Applying Active Vision and SLAM to Wearables*

Walterio W Mayol, Andrew J Davison, Ben J Tordoff and David W Murray

Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK
wmayol,ajd,bjt,dwm@robots.ox.ac.uk
http://www.robots.ox.ac.uk/ActiveVision/

Abstract. This paper reviews aspects of the design and construction of an active wearable camera, and describes progress in equipping it with visual processing for reactive tasks like orientation stabilisation, slaving from head movements, and 2D tracking. The paper goes on to describe a first application of frame-rate simultaneous localisation and mapping (SLAM) to the wearable camera. Though relevant for any single camera undergoing general motion, the approach has particular benefits in wearable vision, allowing extended periods of purposive fixation followed by controlled redirection of gaze to other parts of the scene.

1 Introduction

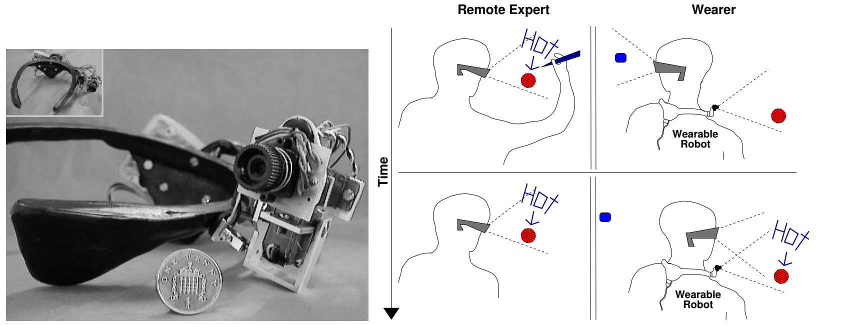
Key to the advancement of wearable computing beyond the superficial stitching of processors and memory into this or that garment is the development of input sensors and output interfaces which provide a rich first-person perspective on the interaction between wearer and environment, and which prove socially acceptable[1]. Vision fits well into the wearable domain, offering many (perhaps too many) non-intrusive ways of recovering the cues that allow us to navigate, to touch and grasp, to follow action, to understand gesturing and signage, and to react to events. It is also a modality whose computational demands are becoming less incompatible with the power of wearable cpu's.

A number of authors have considered deploying static cameras on a wearer's clothing (eg [16,13,4]). Although ideal for recovering ambient information, such as lighting levels, colour histograms, and so on, for more specific tasks the resulting imagery is too dependent on the wearer's exact posture and motions. A substantially different approach is to attach cameras to particular body parts. In [1], for example, a hat mounted camera looks outwards to where the wearer's face points, whereas in [20] the hat mounted camera looks downwards at the wearer's hands and reaching space. In [23] cameras are strapped to the wearer's hands to get this type of view. Now the views are highly specialised, but are often *more* susceptible to the vagaries of the wearer's changing stance.

Many of the issues in wearable computing have been successfully solved in the animal world, and reaching a compromise between high resolution imagery and a small number of low volume sensors has resulted in moving eyes. In recent work [18],

* Research supported by UK EPSRC Grant GR/N03266, by a Mexican CONACYT scholarship to WWM and by an EPSRC Advanced Research Fellowship to AJD.

[1] Social acceptability is not driven by a woolly desire to 'look good', but rather by the experimental imperative of not altering the interaction under observation.



(a) (b)

Fig. 1. (a) The wearable camera has three motorised axes, and is worn on the shoulder. (b) A remote expert annotates the environment he interacts with via a wearable visual robot.

we have begun to marry techniques from active vision with wearable computing, and have produced the lightweight, directable, wearable camera shown in Figure 1(a). By gaining a degree of independence from the wearer, the camera can work in one of three sets of frames. The first set relates to the wearer's activity, for example, sensing the manipulative space in front of the wearer's chest, or pointing where the wearer's head is pointing. The second set includes frames aligned with the static surroundings — for example, the camera might wish to maintain alignment with gravity; and third is the set of frames defined by the wearer's position relative to independent moving objects, particularly relevant to object tracking. Perhaps another point in favour of the animated camera is that a sensor able to indicate where it is looking – and hence where it is *not* looking – appears more acceptable than the unwavering stare of a passive device.

Although a long term goal must be to deploy the device as the front end of an autonomous assistant, a more immediately realisable use is that of the teleoperated eye of a remote human assistant, as depicted in Figure 1(b). The example shows a remote assistant pre-labelling an item of interest in the wearer's surroundings. Note that the example already exploits the decoupling of the wearer's attention from that of the camera and its operator.

Section 2 of the paper reviews some of the issues considered in the design and build of our device, and Section 3 lists aspects of the reactive visual processing explored to date. Sections 4 and 5 describe a novel implementation of single camera simultaneous localisation and mapping (SLAM) and its application on the wearable.

2 The wearable device

Figure 2(a) shows the configuration of elevation, pan and cyclotorsion axes used to give the wearable complete rotational freedom, with the kinematic chain chosen such that singularities occur only in unlikely body postures. The axes are driven by miniature HiTEC HS-50 servos combining motor, controller and gear head in a volume of about 6 cm^3 with a mass of 6 g. Each produces a relatively high axis torque of 0.06 Nm and drives its axis at speeds up to 10 rad.s^{-1} over an angular

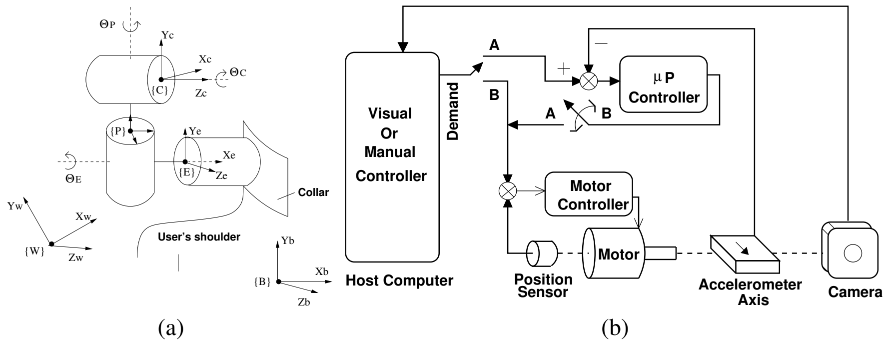


Fig. 2. (a) From collar to camera the axes are arranged as elevation, pan, and cyclotorsion, the last being aligned with the camera's optic axis. (b) The control architecture. Two main switchable control paths are possible. Path A uses inertial information as feedback to refer the camera to a world coordinate frame (gravity). Path B relies on servo-motor control to make camera movements relative to wearer's body. Vision can be combined with either path for tracking.

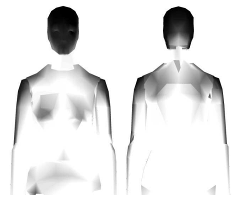
range of 160°. The motors have modest power consumption as they are able to hold position without being energised.

The detailed mechanical layout has been subject to optimisation against seven criteria, all of which should be minimised: (i) working volume; (ii) self collisions between motors and sensors, which would otherwise reduce the angular sensing range; (iii) integrated torque or energy consumption; (iv) maximum torque; and (v-vii) the three radii of rotation with respect to the centre of the image sensor. An optimisation algorithm generated variations of the 3D coordinates of camera and motors and after these seven criteria were evaluated for each variation, a Pareto test of non-dominance was applied to select the best candidates from which the final robot was built [17].

A multiple micro-controller architecture is used to interface the robot with the host computer, and is shown simplified in Figure 2(b). Although vision is also used to stabilise gaze direction, it is also useful to have a sense of absolute direction. The two-axis accelerometers used to stabilise the camera with respect to gravity have recently been augmented with an Intersense InertiaCube2 orientation sensor. The camera head is a 640x480 pixel Sony colour CCD, specially separated from its associated IEEE-1394 transfer electronics.

2.1 Locating a camera on the body

A variety of ad hoc solutions for placing wearable cameras on the human body appear in the literature. A more formal method of assessment has been devised by scoring (i) the absolute field-of-view, (ii) the view of handling space, and (iii) the resilience against wearer motion during locomotion for a simulated camera placed on each of the 1800 polygons of a human body model in turn. The polygons are arranged into 16 body-segments which can be independently rotated to simulate different body poses. For absolute field of view studies, rays are traced from the 4π



Zone	Field of view	Best suited to task	Body Motion	Social acceptability
Head	wide⋆	linked to wearer's attention	large	low
Shoulder	wide⋆	most tasks⋆	small⋆	medium(⋆)
Chest	limited	linked to wearer's workspace	small⋆	high⋆

Fig. 3. Front and back views showing how well body positions score against the criteria, with scores shown darker. The table shows some attributes of placing a camera on the Head, Shoulder and Chest. "⋆" denotes a feature which is desirable in all frames of reference.

solid angle around the camera and tested for intersection with the body. For motion resilience, the mean amount of movement of each polygon is determined over a range of simulated body motions. Maps of the sort shown in Figure 3 are obtained, here showing that the head and shoulders are favoured locations. If one also requires that the camera's motion not be tied to wearer's attention, the shoulder area used in our work wins through.

3 Examples of visual processing

We now outline some of the 2D visual competences implemented on the wearable.

1. **Visual and inertial stabilization.** An example of using both visual and inertial feedback to stabilise against body movement is given in Figure 4. The left column shows stills cut from a sequence during which the wearer sits down. The middle column shows the images from a static wearable camera, and the right column shows the resulting imagery from the active wearable when accelerometers were used to take out gross rotation, and vision is used to remove any remainder. This residual rotation was determined by finding, and matching from frame to frame, corner features $\boldsymbol{x}_i$ imaged in frame i as $\boldsymbol{x}_i = \mathtt{K}_i[\mathtt{R}_i|\boldsymbol{t}_i]\boldsymbol{X}$ from the static background $\boldsymbol{X}$. Here $\mathtt{K}_i$ is the camera's intrinsic calibration, and $\mathtt{R}_i$ and $\boldsymbol{t}_i$ are the rotation and translation relative to the reference frame. For pure rotation $\boldsymbol{t}$ is zero, and the inter-frame matches are fitted by an infinity homography $\boldsymbol{x}_j = \mathtt{H}_{ji}\boldsymbol{x}_i$, which may then be

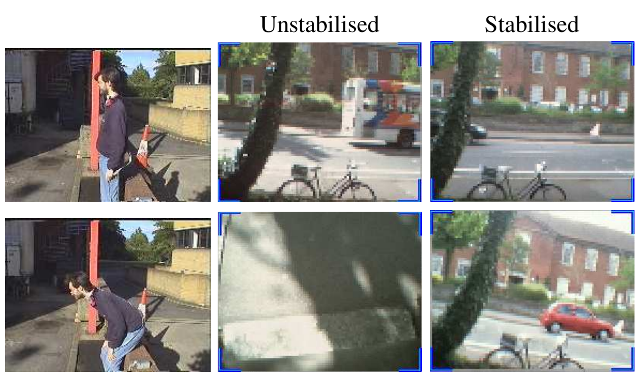


Fig. 4. Samples from a sequence obtained while the wearer sits down in an outdoor scene.

(a) (b)

Fig. 5. (a) The operator's head motion mapped onto the wearable. (b) A polyhedral shape model with a trained colour model superimposed forms the representation for tracking the head. Sample matches (red squares) used to calculate a motion hypothesis and the other matches (black crosses) which support the pose hypothesis.

decomposed as $\mathtt{H}_{ji} = \mathtt{K}_j \mathtt{R}_j \mathtt{R}_i^{-1} \mathtt{K}_i^{-1}$. The calibration is known and fixed, allowing recovery of the inter-frame rotation, which is used here to reposition a software fovea within the full image.

2. **Head tracking and servoing.** Figure 5(a) shows the wearable device being controlled from observations of a moving head. In the context of teleoperation, these may be the head movements of a remote operator, perhaps acting as the wearer's assistant. In other circumstances the head motions might be those of a person whose direction of attention the wearable is following. These movements are often larger than those experienced in typical video-conferencing applications, and tracking frontal facial features proved insufficient. The method illustrated in Figure 5(b) uses a learned colour model of the head to augment a feature tracker [22].

3. **Object tracking.** Tracking a region of interest through a sequence of images is a fundamental task in many vision systems and one for which there exists a plethora of methods. Figure 6 shows stills cut from a sequence taken through the wearable camera while both wearer and object of interest are moving. This particular example uses an algorithm based on first learning expected motions of a planar patch and then estimating the current image transformation based on the learned data. It is able to handle moderate occlusion and permits dynamic change of the scale and size of the sampling pattern. A mean shift tracker [5] has also been implemented. This tracks colour similarity, gaining robustness to 3D deformation by neglecting shape similarity, and proves particular useful when viewing hands.

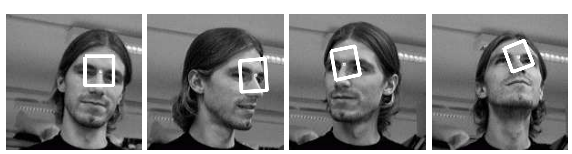


Fig. 6. Tracking an independent object from a non static camera. The polygon showing the approximated affine region recovered.

4 Wearable Vision and Localisation

The visual routines described in the previous section are reactive, in that the wearable camera merely responds to change in the image and scene. To undertake the more purposive movements characteristic of autonomy the camera needs to be able to establish its location in an initially unknown environment, and to be able to re-locate itself within that environment at a later time. Real-time localisation is of considerable concern to the wearable computing community, but current visual methods combine inertial sensing with visual measurement of either special fiducial markers [11] or pre-mapped targets [14]. Here we propose using vision alone, and eschew the use of either fabricated features or a known map, preferring instead to develop a general simultaneous localization and mapping (SLAM) approach.

In earlier work [6], two of the authors noted that the egomotion recovery in visual navigation was rather different problem from that of recovering sequential or dynamic structure from motion (SFM). In SFM the emphasis was on obtaining reasonably dense structure, often for model generation, whereas in navigation the emphasis is on knowledge of the accuracy of location over time. Such accuracy could only be obtained by maintaining a full motion and structure state covariance, and computational complexity demands this be restricted to a relatively sparse set of reliable and relocatable landmark features. Using even large numbers of independent feature-by-feature covariances inevitably leads to motion drift, a defect illustrated by the SFM navigation results in [2]. The contrast between SLAM and SFM approaches persists. For example [3] presents a real-time, full-covariance Kalman Filter approach to sequential SFM from a single camera, but the feature detection and tracking method does not allow features to be re-acquired after periods of neglect, and motion drift must arise. In our previous work it was also noted that relocating landmarks was a problem well suited to the active vision approach where directable cameras were able to seek out matches. Although these arguments were rehearsed independently of the burgeoning work in localization using other sensors (eg [21,10,15]), our resulting single EKF algorithm was nothing but visual SLAM [8,9], and introduced the valuable idea of update postponement [6], pursued more fully in [12].

Here we suggest that sequential localization using landmarks is also a process highly suited to *wearable* vision, the camera making use of the same data type as its human wearer and potential remote collaborator. Landmarks become objects of intrinsic interest, and tasks like that of visual annotation mentioned in the introduction becomes natural. One might imagine that if the wearable camera, even with its ability to redirect gaze, cannot see a particular landmark, it could request the wearer to move so as to allow it to see the feature.

4.1 Single camera SLAM for general motion

Single camera SLAM in which the camera is allowed to move generally in 3D is a challenging problem [7], because neither single nor multiple views of a single point yield depth when the motion is unknown. Information comes from points collectively, but as processing has to be completed in a fixed time a limit must be imposed on the feature map size. To achieve this, here we restrict the workspace to a few cubic metres around the wearer. But this proximity of camera and scene itself introduces a challenge: small motions introduce relatively large changes of viewpoint and occlusion events increase, the more so as hands and objects move into the workspace. We discuss a method to finesse this later.

Full details of the single EKF used here are given in [7], but the state prediction, measurement process and update regime need little comment. The state vector comprises two parts, $\boldsymbol{X}_i$, the fixed 3D locations of map features, and $\boldsymbol{c}_t = (\boldsymbol{t}, \boldsymbol{q}, \boldsymbol{v}, \boldsymbol{\omega})$, the time dependent camera position, orientation, translational velocity and angular velocity, all defined in a fixed world frame. Orientation is represented by a quaternion $\boldsymbol{q}$ whose redundancy is handled by re-normalisation at each step to ensure that $|\boldsymbol{q}| = 1$ and by a corresponding Jacobian calculation affecting the covariance matrix. The update model is one of constant translational and angular velocities, assuming that in each time step, white, Gaussian distributed impulses in velocity and angular velocity of $\boldsymbol{n} = \Delta t\,(\boldsymbol{a}, \boldsymbol{\alpha})$ are applied to the camera. These components are assumed uncorrelated so that the covariance matrix $\mathtt{P}_n$ of the noise vector $\boldsymbol{n}$ is diagonal.

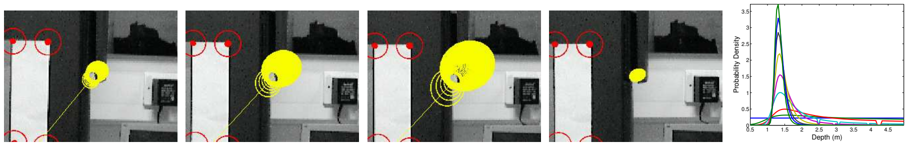
Putative new features for addition to the 3D map are detected initially with the Shi-Tomasi saliency operator [19], and features that are eventually added to the map are stored with a 15×15 pixel appearance template, as in [9]. Given a predicted camera position, each feature $\boldsymbol{X}_i$ is projected into the new image along with its associated uncertainty region derived at the 3σ limit from the innovation covariance matrix $\mathtt{S}_i$, and searches made within the new image for correspondence using normalised sum-of-squared difference correlation. In a fixed-time system, it pays to search first for the feature with highest $\mathtt{S}_i$ as this give greatest scope for reducing the uncertainty in the map as a whole.

4.2 Feature Initialisation Using Factored Sampling

A special aspect of single camera SLAM is that when a salient feature is first found it is known only to lie along a semi-infinite 3D line, with wholly uncertain depth. It cannot be simply inserted into the 3D map with a conventional covariance.

One might track such features in 2D over a number of frames and insert them once the depth is better known, but this is wasteful of available information. Instead we use factored sampling. The feature is initialised in the map as a 3D line along which 100 particles are uniformly distributed over a suitable depth range (here, for indoor operation, 0.5m to 5.0m) to represent depth hypotheses, or equivalently the depth pdf. At subsequent time steps, every hypothesis is tested by projection into the image as an elliptical search region, and a likelihood evaluated that the particular depth is correct, and the particles redistributed using a Bayesian re-weighting. Once the pdf can be well represented by a Gaussian, the feature is changed from line to 3D point with the appropriate Gaussian uncertainty, as shown in Figure 7.

This simple depth prior removes the need to search along the entire epipolar line, and improves the robustness and speed of initialisation. In the frame-rate implementation, the speed of peaking of the particle distribution is increased (and correlation search decreased) by deterministic pruning of the weakest particles at each step. (Note that to avoid an unreasonable computational cost, the parameters of the line are not updated while the particle depth distribution is being refined, except via indirect coupling to the robot state in the Kalman Filter. The feature is not used to update the camera position estimate.)



0.033s 0.066s 0.100s 0.133s Pdf evolution

Fig. 7. A representation in the image of the distribution of particles as the depth pdf is refined. Initially the 3D line projects as the epipolar line, but after very few frames the particles are sufficiently clustered for the line to be replaced by a point. The graph shows the pdf growing out from a flat line to a peaked distribution over a few frame periods.

4.3 Map Management

The map management criteria aim to keep the number of features observable in any one view to the range 6-10, a number compatible with both the desired localization accuracy and maintaining a 30Hz frame-rate on a 1.6GHz Pentium Centrino. A feature is *predicted* to be observable from some particular viewpoint if its projection into a model of the camera lies well within the image, and if both angular and range differences between the viewpoint and the feature's initial viewpoint are not so great as to make the image template valueless. Features are added if the number of those actually observable falls to 5 or less, and a feature is deleted if its long-term ratio of actual observability after search to predicted observability falls below 1:2. Our experience is that these policies allow for the necessary degree of ephemeral clutter in the scene without introducing mismatches (which are of course not acceptable in SLAM).

5 Active control of the wearable

The SLAM procedure just described is applicable to any single camera – a hand-held camera, a static wearable camera, robot-mounted or otherwise [7]. But for an *active* wearable camera, knowledge of its own position within that 3D scene opens up two very promising actions: (i) saccades, or controlled redirections of gaze between known 3D locations, even when the target location is out of the current field of view; and (ii) extended periods of fixation on a 3D location during wearer motion.

Given the 3D positions of the camera and some point in the scene it is a straightforward matter to compute the overall rotation and, via the inverse kinematics, the joint rotations required to fixate it. (The joint angle redundancy is removed by requiring the cyclotorsion axis to maintain the vertical in the image.) Although the motion update model can handle sudden controlled motions, we prefer to limit such demands so that the maximum angular velocity is $\sim 30^\circ\mathrm{s}^{-1}$. This allows the controlled motion to be absorbed by the motion model, thus avoiding the need to feed back detailed odometry from the wearable, and preventing image blur which would in turn require regular vision processing to be suspended. It is worth stressing that regular accumulation of data into the EKF continues uninterrupted.

Figure 8 shows stills clipped from an extended sequence taken through the wearable camera as the wearer moves in front of a workbench scene. The SLAM process determines 3D location and structure, and a remote operator determines when to redirect the wearable camera's gaze and when to use it to fixate on an object

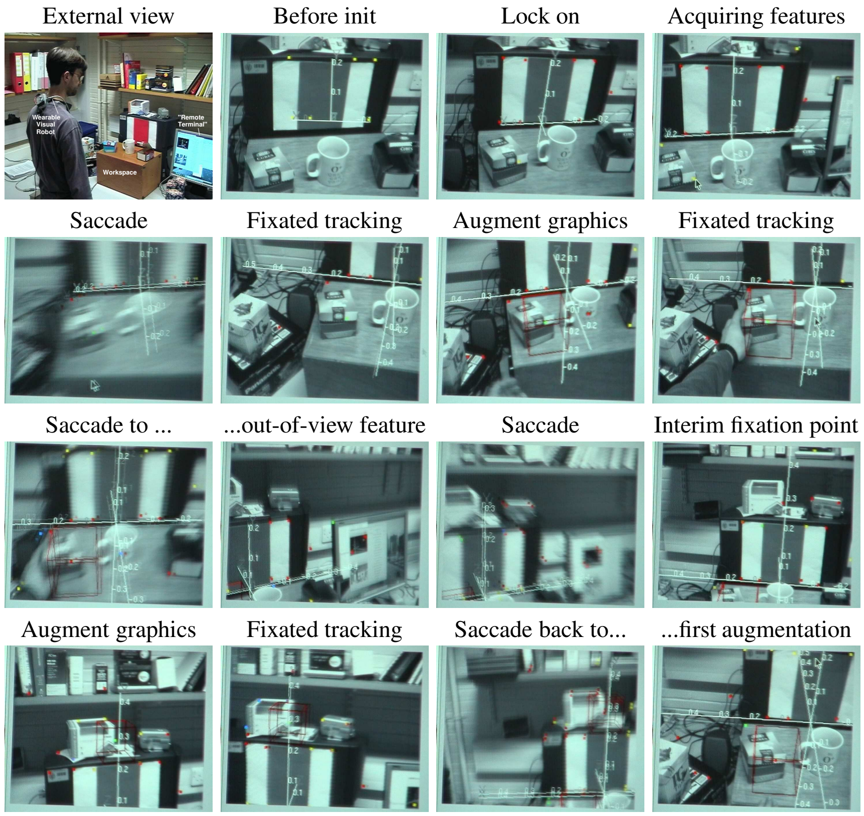


Fig. 8. The experimental setup, followed by the camera's view of initialisation and tracking of features and saccades between them during continuous wearer motion. (The blurring in the saccade frames is an artefact of the video making process.) The video is at www.robots.ox.ac.uk/ActiveVision/Movies/wearableslam.mpg

by clicking on one of two graphical displays. The first shows a three-dimensional reconstruction of the estimated locations of the camera and features, and the other shows the view from the camera augmented with a world coordinate frame and other graphics, including virtual objects. Interestingly, operators find it difficult to click reliably on objects in the rapidly-moving camera view — the stable 3D view in the world coordinate frame proves valuable for this.

6 Conclusions

Wearable sensing appears a fertile area in which to develop active visual processes, perhaps particularly because the presence of a wearer and, if teleoperated, the remote operator, provide a degree of high level decision making that is at present unattainable in autonomous systems. This paper has reviewed the design and construction of an active wearable camera, and has outlined progress in equipping it with reactive and purposive visual processing. The frame-rate visual localisation and mapping system, though relevant for any single camera undergoing general motion, constitutes a particularly significant step towards useful wearable vision.

Perhaps the main challenge that arises here is that using a single camera over a restricted viewing solid angle leads to larger errors in depth. We certainly need more efficient processing and better matching to increase the number of features maintainable at frame rate, but also wish to explore how to make wearable and wearer enter a co-operative dialogue to acquire additional features from a wider viewing cone.

References

1. H. Aoki, B. Schiele, and A. Pentland. Realtime personal positioning system for a wearable computers. In *Proc 3rd IEEE Int Symp on Wearable Computing, San Francisco CA, Oct 18-19, 1999*, pages 37–43.
2. J.-Y. Bouget and P. Perona. Visual navigation using a single camera. In *Proc 5th Int Conf on Computer Vision, Boston MA, June 20-23, 1995*, pages 645–652. IEEE Computer Society Press, Los Alamitos, CA, 1995.
3. A. Chiuso, P. Favaro, H. Jin, and S. Soatto. "MFm": 3-D motion from 2-D motion causally integrated over time. In *Proc 6th European Conf on Computer Vision, Dublin, June/July 2000*, pages II:734–750. LNCS vol 1843, Springer, 2000.
4. B. Clarkson and A. Pentland. Unsupervised clustering of ambulatory audio and video. In *Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, Phoenix AZ, 1999*, volume VI, pages 3037–3040.
5. D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.
6. A. J. Davison. *Mobile Robot Navigation Using Active Vision*. DPhil thesis, University of Oxford, 1998. At www.robots.ox.ac.uk/˜ajd
7. A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc 9th Int Conf on Computer Vision, Nice, Oct 13-16, 2003*.
8. A. J. Davison and D. W. Murray. Mobile robot localisation using active vision. In *Proc 5th European Conf on Computer Vision, Freiburg, June 2-6, 1998*, pages II:809–825. LNCS vol 1407, Springer, 1998.
9. A. J. Davison and D. W. Murray. Simultaneous localization and map-building using active vision. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 24(7):865–880, 2002.
10. H. F. Durrant-Whyte, M. W. M. G. Dissanayake, and P. W. Gibbens. Toward deployments of large scale simultaneous localisation and map building (SLAM) systems. In *Proc 9th Int Symp on Robotics Research, Snowbird, Utah, 1999*, pages 121–127. (Robotics Research: the Ninth International Symposium. Eds Hollerbach and Koditschek, Springer,2000)
11. E. Foxlin. Generalized architecture for simultaneous localization, auto-calibration and map-building. In *Proc IEEE/RSJ Conf on Intelligent Robots and Systems, Lausanne, Switzerland, October 2-4, 2002*, pages 527–533.
12. J. G. H. Knight, A. J. Davison, and I. D. Reid. Towards constant time SLAM using postponement. In *Proc IEEE/RSJ Conf on Intelligent Robots and Systems, Maui, HI, Dec 2001*, pages I:406–412. IEEE Computer Society Press, 2001.
13. N. Kohtake, J. Rekimoto, and Y. Anzai. InfoStick: an interaction device for inter-appliance computing. In H. Gellersen, editor, *Proc Workshop on Handheld and Ubiquitous Computing, Karlsruhe, Sep 27-29, 1999*, pages 246–258. LNCS vol 1707, Springer, 1999.
14. M. Kourogi, T. Kurata, and K. Sakaue. A panorama-based method of personal positioning and orientation and its real-time applications for wearable computers. In *Proc 5th IEEE Int Symp on Wearable Computing, Oct 2001*, pages 107–114.
15. J. J. Leonard and H. J. S. Feder. A computationally efficient method for large-scale concurrent mapping and localization. In *Proc 9th Int Symp on Robotics Research, Snowbird, Utah, 1999*,pages 169–176. (Robotics Research: the Ninth International Symposium. Eds Hollerbach and Koditschek, Springer, 2000)
16. S. Mann. WearCam (the wearable camera): Personal imaging systems for long-term use in wearable tetherless computer-mediated reality and personal photo/videographic memory prosthesis. In *Proc 2nd IEEE Int Symp on Wearable Computing, Pittsburgh PA, Oct 19-20, 1998*, pages 124–131.
17. W. W. Mayol, B. J. Tordoff, and D. W. Murray. Designing a miniature wearable visual robot. In *Proc Int Conf on Robotics and Automation, Washington DC, May 11-15, 2002*, pages 3725–3730.
18. W.W. Mayol, B. Tordoff, and D.W. Murray. Wearable visual robots. In *Proc 4th IEEE Int Symp on Wearable Computing, Atlanta GA, Oct 16-17, 2000*, pages 95–102.
19. J. Shi and C. Tomasi. Good features to track. In *Proc IEEE Conf on Computer Vision and Pattern Recognition, Seattle WA, June 21-23, 1994*, pages 593–600.
20. T. Starner, J. Weaver, and A. Pentland. Real-time American Sign Language recognition using desk and wearable computer based video. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, 1998.
21. S. Thrun, D. Fox, and W. Burgard. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine Learning*, 31:29–53, 1998.
22. B.J. Tordoff, W.W. Mayol, T.E. de Campos, and D.W. Murray. Head pose estimation for wearable robot control. In *Proc British Machine Vision Conference, Cardiff, Sept 2-5, 2002*, pages 807–816.
23. A. Vardy, J. Robinson, and L. Cheng. The WristCam as input device. In *Proc 3rd IEEE Int Symp on Wearable Computing, San Francisco CA, Oct 18-19, 1999*, pages 199–202.