

Shoes as a Platform for Vision

Paul Fitzpatrick and Charles C. Kemp¹
Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Laboratory
Cambridge, MA 02139, USA
{paulfitz, cckemp}@ai.mit.edu

Abstract

We explore the use of a shoe-mounted camera as a sensory system for wearable computing. We demonstrate tools useful for gait analysis, obstacle detection, and context recognition. Using only visual information, we detect periods of stability and motion during walking. In the stable phase, the foot can be assumed to be parallel to the ground plane. In this condition, the floor dominates the lower part of the camera's view, and we show that it can be segmented out from the remainder of the scene, leaving walls and obstacles. We also demonstrate floor surface recognition for context awareness.

1. Introduction

Costs for digital cameras and computation continue to be driven lower by technological advances and strong demand. Future wearable computing system can benefit from these trends by applying cameras to new, more specialized, and less traditional sensing tasks. In this paper, we explore the use of a shoe-mounted camera for gait analysis, obstacle detection, and context recognition.

Wearable computing on shoes has been used for a variety of purposes, including user interfaces [6], power production [9], and gambling [10]. We show that visual processing can also benefit from this prime location. The planted foot is the only part of the body that is reliably stationary with respect to the world during walking and standing. When we walk, our feet come into contact with the ground in an alternating pattern. Each foot swings swiftly through the air, then is pressed against the ground as the weight of the body is transferred onto it [7]. During these key moments within a person's stride, the planted foot tends to be in a canonical orientation with respect to the floor and relatively motionless, which leads to simplified vision processing.

¹authors ordered alphabetically



Figure 1. The system. A camera and inertial sensor are mounted on a sandal.

In this paper, we analyze the wearer's gait and pick out frames corresponding with moments of stability when the foot is pressed against the floor. During these moments, the floor dominates the lower part of the camera's view, and we show that the floor can be segmented out from the remainder of the scene, leaving walls and obstacles. Next, we demonstrate floor surface recognition for context awareness. We conclude by speculating about the role a foot-mounted camera could play in future wearable systems.

2. The platform

Our system consists of a camera and three inertial sensors. The camera is mounted at the very front of a sandal as shown in Figure 1. It is rigidly attached to an inertial sensor. The remaining two inertial sensors are attached to the leg as shown – however they do not play a role in the component of this project described here. Data is logged on a laptop carried in a backpack. The system was tested on two floors of a building, on eight different surfaces (see Figure 6).

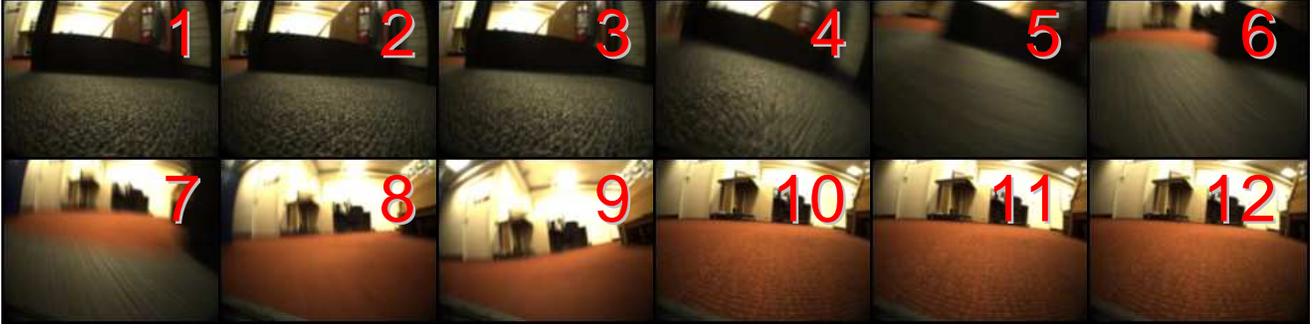


Figure 2. This figure shows a sequence of frames taken during a single step. In that step, the wearer moves from a lobby area into a corridor, and from a blue to red carpet. The main swing phase of the step occurs in frames five to nine.

3. Gait analysis

When the foot is pressed against the ground, the camera is in a fixed orientation with respect to the ground plane, and so this is an ideal opportunity for visual processing. Figure 2 shows images recorded during a single step. In the initial part of the swing phase, the floor in these frames becomes blurred (low spatial derivative ΔI_x), the view changes rapidly (high temporal derivative ΔI_t), and the foot turns downwards towards the floor (low average luminance I_0). Each of these cues could be used individually to identify the swing phase of walking. We combine them for robustness into a single measure s .

$$I_0 = \frac{1}{N} \sum_{x,y} I(x,y), \quad N = \sum_{x,y} 1 \quad (1)$$

$$\Delta I_t = \frac{1}{I_0 N} \sum_{x,y} |I(x,y,t) - I(x,y,t-1)| \quad (2)$$

$$\Delta I_x = \frac{1}{I_0 N} \sum_{x,y} |I(x,y,t) - I(x-1,y,t)| \quad (3)$$

$$s = \alpha \Delta I_t - \beta \Delta I_x - \gamma I_0 \quad (4)$$

Typically, during the swing phase, the summed magnitude of the temporal derivative of the pixels in the image increase dramatically, and the summed magnitude of the spatial derivative falls because blurring removes high-frequency edges. The spatial derivative is normalized for the overall average luminance since this changes as the camera moves from pointing towards the floor to pointing towards the ceiling (and lights). Since the wearer may move between different surfaces, the average component of ΔI_x and I_0 is removed using a running average.

Plots of these measurements for the step in Figure 2 are shown in Figure 3. As the step begins, the temporal

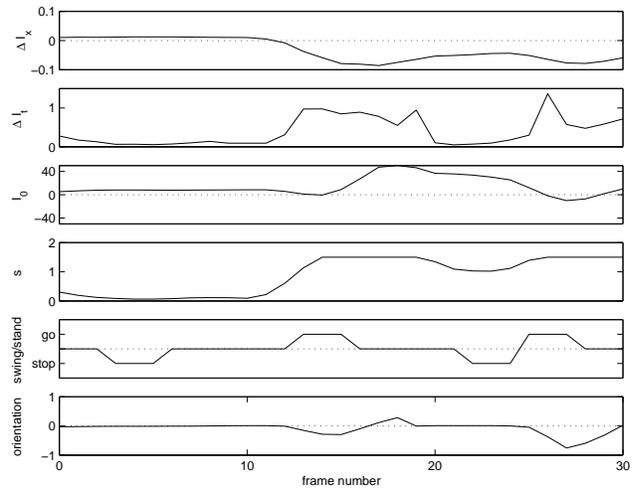


Figure 3. Visual and orientation information recorded during a single step (as shown in 2). The stride begins as the orientation value goes negative.

derivative increases (due to motion), the spatial derivative decreases (due to blur), and the mean luminance tends to fall (due to the camera looking towards the floor). Pitch information from the inertial sensor attached to the camera is also shown. We used this as an independent measure to verify the visual gait analysis. Whether the foot is in full swing, standing, or in an intermediate state is determined by analyzing s . A transition between steps is assumed to occur whenever this drops and rises again by at least 5% of its maximum range. To demonstrate this, a longer walking sequence is shown in Figure 5. Periods of stability detected from the gait analysis are processed to achieve floor segmentation and recognition.

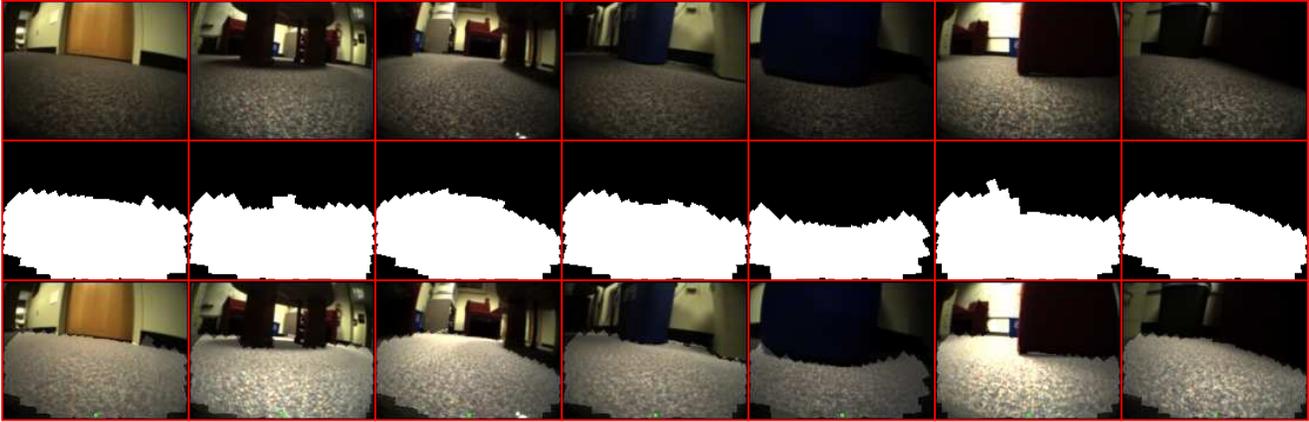


Figure 4. Floor segmentation in action. The top row shows original images, the second row shows masks corresponding to the floor, and the bottom row overlays the two.

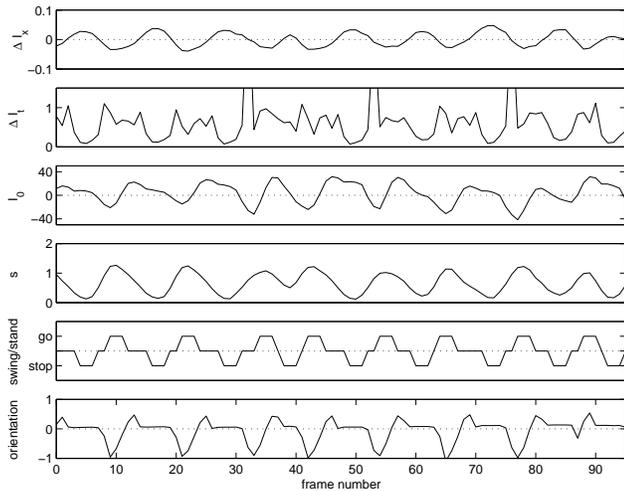


Figure 5. Visual and orientation information recorded while walking down a corridor. The stride period is visible in all modalities.

4 Floor segmentation

Due to the canonical orientation of the stable images selected by our gait analysis, we know with high probability that the bottom quarter of the image is floor and that the top quarter of the image is not floor. The bottom quarter of these special images corresponds with the area from the toe out to 3 inches on the floor. The top quarter of these images is very far above the horizon line. In this section and the next, we exploit this property to perform segmentation and recognition. With segmentation, this observation allows us

to collect a significant sample of the appearance of the floor and non-floor parts of the image. With recognition, we are able to reliably sample from the floor without the need for any floor detection algorithms.

These images are well suited to wearable computing applications that benefit from detailed sensing of the wearer’s nearby environment (for example, detection of walking hazards [5, 8]). Segmenting the floor in the image can serve as a first step to analyzing the free space, objects, and obstacles close to the wearer, out to about 6 feet with our wide angle lens. We show example results from our floor segmentation algorithm in Figure 4. If we assume flat floors, we can construct a function that for each pixel gives the distance from the toe to the corresponding point on the floor [3], which could be useful for object avoidance.

The top quarter and bottom quarter of the image are used to initialize two probabilistic appearance models, one for the floor and one for the non-floor parts of the image. These two appearance models and the resulting segmentation are iteratively optimized using EM (expectation maximization) to find a maximum likelihood segmentation of the floor and non-floor. The segmentation is constrained to be a set of radial distances emanating from the center of the bottom of the image [4].

5. Recognition

Areas with different functions often have distinct floor surfaces (see Figure 6). For example, a wash room floor is unlikely to be carpeted so that it can be easily mopped. Floor recognition is therefore a valuable cue for localization and context awareness [1].

We have used camera placement to essentially solve the problem of floor detection, as detailed in Sections 3 and 4.



Figure 6. Floors encountered. Four are carpeted, four are not. The floors are drawn from corridors, an office, lab space, a kitchen area and a wash room. There is considerable variety in color and texture.

If we model the appearance of the lower part of the image, this will be dominated by the floor. As a proof of concept, we simply compute the average color of this part of the image (normalized for luminance) and use it to represent the floor. We took a training sample from each of eight qualitatively different floors in our building. We tested a large set of other samples, comparing them with the models using a simple Euclidean distance metric. The confusion matrix is shown in Table 1. A total of 86.3% of the classifications are correct. A classifier that always guessed “floor number 1” (the most frequent case in the data) would have a 44% success rate. The number of samples of each floor are different since the data was collected by simply walking around, and the areas covered by the different floor types are of different sizes. The greatest confusion present is between two similar reddish-hued carpets (floors number 7 and 8).

6. Discussion and conclusions

We have shown that a foot-mounted camera is well placed for a number of sensory tasks relevant to wearable computing applications. Specifically, we have presented methods and results for gait analysis, floor segmentation, and floor recognition based solely on images from the camera. In general, as cameras and computation become less costly we expect for more specialized camera sensing, such as this, to become practical for wearable computing. Issues of privacy and misuse could be mitigated by making a closed sensory system. A camera on each foot would make several applications easier by allowing for nearly uninterrupted acquisition of closely-spaced stable images. Several interesting future applications might be built on top of the results we have presented including automated cartography, localization, detection of nearby people by their feet and legs, more powerful floor recognition systems, outdoor operation, and recognition of common nearby objects such as chairs, tables, walls and trash cans.

floor	classification frequencies for floor samples							
	1	2	3	4	5	6	7	8
1	194	0	28	14	2	13	0	0
2	0	30	0	0	0	0	0	0
3	0	0	4	0	0	0	0	0
4	2	0	3	37	0	0	0	0
5	0	0	0	0	31	0	0	0
6	0	0	0	0	0	149	0	0
7	0	0	0	0	0	0	36	15
8	0	0	0	0	0	0	0	6

Table 1. The confusion matrix for floor recognition. All off-diagonal entries are errors.

Acknowledgements

We would like to thank the reviewers for their many helpful comments. Funds were provided by DARPA (DABT 63-00-C-10102), and by the Nippon Telegraph and Telephone Corporation under the NTT/MIT Collaboration Agreement.

References

- [1] B. Clarkson, A. Pentland, and K. Mase. Recognizing user context via wearable sensors. In *Proc. of the Fourth Intl. Symposium on Wearable Computers*, pages 69–76, 2002.
- [2] F. Dellaert, W. Burgard, D. Fox, and S. Thrun. Using the condensation algorithm for robust, vision-based mobile robot localization. In *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 1999.
- [3] I. Horswill. Polly: A vision-based artificial agent. In *Proc. of the 11th National Conference on Artificial Intelligence*, pages 824–829, Menlo Park, CA, 1993.
- [4] C. C. Kemp. *Thesis Proposal: Humans as Robots*. PhD thesis, MIT, December 2002.
- [5] C. M. Lee, K. E. Schroder, and E. J. Seibel. Efficient image segmentation of walking hazards using IR illumination in wearable low vision aids. In *Proc. of the Sixth Intl. Symposium on Wearable Computers*, pages 127–128, 2002.
- [6] J. A. Paradiso, K. Hsiao, A. Y. Benbasat, and Z. Teegarden. Design and implementation of expressive footwear. *IBM Systems Journal*, 39(3/4):511–529, October 2000.
- [7] J. Pratt. *Exploiting Inherent Robustness and Natural Dynamics in the Control of Bipedal Walking Robots*. PhD thesis, MIT, Cambridge, Massachusetts, 2000.
- [8] S. Shoval, I. Ulrich, and J. Borenstein. NavBelt and the GuideCane. *IEEE Robotics and Automation*, 10(1):9–20, March 2003.
- [9] T. Starner. Human-powered wearable computing. *IBM Systems Journal*, 35(3/4):618–629, 1996.
- [10] E. Thorp. The invention of the first wearable computer. In *Proc. of the Second International Symposium on Wearable Computers*, pages 4–8, October 1998.