On the Choice and Placement of Wearable Vision Sensors

Walterio W. Mayol-Cuevas, Member, IEEE, Ben J. Tordoff, and David W. Murray, Member, IEEE

Abstract—This paper discusses two of the most important design considerations for a wearable device with visual sensing: what kind of sensor to use and where to place it. While nature and computer vision have explored a wide range of imaging techniques, wearables have mostly viewed the world through conventional narrow-view passive cameras designed for nonwearable applications, which are attached to the wearer’s head. The rationale presented here for sensor selection and the novel methodology developed for objectively studying sensor placement have informed the development of a number of visual wearables.

Index Terms—Sensor placement, wearable computing, wearable vision.

I. INTRODUCTION

Most advanced animals are endowed with some degree of visual faculty, assisting in the key tasks of predicting temporal events, the recovery of surrounding scene structure, navigation through that structure, and the categorization and recognition of objects. In response to particular environments and creatures’ roles within them, a remarkable variety of ways of forming images and of locating sensors have evolved [14]. Our own visual system is, of course, part of this spectrum, but if offered an extra wearable visual sensor, what form should it take and where should it be placed? Should, for example, our extra eye be omnidirectional and placed near the ones we already have? What opportunities are gained and lost by doing so?

Optimal choices for a sensor’s type and location certainly depend on the visual tasks envisaged, but in the domain of wearables, this has often been the only criterion considered and, even then, outside of any formal design process. Indeed, much literature on visual wearables describes the attachment of standard video cameras to the wearer’s clothing. Although always satisfactory for recovering ambient information such as lighting levels and color histograms, for more specific tasks, the resulting imagery may be too dependent on the wearer’s exact posture and motion. Often, a more considered approach is to attach cameras to particular body parts. In [2] and [15], for example, a hat-mounted camera looks outward to where the wearer’s face points, whereas in [3], the hat-mounted camera looks downward at the wearer’s hands, and in [4], the camera points upward toward the ceiling. In [16] and [11], cameras are strapped to the wearer’s hands themselves, and in [10], cameras are strapped to shoes. Now, the captured views are highly specialized but, consequently, more susceptible to the vagaries of the wearer’s changing stance.

Fig. 1 summarizes positions for the placement of wearable 49 cameras reported in the literature. While intuition, experience, and, in some cases, comparative field trials of alternatives might have led to the correct design choices being made, we suggest that an alternative during prototype development is to use a computer-based simulation with models of both the wearer and the device. Quantitative assessments provide a common ground on which to make objective comparisons to inform and shorten the development cycle.

This is the approach taken here, and two areas have been chosen as particularly relevant and amenable to quantitative evaluation. The first is sensor field of view (FOV), where an assessment is made of how much of the world the sensor can see from a given position or, equivalently, how vulnerable is its view to occlusion by the wearer’s body. The second is the level of positional disturbance suffered by the sensor as the wearer moves. Although some motions of the wearer’s body may be helpful in redirecting the gross viewing orientation (for example, with the hat-mounted camera), other movements caused by locomotion, or even breathing, are nothing but noise during visual tasks.

In Section III, a simulation is devised, which allows cameras to be placed at arbitrary locations on an articulated 3-D model of the human body. Figures of merit can then be calculated for the criteria in the areas mentioned earlier, and after this first quantitative evaluation is done, further design considerations are introduced. The description is centered on cameras, but the methodology is equally applicable to other optical devices. It constitutes a novel approach to the issue of device placement for wearables.

Section IV details experiments which examine suitability in the areas of absolute FOV, FOV of the handling space, and body motion. The vexing issue of how to combine those results is addressed in Section V, and their application to the design of a particular wearable is outlined in Section VI. In Section VII, the results are discussed and conclusions are drawn. First, however, broader issues in the selection and sighting of wearable vision devices are introduced.

Manuscript received September 11, 2006; revised November 28, 2007 and April 11, 2008. This work was supported by the U.K. Engineering and Physical Science Research Council under Grants GR/N03266 and GR/S97774. The work of W. W. Mayol-Cuevas was supported by a Mexican CONACYT scholarship. This paper was recommended by Associate Editor C. M. Lewis.

Digital Object Identifier 10.1109/TSMCA.2008.2010848
II. ISSUES IN WEARABLE CAMERA SELECTION

As just mentioned, nature has found a number of ways of realizing the mapping of spatially distinct parts of the scene onto spatially distinct and addressable parts of visual sensors. The flatworm distributes its photoreceptors around its otherwise opaque body, whereas the marine nautilus requires incoming light to pass through a pinhole before hitting its sensor array. Like the nautilus, shrimps avoid the need for lenses, but cleverly, by using mirrored frustums of pyramids to focus light. Insects have evolved compound lens eyes, with each lens-piece and receptor oriented in a fixed direction relative to the body, the resulting wide angle of view being highly suited to ego-motion recovery and collision avoidance. In more sophisticated creatures, the need to compromise between increasing resolution with high angular spread and low sensor volume has led to the evolution of the moving eye, with a high-resolution, but narrow-angle, central fovea complemented by a surrounding lower resolution and wider angle periphery. Moving eyes also provide resilience against image blur during rapid head motion, an enhanced ability to discern between foreground and background motions, and a facility to estimate heading direction by actively preventing contamination of the translational flow field. The example of the body, the resulting wide angle of view being highly suited to ego-motion recovery and collision avoidance.

By contrast, in wearable computing, it is noteworthy that the variety of visual sensors used to date has been much smaller, with the great majority of researchers using standard video cameras with a modest FOV (see Table I). There are a few exceptions. Clarkson and Pentland [6] used a wide-angle camera mounted on the shoulder, but this was only to measure lighting levels in coarsely divided parts of the scene, and so, a small and low-quality device sufficed. Rungsarryotin and Starner [7] used a catadioptric omnidirectional camera, but its bulk requires it to be “worn” on top of a pole attached to the wearer’s back and sticking out above the head.

Table I: Visual Sensors and Examples of Their Use in the Literature

<table>
<thead>
<tr>
<th>Wearable Sensor Type</th>
<th>Used By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive Narrow View Camera</td>
<td>Mann [1], Starner et al. [5]</td>
</tr>
<tr>
<td>Passive Wide Angle</td>
<td>Clarkson and Pentland [6], Foxlin [4]</td>
</tr>
<tr>
<td>Catadioptric (mounted on pole at the back)</td>
<td>Rungsarryotin and Starner [7]</td>
</tr>
<tr>
<td>Active Narrow View Camera</td>
<td>Mayol et al. [12]</td>
</tr>
<tr>
<td></td>
<td>Kurata et al. [13]</td>
</tr>
</tbody>
</table>

Taking a little license, the range of imaging devices found in engineered vision systems is similarly varied. At the level of the flatworm, one might pick out daylight detectors placed on top of buildings and streetlamps. Static arrays of cameras used, for example, by Aloimonos et al. [19], [20], have some resonance with the compound insect eyes. Schenato et al. [21] have applied biomimetic, but much simplified, compound eyes for the control of micromechanical flying insects. Omnidirectional cameras using fish-eye lenses have had frequent application in navigation [22], [23], and omnidirectional imaging using reflection has been much advanced by Nayar and Peri [24] using catadioptric cameras. Imagers with graduated log-polar resolution were designed and fabricated by Sandini et al. [25], although their use has not been widespread. Visual and control aspects of the active camera were much researched in the 1990s [26]–[30].

While most workers have used static wearable cameras, some have opted for an active vision approach by developing miniature directable cameras that actively reorient themselves depending on the task and offer a level of decoupling between wearer motion and sensor [9], [12], [31].

To give weight to the argument for an active camera, consider the physical requirements of a passive wide-angle (hemispherical) sensor with an angular resolution comparable to that of the human fovea, about \( \frac{1.6 \times 10^{-2} \, \text{c}}{} \). To cover the visible spectrum, \( 0.3 \leq \lambda \leq 0.8 \, \mu\text{m} \), the minimum receptor size is 146 \( \mu\text{m} \) on the order of 1 \( \mu\text{m} \times 1 \, \mu\text{m} \), although to prevent charge leakage to neighboring photoreceptors, the linear dimension might be doubled. Packing 2-\( \mu\text{m} \times 2-\mu\text{m} \) receptors to 149 give the required angular resolution yields a sensor area of 150.
22.5 mm × 22.5 mm, with a total of some 1.3 × 10^8 pixels. This hypothetical sensor has an acceptably small size for wearable applications, although issues of wiring, power consumption, and lens distortion are not easily dismissed.

Further problems with a nonmoving sensor start to appear when sampling times and scene on sensor motion are taken into consideration. If Δφ is the receptor’s sensing angle and Δt is the receptor’s response time, then

\[ \Omega_{\text{res}} = \frac{\Delta \phi}{\Delta t} \]

is the maximum angular speed that can be tolerated before blurring occurs. For a conventional 640 × 480-pixel video camera with a refresh rate of 30 Hz, FOV of 42° horizontal, and maximum light-gathering time of 33 ms, (1) indicates that detail begins to be lost at relative angular velocities of just 2° s⁻¹. Note that a person walking past 2 m away from the camera corresponds to an angular velocity of about 40° s⁻¹. The situation is worse, of course, for the hypothetical wide-angle high-resolution sensor. With an integration time of 33 ms, it would start to lose detail at angular velocities of less than 0.5° s⁻¹. To avoid blur, one could integrate for a shorter time. In the example of the walking person, the video camera above would need to collect light in Δt = (1/800) s, and the hypothetical sensor in just Δt = (1/3000) s. Current camera electronics can certainly deliver such times, but shorter integration times require either a better light environment or a more sensitive receptor, achievable perhaps by increasing its area.

We conclude that only poorly resolving systems (large Δφ) can afford being passive, the more so as motions of 40° s⁻¹ and beyond are easily removable by mechanical rotation of the sensor. Moreover, the only way that the hypothetical passive sensor can have omnidirectional zoom is by increasing sampling resolution. Since we have already considered the physical limits of receptor’s size, to achieve a required magnification, the 22.5-mm × 22.5-mm sensor has to be scaled. A modest 10× zoom would produce an area of 225 mm × 225 mm and some 1.3 × 10^9 pixels. Another reason to avoid wide-angle passive sensors is that the optics of getting light into the camera, whether by lens or mirror, are still too bulky. We argue that the obvious step in advancing wearable vision sensors is to introduce controllable active cameras.

The choice of a directable sensor is not based entirely on the balance of volume with resolution and on optical theory. Just as Land notes the visual information processing benefits in animals, so too have active vision researchers pointed out the benefits of active sensing to visual processing [26], [32]–[34]. These transfer without modification into the wearable domain, as can be seen by making explicit the set of frames a wearable camera must be able to reference. There are three frames: 1) frames attached to the wearer’s body; 2) frames attached to the static world; and 3) frames relating independent objects with the wearer. These sets of frames involve fundamentally different kinds of actions by a sensor. The first set is involved in tasks such as viewing the manipulative space in front of the wearer’s chest or viewing the wearer’s head. Here, there is less requirement for camera movement and less requirement for a wide or omnidirectional view. Occlusion of the camera’s view by the wearer is less likely to occur. The second set is involved in tasks like maintaining viewing direction or looking in “global” directions like “up” and “down” or “along the street.” The need to see all around the wearer becomes more important, and the ability to decouple from the wearer’s motion is useful. The last set is the most demanding of all. Both reducing occlusion and being independent from the wearer’s underlying motion are important when attempting to fixate on a particular part of the surroundings or to track an independently moving object.

Two points begin to emerge from this discussion. First, an active camera with an ability to select motion behavior or control modes can satisfy the needs in these different frame sets [12], [31]. Second, key criteria in the objective evaluation of a wearable camera are its FOV and its resilience to positional disturbance introduced by the wearer’s own motion. Both of these are functions of the positioning of the camera on the body, and exploring these dependences is the purpose of the remainder of this paper.

III. MODELING FOV AND MOTION DEPENDENCE

A. Simulation Model

A female model from the Humanoid Animation Working Group [35] is used as the basis of the comparison for the functionality of camera positions around the human body. It is a facet model, coded in Matlab, consisting of about 1000 2D points connected into 1800 triangular facets, and arranged into 250 16 body segments, each of which can be independently rotated to simulate different poses. Its scaled height of about 1.70 m is typical of Caucasians [36], [37]. The model is one example of the shapes that can be considered, but the approach and methodology is applicable to other shapes and sizes.

The simulated optical device can be positioned arbitrarily in 2D space around the body but, for convenience in tests, is placed some distance d above the centroid of each of the model’s facets in turn, as shown in Fig. 2(a) and (b). Given a patch defined by its vertex positions P₁, P₂, and P₃ in anticlockwise order, the camera is placed at

\[ C = P₀ + d\hat{N}, \quad d > 0 \]

where P₀ is the centroid \( P₀ = (P₁ + P₂ + P₃)/3 \) and the outward unit normal vector is

\[ \hat{N} = \frac{(P₂ - P₁) \times (P₃ - P₁)}{|(P₂ - P₁) \times (P₃ - P₁)|} \]

B. Determining Occlusion

To determine body self-occlusion, and for subsequent visualization, the unit radius view sphere is constructed around the camera, and its center is set to the origin. The vertices of any triangle [Fig. 2(b)] are projected onto the sphere as

\[ S_i = \frac{T_i - C}{|T_i - C|} \]

The collection of body segments and polygons was modified and used with permission from an original VRML file by C. Ballreich copyright 1997 3Name3D.
Fig. 2. (a) Device is positioned above the centroid of a model facet, at a small distance (here, 37.5 mm) along the surface normal. (b) Vertices $T_i$ are mapped to $S_i$ to project the model into a unit sphere with origin at $C$ situated a distance $d$ above another triangular facet. (c) View sphere is derived from the camera position in (a).

249 and the vectors

$$E_{12} = S_1 \times S_2 \quad E_{23} = S_2 \times S_3 \quad E_{31} = S_3 \times S_1$$

are constructed. A ray $R$ is occluded by that triangle if

$$R \cdot E_{12} \geq 0 \quad R \cdot E_{23} \geq 0 \quad R \cdot E_{31} \geq 0.$$

A subdivided icosahedron is used to generate sample rays $R$. Because the icosahedron is the highest order platonic solid, it is impossible evenly to space more than 20 points around a sphere, but repeated resection of the icosahedron’s faces into equilateral triangles gives a good approximation to uniformity. Fig. 2(c) shows a result derived from the camera position given in Fig. 2(a) and obtained by projecting all body facets into the camera’s view sphere. The head and shoulder are visible, but the lower body is obscured. The percentage of the $4\pi$ solid angle occluded is about 23%. Fig. 3 shows further examples of mapping for some camera placements used in the experiments for Figs. 7 and 12.

IV. EXPERIMENTS AND RESULTS

Experiments have been performed to assess how the absolute FOV, the view of the manipulation space, and the degree of motion change as the camera is moved to different positions around the body.

Fig. 3. Further examples of placements of an optical device on the head and chest with their corresponding projections onto the unit sphere centered on the device (small circle). Placements are (a) temple, (b) forehead, and (c) chest; in each case, the center of projection is at 37.5 mm from body surface.

As noted earlier, the facet model is articulated into 16 segments, and it is possible to change its pose by rotating the respective body sections. Although any representation of rotation could be adopted, here, we use quaternions of the form $\hat{q} = [q^\top, q_0]^\top$, with the conjugate denoted as $\bar{q} = [-q^\top, q_0]^\top$. In the case of the articulated model, to rotate a point $P$ from a body section around a joint with center $J$ using quaternion $\hat{q}$, one derives the pure quaternion $[(P - J)^\top, 0^\top \hat{q}]$, then rotates it before adding the vector part to $J$

$$[r, 0] = \bar{q} [\bar{q}^\top, 0] [\bar{q} \cdot P - J]$$

$$P' = r + J.$$
Fig. 4. Kinematic chain for the polyhedral human model. Joints are labeled in lowercase, and body parts in uppercase. The arrows indicate the chain direction from the root.

A. Absolute FOV

For the first of these tests, the model was set in a straight, so-called anatomical configuration, and the camera was placed first at a distance of 12.5 mm and then at 37.5 mm above each facet in turn. The absolute FOV was measured for each position by casting rays in directions nearly uniformly spaced around the sphere and deriving the fraction of those occluded. For each camera position, 272 rays were cast, corresponding to an average angular separation of 12.5°. Note that it was assumed that the body was in free space, i.e., that there was no other object adjacent to it and no floor beneath it. Measuring body occlusion alone is relevant, since, for example, a foot-mounted camera can be useful for detecting some level of wearer’s activity and context through observing the floor, an example of which is described in [10].

Fig. 5(a) and (b) shows the frontal projections of the model for 12.5 and 37.5 mm, respectively. The shade of each facet indicates the FOV, with light areas having a narrow FOV, and vice versa. As one would expect, the areas under the arms and between the legs are heavily occluded, and highly convex sections such as the head, shoulders, and even the feet score well. In fact, as the human model is quite thin, any facet on the periphery of the coronal anatomical plane has a wide view, including the outside edges of the hands, arms, legs, and shoulders.

Fig. 5(c) redisplays the results for $d = 37.5$ mm, but with the scores quantized to highlight the best and worst areas, and using a more informative viewpoint. In particular, the head, the position of choice both for nature and for many researchers, is favored, as are the shoulders. This comes as no surprise; however, it is important to note which alternative placements are also favorable if these locations cannot be used for reasons related to general social acceptability or the specific application.

The chosen posture in these experiments was the standard anatomical position, but the analysis can, of course, be made for any other posture once a specific application is under consideration. An example of a different posture and its map of absolute FOV is shown in Fig. 6.

To examine how the degree of occlusion changes with distance $d$, some of the favored positions from the previous results were selected and the sensor distance was varied from essentially resting on the surface, $d = 5$ mm, too far above it, $d = 150$ mm. The results are shown in Fig. 7. At all of the chosen positions, the occlusion reduces as the height is
Fig. 6. Alternative posture (Rodin’s thinker) and its associated map of absolute FOV displayed on the standing figure. The camera was at $d = 37.5$ mm above the surface of each facet.

Fig. 7. Fraction of occlusion for five favored positions plotted as a function of distance $d$ above the body surface. The rates at which the occlusion decreases (or absolute FOV increases) are highest for the three head and shoulder locations but fall to a steady value for $d > 50$ mm (see Figs. 2(a) and 3 for the actual camera placements).

increased, but the most significant reductions are made for the two selected head positions and the shoulder position. In these, the rate of change of decrease has fallen to a constant value by around $d = 50$ mm.

B. View of the Handling Space

Of particular importance in a number of applications is the visibility of the wearer’s hands and the objects that they are manipulating. The area immediately in front of the chest is the region in which the majority of manipulation occurs. Biomechanical studies, such as those of Marras [38], describe volumes for “precision,” “medium,” and “heavy” work, and these have been merged here into a volume called the handling space. The handling space has a volume of $500 \times 500 \times 250$ mm$^3$ located some 300 mm in front of the chest and at about 775 mm from the floor. Note that this space is not the envelope of all the possible places where manipulation can occur.

To compute the visibility of the handling space, its volume is evenly filled with points and a ray is cast from each of them to the viewpoint sphere center to measure the proportion of the volume visible from the camera. As before, the camera is placed at every position around the body, and the results are used to shade the facets in Fig. 8. The handling space is indicated by the dashed box.

C. Resilience to Wearer Motion

If the interest is to image a particular section of the wearer’s body, the ideal setup to null body motion is for the camera to be worn on the same section. If this placement is not possible, the next best candidate to reduce relative camera motion is likely to be another body part that is connected to the one of interest by the simplest (lowest degree of freedom) joint. However, when the interest is in making measurements outside the set of user frames, such as when viewing an object in the scene, the amount of user motion that must be compensated for becomes more significant, and its degree and character will vary considerably, depending on what that part is and what the wearer is doing.

By using the articulated model, it is possible to study any motion for which the evolution of joint angles over time is known. Here, walking is used as an example. Fig. 9(a) shows a half period’s worth of joint angle data applied to the model.
Fig. 9. (a) Key frames from a half-period of walking motion and (b) the corresponding motion map. Darker shading indicates less motion.

365 (these particular data were derived from a chronophotographic 366 sequence by Eadweard Muybridge in the late 1800s [39]).
367 The relative figure of merit for motion is evaluated as the 368 length of the path described by each facet’s centroid as it 369 moves through space over one period, normalized by the overall 370 translation. The results for the sequence are shown as intensities 371 in Fig. 9(b).
372 In the walking example, the motion of much of the upper 373 body and head is similar, but only because the wearer’s attention 374 is fixed forward. More typically, the wearer’s attention would be 375 drawn by surrounding objects or events, and a camera mounted 376 on the head would certainly experience a greater degree of mo- 377 tion, with typical positional disturbances of ±30° at velocities 378 of 90° s^{-1} and accelerations up to 600° s^{-2} [18]. Any device 379 mounted on the head that needs to maintain its own direction 380 of sensing, independent from that of the wearer, would need 381 either to be agile or to be able to handle the large image errors 382 that would occur as the wearer’s head moved.

383 D. Multicamera Arrangements

384 An extension of the monocular case involves the computation 385 of the joint FOV of two or more cameras worn at different 386 body locations. This joint FOV is computed as the union of 387 the individually visible regions. Choosing a location that has 388 a larger choice of where a second camera can be placed could 389 be valuable when designing a visual vest or shield.
390 Fig. 10 shows example configurations for pairs of cameras 391 with joint full unobstructed spherical views, calculated for 392 the cameras at 37.5 mm above the body surface, where the 393 “master” camera is at one of four fixed positions on the head 394 and shoulder. Fig. 11 shows the results of a simulation where 395 the master camera was placed at each facet of the head and torso 396 in turn. The shade of each facet is proportional to the cumulative 397 area of other facets where the second camera can be placed to 398 produce a joint FOV with 0% occlusion. The darker the facet, 399 the larger the associated area.

400 These simulations ignored potential occlusion by the com- 401panion camera and were also restricted by placing the second 402 camera within the same body part as the master (i.e., if the 403 master camera is placed on the torso, a companion is also 404searched somewhere else on the torso). The angular resolution 405 for rays on the sphere was, as earlier, 12.5°, and each cast ray 406was assumed to be from a point at infinity. Finally, the posture 407analyzed was the anatomical posture.

Fig. 10. (Cross) “Master” camera and (circles) potential couples that complement the FOV. Master cameras at (a) the forehead, (b) temple, (c) shoulder close to neck, and (d) shoulder close to arm. All joint FOVs have 100% visibility, except in case (c) which has the best companion (around the lowest rib) with a joint occlusion of 1%. Moving the sensor toward the arm, as in case (d), generates further areas for companion cameras with joint 100% of visibility.

An interesting outcome for the torso is that the locations that 408 score well for a single camera—i.e., along the shoulder—have 409 the worst performance in terms of finding a single second 410 location to complement the FOV. Fig. 11(c) and (d) shows 411 images that are almost the negatives of the torso map shown 412 in Fig. 5(c). This is a result both of occlusion by the neck and 413 head and the lack of location from which a camera can view 414 downward. In contrast, a camera placed on the chest, which 415 does not score well as a single camera, has its FOV more 416 easily complemented by one placed anywhere on the back. The 417 outcome for head locations differs again, a result of the head’s 418 more spherical shape and its smaller radius of curvature. Here, 419 the locations that score highly for the single camera also score 420 well as locations for which the FOV can be complemented by a 421 second camera on the head.

Although the rays were here assumed to be cast from points 423 at infinity, one could easily adapt the method to compute, for 424
Fig. 11. Comparing the value of a facet as a site for a master camera. The darker the shade of a facet, the larger the cumulative area elsewhere on that body part where a second camera can be placed to generate a joint FOV of 100%.

Example, optimum placements for cameras to have stereo views of a given region in space.

V. Fusing the Criteria

Having computed quantitative figures of merit over every candidate placement for a number of criteria—here for the absolute FOV, view of the handling space region, and resilience to body motion—it would be useful to fuse these together into a single quality map. Obviously, hard constraints on the acceptability of any single quantity can be used immediately to rule out a position, but, in the absence of further evidence indicating relative importance, one is reduced merely to exploring possible combinations. Here, a simple thresholded linear combination is considered

\[ m(k) = \sum_i g_i m_i(k) \]  

where \( m_i(k) \) is the normalized value of the \( i \)th criterion at facet \( k \) and \( g_i \) is its corresponding gain.

Fig. 12 shows the sensitivity surfaces generated by every candidate placement as the gains \( g_{FOV} \) and \( g_{motion} \) for absolute FOV and motion resilience are varied, taking the value of \( g_{handling} \) to be unity (for this analysis, only a single camera, anatomical posture, and walking motion are considered). What is apparent is that if one places relatively little emphasis on the absolute FOV, the chest area will win. However, if emphasis is placed on the absolute FOV, then the head, temple, and shoulder areas are favored, no matter what emphasis is placed on resilience to motion.

Fig. 13 shows this latter situation, for the particular gains of \( g_{FOV} = 50, g_{motion} = 30, \) and \( g_{handling} = 1 \), but mapped onto the body model. The favored areas are the head along with the 452 shoulders.

Fig. 14. Wearable visual robot. (1) Elevation axis, (2) pan axis, (3) cyclostoration axis, and (4) CCD and optics.
Fig. 15. Rows: First, middle, and last frames of a sequence as the wearer turns swiftly to his right at about 60° s\(^{-1}\) under two modes of shutter and stabilization. (a) If the camera is passive and the shutter is the slowest, detail is lost and the disparity is significant. (b) Faster shutter keeps detail, but disparity obviously remains high. (c) Detail is better than in (a) and disparity controlled. (d) Both detail and disparity are controlled (although at the expense of increased pixel noise: The 15-ms shots had their brightness compensated with increased gain to provide a similar brightness to the other tests).

Although our aim in this paper is to discuss a general methodology, one can readily see that if use of the head is vetoed—either because of its over strong coupling to the wearer’s attention or because of its poor social acceptability, or, returning to the discussion of Section II, because high angular velocities and accelerations can easily break visual tracking—then the shoulders become the most favored region.

VI. WEARABLE VISUAL ROBOT

The sensor selection and placement methodology discussed in this paper have been applied to the development of the shoulder-mounted wearable active camera shown in Fig. 14. It uses three servomotors with 160° of motion range to provide orientation around elevation, pan, and cyclotorsion axes; it has a 640- × 480-pixel noninterlaced image sensor with 42° of FOV; and it has a control interface connected to a laptop computer. An inertial sensor package is used to recover 3-D bearings to stabilize images based on wearer’s orientation, while visual tracking of objects allows tracking during translation. The device is the front end of a wearable computer able to sense the wearer’s environment and activities and has been applied (through several design iterations) to large-displacement image stabilization [31], gesture recognition [40], hand activity summarization [41], and scene annotation via augmented reality [42].

A. Methodology Applied

Section II of this paper argued that the fundamental benefit of an active camera is the concentration of high resolution and high FOV in a small volume, and noting the greater degree of independence, this wins the sensor from the posture of the passive camera with that from the active wearable for two shutter speeds. The wearer performed turns of his torso at about 60° s\(^{-1}\). Column (c) shows that even at the longer aperture time of 30 ms, the active camera preserves much of the detail. Graphs of the angular stabilization performance are shown in Fig. 16. By using inertial sensing alone, displacement error is controlled to a few degrees but is further reduced by an order of magnitude when a vision-based tracker is used.

The proposed methodology was used to assess location of the wearable using the attributes that resulted in Fig. 13, i.e., emphasis on overall FOV, resilience to walking motion, and visibility of the handling space, exactly those expected to be important in applications such as hand tracking and gesture recognition [40], or as an interface for a remote collaborator annotating the wearer’s workspace [42]. We note that in the resulting location on the shoulder, the FOV is reduced, as
Fig. 17. (a) Manipulation event distribution around the wearable camera’s FOV sphere for a particular task of handling office objects (a calculator, keyboard, etc., and whether the hand in view is only one or at rest). (b) FOV sphere (as shown from outside-in) with crosses indicating the mean positions of hand activity detections and the border line computed by the convex hull of all the observed positions during the sequence (see [41] for details).

Wearable sensor development can benefit from incorporating experiments on the human form through modeling and simulation. A computational way to quantify the performance of sensor placement for a wearable camera has been proposed. The simulation environment featured a human-shaped triangular facet model, divided up into body segments that could be independently rotated, allowing quantitative analysis and essentially exhaustive testing. The results presented here consider the effects of absolute FOV, the quality of the view of the handling region, and dependence on the wearer’s own motion. However, other criteria might also be taken into account; for example, areas related to comfort [44]. The simulation environment allowed clear answers to be given on how far the sensor should be raised from the body’s surface to fulfil a given performance.

Pursuing the issue of occlusion led to the analysis of two-camera configurations, determining how effective a given placement would be if we consider multiple cameras to complement the FOV. It is worth noting that occlusion is not just a matter of the effectiveness of an individual sensor. Overcoming it requires the use of more sensors, which in turn impacts on power consumption, of key concern in wearable computing.

A brief exploration of the linear combination of the figures of merit for the various criteria was conducted. What emerges is that if the absolute FOV is regarded as paramount, then positions near the head and the shoulders are favored, no matter what emphasis is placed on resilience to motion. Of course, without further empirical evidence, it is impossible to use the output as other than a guide, but corroboration of intuition is a worthwhile research outcome as it builds confidence in the reliability and realism of the simulation model.

While the discussion here has concentrated on cameras, the developed methodology could readily be adapted to study other systems, such as compound eyes or laser pointers. Further enhancements to the modeling might involve, first, simulating
It is worth emphasizing again that the aim here has been to present a general methodology, but the one illustrated by constraints was derived from particular applications of our wearable camera system. While these particular applications argue for an active camera located at the shoulder, it should not be used to conclude that there is an overall best or second best placement for a wearable camera.

VIII. SIMULATION SOFTWARE

The simulation tool (coded in MATLAB), human figure, and motion capture sequences are available at http://www.cs.bris.ac.uk/~wmayol/VirtualWearer.tgz.

TABLE II

<table>
<thead>
<tr>
<th>Zone</th>
<th>FOV</th>
<th>Best suited to task</th>
<th>Body Motion</th>
<th>Social acceptability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>wide*</td>
<td>linked to user’s attention</td>
<td>large</td>
<td>low</td>
</tr>
<tr>
<td>Shoulder</td>
<td>wide*</td>
<td>most tasks*</td>
<td>small*</td>
<td>medium(*)</td>
</tr>
<tr>
<td>Chest</td>
<td>limited</td>
<td>linked to user’s workspace</td>
<td>small*</td>
<td>high*</td>
</tr>
</tbody>
</table>

more varied motion models related to particular work environments and to include stochastic elements.

It is certainly possible to incorporate prediction to assist compensation. For example, in [45], head motions were predicted to reduce latency in VR applications, and in [46], muscular activation was accounted for when slaving robot motion from motion of the head. It would also be valuable to consider more detailed spatial models, to include, for example, realistic hair, clothing, and work gear.

Here, we have considered only one area of interest around the wearer, the manipulation space. However, the simulator could readily be turned to applications which require, for example, a clear eye-level view to locate faces, combined with an uninterrupted upward view to determine orientation from sky illumination.

Table II summarizes the attributes of sensor placement based on the current study and for a device intended to be competent in addressing all three sets of frames of reference for wearable cameras: user-, world- and object-centered. The last column of the table remarks on a difficult factor in sensor choice and placement, that of social acceptability. A researcher working in wearable computing wants his devices to survive the changing waves of fashion, and the chances of achieving this are improved if devices do not interfere with social interaction. In this study, head-mounted devices are shown to be able to image the largest part of their surroundings but, consequently, are the most visible and, anecdotally, the most disturbing. The opposite is true of cameras worn on the chest. A camera mounted at the shoulder appears to present a good compromise. Turning to the choice of active versus passive cameras, it may well be that an active and animate device, able to show where it is looking, is more acceptable than the blank unwavering stare of a passive camera.

It is worth emphasizing again that the aim here has been to present a general methodology, but the one illustrated by constraints was derived from particular applications of our wearable camera system. While these particular applications argue for an active camera located at the shoulder, it should not be used to conclude that there is an overall best or second best placement for a wearable camera.

REFERENCES


Walterio W. Mayol-Cuevas (M’94) received the degree in computer engineering from the National University of Mexico (UNAM), Mexico City, Mexico, in 1999, and the Ph.D. degree, with a thesis investigating wearable active vision, from the University of Oxford (St. Anne’s College), Oxford, U.K., in 2004. He was a Founding Member of the LINDA Research Group, UNAM. In September 2004, he was appointed as a Lecturer (Assistant Professor) with the Department of Computer Science, University of Bristol, Bristol, U.K., where he is currently engaged in research on active vision and wearable computing and robotics.


He was with the Active Vision Laboratory as a Postdoctoral Research Fellow until 2004. He re-joined the University of Cambridge and worked in image matching for visual interaction as part of the Cambridge–MIT Institute. Since 2005, he has been with MathWorks, Cambridge, where he is currently a Senior Consultant Engineer.

David W. Murray (M’93) received the degree in physics (with first class honors) and the Ph.D. degree in low-energy nuclear physics from the University of California, Berkeley, CA, in 1977 and 1980, respectively.

He was a Research Fellow in physics with the California Institute of Technology before joining the General Electric Company’s research laboratories in 1982. He moved to the University of Oxford, Oxford, U.K., in 1989, as a University Lecturer in engineering science and a Fellow of St Anne’s College. Since 1997, he has been a Professor of engineering science with the Department of Engineering Science, University of Oxford. His research interests continue to center on active approaches to visual sensing, with applications in surveillance, navigation, telerobotics, and wearable computing.

Prof. Murray is a fellow of the Institution of Electrical Engineers in the U.K.