Karmarkar. 360 x 360 mosaics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina*, volume 2, pages 388–395, June 2000.
16. T. Pajdla and V. Hlaváč. Zero phase representation of panoramic images for image based localization. In *Proceedings of the Eighth International Conference on Computer Analysis of Images and Patterns, Ljubljana, Slovenia, Springer LNCS 1689*, pages 550–557, September 1999.
17. D. Scharstein and A. Briggs. Real-time recognition of self-similar landmarks. *Image and Vision Computing*, 19(11):763–772, September 2001.
18. S. Se, D. Lowe, and J. Little. Global localization using distinctive visual features. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems EPFL, Lausanne, Switzerland*, pages 226–231, October 2002.
19. S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, pages 735–758, August 2002.
20. T. Svoboda, T. Pajdla, and V. Hlaváč. Motion estimation using central panoramic cameras. In *Proceedings of the IEEE International Conference on Intelligent Vehicles, Stuttgart, Germany*, pages 335–340, October 1998.
21. C. Taylor, D. Kriegman, and P. Anandan. Structure and motion in two dimensions from multiple images: A least squares approach. In *IEEE Workshop on Visual Motion*, pages 242–248, Princeton, New Jersey, October 1991.
22. S. Thrun, D. Koller, Z. Ghahramani, and H. Durrant-Whyte. Simultaneous mapping and localization with sparse extended information filters: Theory and initial results. In J.-D. Boissonat, J. Burdick, K. Goldberg, and S. Hutchinson, editors, *Algorithmic Foundations of Robotics V*, pages 363–380. Springer-Verlag, 2004.
23. A. Witkin, D. Terzopoulos, and M. Kass. Signal matching through scale space. *International Journal of Computer Vision*, 1:133–144, 1987.
24. A. P. Witkin. Scale-space filtering. In *Eighth International Joint Conference on Artificial Intelligence (IJCAI-83)*, pages 1019–1022. Morgan Kaufmann Publishers, August 1983.
25. Y. Yagi, S. Kawato, and S. Tsuji. Real-time omnidirectional image sensor (COPIS) for vision-guided navigation. *IEEE Transactions on Robotics and Automation*, 10(1):1–12, 1994.
26. Y. Yagi, Y. Nishizawa, and M. Yachida. Map-based navigation for a mobile robot with omnidirectional image sensor COPIS. *IEEE Transactions on Robotics and Automation*, 11(5):634–648, 1995.
27. J. Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, 9(1):55–76, 1992.
28. Z. Zhu, G. Xu, and X. Lin. Panoramic EPI generation and analysis of video from a moving platform with vibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins*, pages 531–537, June 1999.