The Accuracy-Obtrusiveness Tradeoff for Wearable Vision Platforms

David S. Hayden, Carl Vondrick, Stella X. Jia, Yafim Landa,
Robert C. Miller, Antonio Torralba, Seth Teller
Massachusetts Institute of Technology
{dshayden, vondrick, xiaodan, landa, rcm, torralba, teller}@mit.edu

1. Introduction

We present our ongoing efforts to develop an open wearable platform for egocentric vision research. Our primary goal is to build a physically and socially comfortable wearable suit that still provides useful data on everyday activities for egocentric vision algorithms. Unfortunately, we have found that visibly-worn cameras are not socially acceptable. While mounting fast, wide-angle lenses to the head is ideal for vision algorithms, our results indicate that this will cause the wearer to be ostracized even though performance gains are limited.

In response, we present a simple, inconspicuous wearable platform (shown in Fig.1) that balances the performance of vision algorithms with a discreet design that allows the operator to blend into everyday life. We motivate and describe its design in the next three sections:

Effects of Cameras and Lenses on Vision Algorithms: While it is obvious that having cameras with fast, wide-angle lenses would be ideal for wearable vision, it is less clear to what degree they are beneficial. In section 2, we take a first step at quantifying this. In particular, we demonstrate how variations in camera resolution, shutter type, shutter speed, and lens speed affect the performance of Viola-Jones face detection [4] when applied to unconstrained wearable data. Our results suggest that although performance is improved by using cameras with fast lenses and shutter speeds, the extra bulk may not be justifiable.

Social Acceptability of Wearable Cameras: We have found that building a socially acceptable wearable platform is crucial to broadly explore applications of egocentric vision in daily life. While [2] discusses how face-to-face communications are affected by a head-mounted display, we more broadly investigate how wearable cameras are perceived by the public. Results are in section 3.

Platform for Egocentric Vision: In section 4, we introduce a discreet wearable jacket platform for egocentric vision research. Our preliminary findings indicate that this jacket is a reasonable tradeoff between what is physically and socially comfortable, to what is effective for egocentric vision. We believe this jacket will provide a productive avenue for vision researchers to share code, evaluate methods, and develop interactive egocentric algorithms.

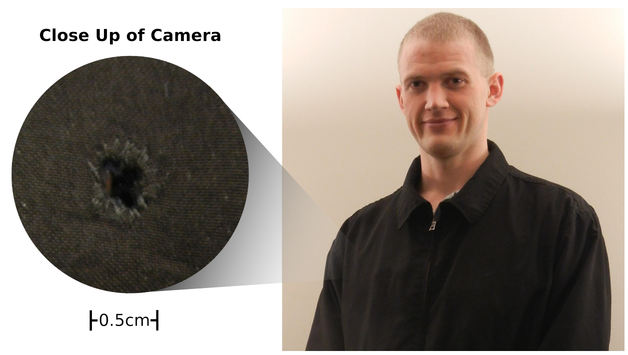

Figure 1: Our wearable prototype. By trading a limited amount of detection performance for a discreet design, we're able to comfortably use the prototype in everyday activities.

2. Egocentric Vision

Designing a reliable wearable assistant is challenging due to the unconstrained nature of egocentric vision. While the state-of-the-art object detectors perform well on standard vision benchmark datasets, the objects in these data sets are typically well-framed and well-exposed, with little to no motion blur. In contrast, an egocentric camera and object detector must be able to handle extreme motion, adversarial views, and dramatic changes in lighting. While in theory each could be addressed by mounting large, wide-angle lenses atop the head that quickly admit substantial light, enough to keep motion blur down, the result would then be bulky, inconvenient, and socially awkward. Rather, we must strike a balance between what is comfortable and socially acceptable to what is effective for object detectors.

Dataset: In order to evaluate vision algorithms on wearable cameras, we collected a difficult, real-world dataset of an operator wearing varying cameras shown in Tab.1. Cameras and lens combinations were chosen for their small size, each weighing less than a third of a pound in total. Cameras

Figure 2: Sample face detections in our wearable dataset; images are ordered by camera, as listed in Tab.1

were placed on the chest, which mimics an appealing wearable device where the camera and computers are sewn into the clothing. [1] presents geometric arguments for the chest being the second-best location for a wearable camera with respect to field of view and expected motion (following the head). The data was collected in MIT's high-traffic computer science building during normal business hours.

Results: We evaluated Viola-Jones face detection on 100 randomly selected frames from each of the wearable camera video sequences. Detection rates are summarized in Tab.1 and sample images are shown in Fig.2. Detections were correct so long as they fired on faces turned within 90°.

Image Quality: Images from the first Firefly trial are noticeably darker than the others and images from the Playstation Eye are significantly more noisy. Although the iPhone had a slower lens than the Lifecam, its images remained brighter, likely due to its backlit sensor and increased resolution, providing additional tolerance for noise caused by gain. Severe blur was common among the cameras with 34ms exposures, but even they had some clear shots, likely taken when one foot was on the ground, and the other rising.

Detection Performance: Precision and recall is very poor for all cameras, with none exceeding 0.4 precision or recall. Abstract faces were found in many and varied objects–trash cans, billboards, lighting installations, and ATMs among them. Among the correct person detections, most were within 10 feet and walking towards the camera but variations in pose that approached 90° were sometimes tolerated. The seemingly low false-negatives of the first and fourth trials can be explained by the severely limited contrast and resolution, respectively–in either case there simply was not enough signal for Viola-Jones to misfire as frequently as it did on more reasonable trials.

Conclusion: The Firefly gave the worst performance when its shutter speed was an aggressively fast 3ms, and the best performance when it was a more moderate 15ms. All other cameras yielded equivalent performances despite variation in resolution and lens speed. Rolling shutters are acceptable since, empirically, only 1 in 300 frames had obvious shutter artifacts. Our results suggest that anything short of a camera with a lens faster than F/2.0 and shutter speed around 15ms may perform similarly, and that all perform poorly on unconstrained wearable data. In light of this, it may be best to design with failure in mind, and to choose cameras based on other criteria.

Figure 3: Pictures of wearable configurations shown to the public for comment, ordered by social acceptability. From left to right, we have control (no wearable), ear (Bluetooth camera), chest, pinwheel, head, and face. Though perhaps absurd, the pinwheel is used as a point of reference.

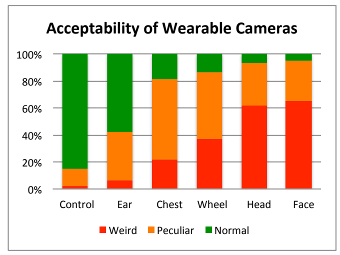

Figure 4: How does the public feel about wearable cameras? These results suggest that the best wearable camera may be the one that is not seen.

3. Social Acceptability of Wearables

In order for a wearable assistant to be useful in daily living, it must not interfere with the user's ability to engage in their environment. Ideally, it must not provide useless or false information, nor should it interrupt the user from daily tasks such as conversation. Yet, even if wearable devices can be made useful, their adoption will be limited unless the public accepts their use. In this section, we attempt to quantify public opinion on wearable cameras with a Mechanical Turk study and by shadowing subjects wearing cameras.

Acceptability Study: We asked 100 US MTurk workers to comment on wearable cameras. We showed workers 4 subjects (3 male, 1 female) wearing different cameras as shown in Fig.3. Workers were asked to rate each photo as normal, peculiar, or weird, with the setting being that they encountered the subject in a grocery store. Our results in Fig.4 indicate that while cameras visibly mounted to the head or chest may offer the highest-quality data, they simply are not socially acceptable. If a camera must be visible, then a Bluetooth-sized camera on the ear is preferable. Interestingly, the pinwheel, although perhaps the most absurd, is more acceptable than are cameras mounted on the head!

Privacy: Privacy was a major concern for the workers:

> Any guy that approaches me in a store with an attachable camera to his head pointed at me, is not going to get anything other than small talk from me. Also, I will be sure to allow them to walk to their car first in the parking lot. No way will they film my license plate number.

While a wearable assistant would likely discard camera data after processing, broadly convincing people that the device is actually doing so may prove to be difficult.

	Parameters				Detection Results				
Camera	Resolution	F-Number	Max Exposure (ms)	Shutter	TP	FP	FN	Precision	Recall
Point Grey Firefly	640 x 480	f/1.3	03	Global	7	98	27	0.0667	0.20
Point Grey Firefly	640 x 480	f/1.3	15	Global	35	75	72	0.3182	0.3271
Microsoft Lifecam	640 x 480	f/2.0	34	Rolling	20	89	77	0.1835	0.2062
Sony Playstation Eye	320 x 240	f/2.1	09	Rolling	20	80	33	0.20	0.37
Apple iPhone 4S	1920 x 1080	f/2.4	34	Rolling	24	109	83	0.1805	0.2243

Table 1: Face detections on 100 randomly selected frames from sequences of wearable video data collected from low-profile cameras. All were operated with fixed focus and all but the Firefly were operated with auto exposure and gain.

Public Use: Many retail and entertainment venues do not allow the use of cameras. In our testing, we found that subjects wearing cameras were quickly asked to put them away or to leave when wearing them in public venues including grocery stores, restaurants, and bars. Owners were worried that the camera would record credit card numbers or steal product information. Moreover, we found that police refused to talk to subjects who wore a camera.

Status: Finally, wearables are unacceptable to some:

> Why is the camera strapped to him like a baby? The apparatus is way too strange. Despite his friendly expression, I might be too concerned with his mental well-being to be very forthcoming in a conversation.

Conclusion: Visible wearable cameras are socially detrimental to the wearer. We believe that making them invisible, or nearly so, will lead to broader adoption.

4. Wearable Platform

Our preliminary experiments suggest that fitness for egocentric vision must be balanced with social acceptability. To that end, we are developing a wearable jacket that is socially acceptable in the sense that it goes unnoticed to the casual observer. Fig.1 shows this prototype where one or two cameras are mounted to the chest, obscured by the wearer's clothing. Although the form factor restricts us from using faster lenses, we found social acceptability to be paramount.

The Jacket: Our wearable platform is simple to build and seamlessly integrates into the operator's clothing: one or more phones are securely contained with a zippered pouch that is sewn into the interior lining of a light jacket at the breast. The phones are vertically oriented for comfort. Though smartphones are not built for this purpose, they are well-suited to it. The Linux environment coupled with multicore processing, camera, microphones, GPS, motion sensors, Bluetooth, 4G, built-in power, and dimensions below 10mm thickness and 130g weight permit a comfortable, discreet jacket that only needs to charge at night.

The Wearable Assistant: We believe our jacket provides an interesting platform for the development of wearable assistants that can provide information, such as peoples' identities (e.g. [3]), at opportune times. By using smartphone cameras mounted to the chest, we have compromised the accuracy of egocentric vision algorithms—camera framing will frequently be poor and motion blur heavy. However, we note that egocentric vision appears to perform poorly on unconstrained data, regardless of camera. By offering a modicum of physical and social comfort, our jacket makes for a platform that can be worn regularly and exposed to unconstrained data outside of lab settings.

5. Discussion

We have provided a baseline on how cameras with differing resolutions, shutter speeds, and lens properties affect Viola-Jones face detection when applied to unconstrained wearable data. Our analysis reveals that anything short of a camera with a fast lens and a shutter speed around 15ms performs about the same, and that performance is poor regardless. We have additionally addressed how socially acceptable wearable cameras are to the general public. Our experiments suggest that the best wearable cameras may be those that are not seen. Armed with this knowledge, we propose a wearable platform that optimizes for the physical and social comfort of the wearer rather than for camera placement and computational ability.

References

[1] W. Mayol-Cuevas, B. Tordoff, and D. Murray. On the choice and placement of wearable vision sensors. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(2):414–425, 2009. 2

[2] G. McAtamney and C. Parker. An examination of the effects of a wearable display on informal face-to-face communication. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 45–54. ACM, 2006. 1

[3] B. Singletary and T. Starner. Symbiotic interfaces for wearable face recognition. In *HCII2001 Workshop On Wearable Computing, New Orleans, LA*. Citeseer, 2001. 3

[4] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153–161, 2005. 1