2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops

An Introduction to the 3rd Workshop on Egocentric (First-person) Vision

Steve Mann, Kris M. Kitani, Yong Jae Lee, M. S. Ryoo, Alireza Fathi

Abstract

Egocentric vision provides a unique perspective of the visual world that is inherently human-centric. Since egocentric cameras are mounted on the user (typically on the user's head), they are naturally primed to gather visual information from our everyday interactions, and can even act on that information in real-time (e.g. for a vision aid). We believe that this human-centric characteristic of egocentric vision can have a large impact on the way we approach central computer vision tasks such as visual detection, recognition, prediction, and socio-behavioral analysis. By taking advantage of the first-person point-of-view paradigm, there have been recent advances in areas such as personalized video summarization, understanding concepts of social saliency, activity analysis with inside-out cameras (a camera to capture eye gaze and an outward-looking camera), recognizing human interactions and modeling focus of attention. However, in many ways people are only beginning to understand the full potential (and limitations) of the first-person paradigm. In the 3rd workshop on Egocentric (First-Person) Vision, we bring together researchers to discuss emerging topics such as: Personalization of visual analysis; Socio-behavioral modeling; Understanding group dynamics and interactions; Egocentric video as big data; First-person vision for robotics; and Egographical User Interfaces (EUIs).

1. Introduction

Egocentric *vision* systems have recently been made practical through portable wearable cameras and now enable a wide range of computer vision technologies. Traditionally applications of computer vision have been from a 3rd person point-of-view (archicentric), through cameras affixed to architecture, *i.e.*, outside or inside buildings, or on property (*e.g.* on other structures like lamp posts). More recently, miniaturization of cameras has made human-centric vision possible, with cameras worn by humans [40, 41].

Recent work in egocentric computer vision has focused particularly on a head-worn ('outside') sensor enabling the camera to capture information about the users motion and focus of attention [26, 57]. In some cases, the outside looking camera is supplemented with an inside looking camera to measure a user's eye gaze [76, 59]. Using this 'inside-out' sensing platform, computer vision researchers have proposed novel tasks such as social saliency estimation as well as addressing traditional tasks such as activity recognition and video summarization. This year's workshop features new emerging topics such as:

- Multi-agent egocentric vision systems;
- Privacy preserving techniques and applications;
- Attention-based activity analysis;
- Social interaction analysis;
- Navigation for the blind;
- Hand pose analysis.

Of particular interest to this year's workshop are the following topics:

- Egographical User Interfaces (EUI);
- Assistive technologies for the blind;
- Understanding social dynamics and attention;
- Revisiting robotic vision as egocentric sensing.

In the following we give a brief review of vision-based work leading up to this workshop, followed by an introduction the concept of Egography written by Steve Mann.

2. A Brief History

The idea of using a wearable camera as a sensing modality has a history of at least 35 years dating back to the WearComp work of Mann [41, 40] in the 70s, with wearable computational photography in the 1980s (High Dynamic Range seeing aid, WearCam panoramics, etc.). In the 90s with the introduction of digital cameras and video recording devices like Sixth Sense [42], and later, Microsoft SenseCam (Fig 2), researchers began experimenting with large scale egocentric recordings of human life – lifelogging – which can also function as a visual memory aid [37]. Much of the early work from the late 90s using computer vision techiques were focused on the wearable camera as a gestural interface [40, 21, 42, 69, 70]. While early work helped highlight the opportunity to develop computer vision algorithms for mobile egocentric vision systems, the focus was primarily on analysis of fingers, hands and faces [70, 29, 37, 40].

Transitioning into the early 2000s, contextual visual scene information was automatically analyzed to a limited extent for place recognition [69], localization [28] and video

DOI 10.1109/CVPRW.2014.133

tagging for lifelogging [2]. Place recognition was later explored more fully by Torralba *et al*. [75] to classify scenes with a wearable camera. These pioneering works implicitly took advantage of the characteristics of the egocentric perspective (*i.e*. scenes and people are observed roughly from the same human-centric perspective). During this time, Land and Hayhoe [30] also began exploring the relationship between eye gaze and hand motion, laying the conceptual groundwork for later integrating outside and inside egocentric vision systems [54, 17, 34].

In the mid-2000s, early work by Mayol and Murray [49] began to explore the more general task of visually recognizing hand-object interactions from the first-person perspective. The work of Mayol *et al*. [50, 51] also examined the placements of wearable cameras in the context of active vision. During this time, more robust methods for tracking hands from the egocentric perspective were also developed [27] influencing the development of many HCI applications.

In 2009, the first workshop on egocentric vision was held in conjunction with CVPR, raising the visibility of the egocentric sensing paradigm. The first workshop organized by Philipose, Hebert and Ren, featured topics such as object analysis [63, 72], activity analysis [74, 68] and scene understanding [25, 16, 20]. The first workshop helped to bring together computer vision researchers to develop more advanced component-level technologies and to understand the challenges of working with egocentric vision.

Entering into the 2010s, the computer vision community began to revisit the egocentric paradigm, proposing new computational techniques for egocentric analysis. Initial approaches for such tasks as place recognition [23], object detection [53] and temporal segmentation for sports videos [26], addressed component technologies reminiscent of lifelogging and wearable computational photography applications of the past. The work of Fathi *et al*. further explored basic tasks such as object recognition [14] and activity recognition [11]. The main technical developments with respect to egocentric vision, were the adaptation of image/video features developed for 3rd person POV activity analysis and image understanding to egocentric vision.

In 2012, the second workshop on egoecentric vision was organized by Rehg, Ramanan, Ren, Fathi and Pirsiavash, to gather researchers to discuss the challenge and future directions of egocentric vision [62]. In a 2012 publication, Kanade and Hebert [24] argued that the egocentric perspective is an inverse to the traditional surveillance perspective, and that it "senses the environment and the subject's activities from a wearable sensor, is more advantageous [than surveillance] with images about the subject's environment as taken from his/her view points" (further discusion in Section 3.1). Based on this insight, researchers began to uncover some of the unique properties of egocentric vision to define and address new problems previously not applicable to 3rd POV sensing.

To date researchers have explored the use of egocentric vision for activity recognition [60, 11, 10], object recognition [63, 14], summarization [31, 35], temporal segmentation [68, 26, 61], scene understanding [65], interaction analysis [13, 66], hand detection [33, 32], gaze estimation [76], gaze analysis [54, 80, 34, 10], visual saliency [78, 79], social saliency [57] and motion capture [67].

While many previously proposed approaches have been adapted from traditional computer vision tasks using known techniques from the 3rd POV, approaches addressing the unique properties [24] of the egocentric paradigm (*e.g*., saliency, focus of attention) are beginning to emerge. With this historical context in mind, it is the goal of this third workshop to elucidate directions and challenges for computer vision researchers which are unique to the egocentric vision paradigm.

3. Introduction to Egography (by Steve Mann)

Egography (Greek for "ego"="self" and "graph"="instrument for recording" or "something written") is the capture, processing, transmission, display, etc., of, or computationally interacting with, egocentric (first-person) photographic, videographic, etc., visual information. It is also known variously as "first-person vision", "inside-out vision", sousveillance (inverse or "inside-out" surveillance), or "personal imaging". "Surveillance" is a French word, which means "to watch" ("veillance") "from above" ("sur" as in words like "sur tax" or "sur charge"). The opposite of surveillance is "sousveillance", formed by replacing the "sur" ("from above") with "sous" ("from below" as in words like "sous-chef" = under-chef or "sous-vide" = under-vacuum). See Fig 4,2,3.

3.1. Surveillance and Sousveillance

Computer vision may be divided into two broad categories:

- surveillance: cameras/sensing, on fixed objects, such as property (e.g. land or buildings); and
- sousveillance: cameras/sensing on people, e.g. "quantified self" (http://en.wikipedia.org/wiki/Quantified_Self).

The primary (#1) definition of **"surveillance"** is:

1. "**a watch kept over a person, group, etc., especially over a suspect, prisoner, or the like**: *The suspects were under police surveillance.*" [1]

Surveillance often consists of cameras affixed to property, i.e. real-estate: either buildings (e.g. mounted to inside or outside walls or ceilings), or to land (e.g. mounted to lamp posts, poles, and the like) [44, 52, 7, 15, 64, 4]. In this sense, surveillance is typically initiated by property owners or property custodians such as governments. Surveillance is well-known and well-studied [36], and there are numerous conferences, symposia, etc., on surveillance.

Figure 1. Sixth Sense is an example of an Egographical User Interface in which first-person gestures are sensed and used as a way of interacting with a computer, as well as with other people, i.e. telepresence, collaborative shared interactive communication, etc.. Picture from Wikimedia Commons.

Figure 2. Camera necklace, 16 years ago and today:Archicentric vision (surveillance) and humacentric vision (sousveillance)

Sousveillance (“undersight”) refers to the less hierarchical and more rhizomic veillance of social networking, distributed cloud-based computing, self-sensing, body-worn vision systems, wearable cameras [48, 44, 45, 77, 15, 52, 22, 4], ego-centric vision (i.e. Personal Imaging) [40, 37, 58, 38, 31, 12, 79], implantable vision (first-person point-of-eye sensing), and Mobile Pervasive Sensing [6]. Sousveillance with social networking is also an important area of study regarding privacy, security, and trust [19].

The term *veillance* is now used, more broadly, to describe a politically-neutral watching or sensing that does not necessarily involve a social hierarchy [8, 9], inclusive of both surveillance (architecture-mounted cameras, sensing and computation) and sousveillance (human-borne cameras, sensing, and computation).

Sixth Sense, as mentioned earlier, is an example of an Egographical User Interface. See Fig 1

The name “Sixth Sense” for egographical user interfaces was first coined by S. Mann, as “Synthetic Synesthesia of the Sixth Sense” [18]. A good literature review of the Sixth Sense system is provided in [81], which itself represents an important contribution to the field of egography by A. Yeole *et al*.

Egographic Systems have many applications, including:

- Research, such as activity detection and sensing social ineractions [13, 11, 60, 66, 63, 5], analysis of sports videos [26], etc.;

Figure 3. Egographic computing at point-of-eye (EyeTap)

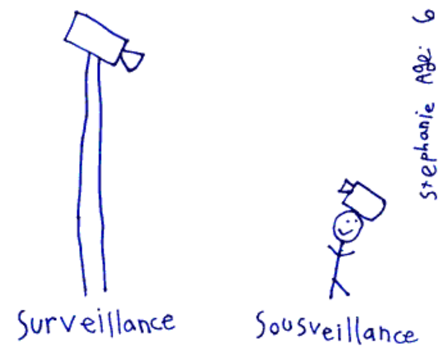

Figure 4. Child’s drawing depicting traditional land/building-centered vision from above (Surveillance) and its “inside-out” reversal (human-centered Sousveillance).

- Personal Safety and Security [39, 43];
- Health, e.g. automatic sensing of dietary intake [43, 55, 73] (first proposed by S. Mann in 2002 [43]), and assisting the blind and visually impaired. See, for example, Fig 7;
- Improved eyesight, Augmediated Reality, e.g. seeing in HDR (High Dynamic Range) for welding (being able to see the electric arc of the welding process clearly, and at the same time being able to see in complete darkness). See Fig 8;
- Communications and remote assistance, e.g. by automatic generation of egocentric panoramas [11, 56, 41] See Fig 5;
- Egographical User Interfaces, e.g. first-person gesture-based interaction, as for example, implemented in Metaview Spaceglasses [47]. See Fig 6.
- Surveillometry: using egography to sense, measure [22], observe, and understand surveillance (e.g. through abakographic [47] imaging processing). See Fig. 9, where a “surveillight” is used to trace out a locus of points in space that are under surveillance. This example is doubly egographical, in the sense that it combines egographic surveillometry with egographical user-interfaces (e.g. ego-gesture-controlled helicopter to create egometrical function spaces).

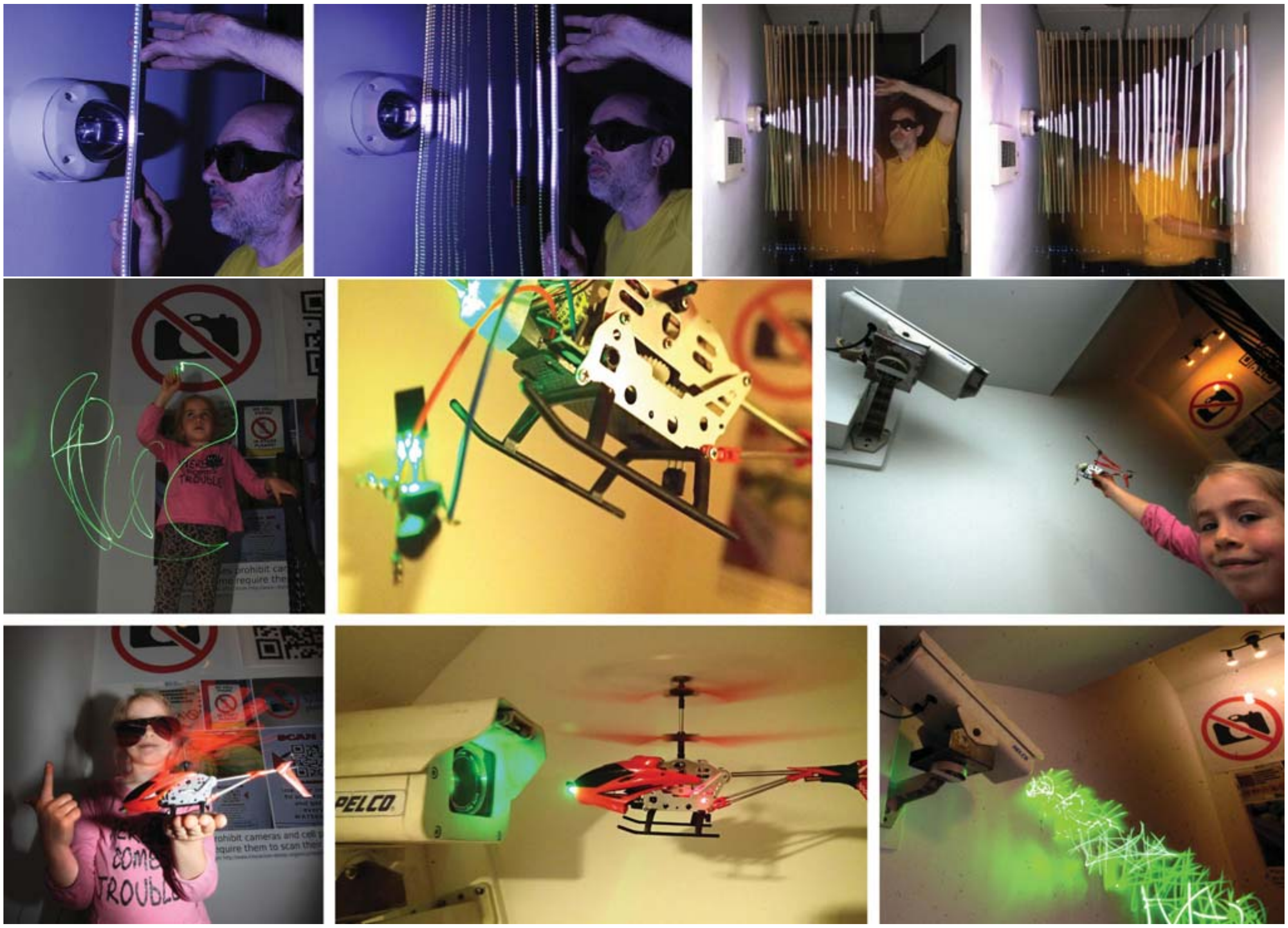

Figure 9. Egographical User Interfaces (sousveillance) for visualizing surveillance. **Top row:** Surveilluminscent stick ("bugbroom") sweeps out the sightfield of a surveillance camera as the stick is waved in front of the surveillance camera, while the sightfield is viewed on another camera (e.g. egographic Spaceglasses) as a time-integrated exposure (scotographic "darkpainting"). **Middle row:** Leftmost: Surveilluminescent "SmartDust" mote that glows green when it is being watched by a surveillance camera. When the surveilluminescent mote is affixed to a hellicopter ("surveillacopter") it "paints out" the sightfield of a surveillance camera. Thus sightfields of surveillance cameras may be made visible using egographic Spaceglasses (spatial imaging glasses) as sousveillance. Housing removed to reduce weight for increased battery life with surveillographic payload. **Bottom row:** Egographic User-Interface using Spaceglass for gesture-based control of the completed surveillometric helicopter. Bottom right image: sightfield visualized using sightpainting (as the light source moves around in front of the camera and glows only when the camera "sees" it). Surveillicopter by Stephanie Mann, Age 7, at the Tinquiry GENIUSchool.

4. The Future of Egographic Systems

Egographic Systems such as Egographic User Interfaces show great promise in a wide range of applications, and egography has become an important field of research. Much remains to be done before egography becomes widespread. One important area of work remains in matters of public acceptability, privacy, security, trust, and the like [3, 71]. For some future predictions, see [46].

References

[1] Online etymology dictionary, douglas harper. 2010.

[2] K. Aizawa, K. Ishijima, and M. Shiina. Summarizing wearable video. In *ICIP*, 2001.

[3] M. A. Ali and S. Mann. The inevitability of the transition from a surveillance-society to a veillance-society: Moral and economic grounding for sousveillance. In *Technology and Society (ISTAS), 2013 IEEE International Symposium on*, pages 243–254. IEEE, 2013.

[4] V. Bakir. *Sousveillance, media and strategic political communication: Iraq, USA, UK*. Continuum International Publishing Group, 2010.

[5] A. Betancourt, M. M. López, C. S. Regazzoni, and M. Rauterberg. A sequential classifier for hand detection in the framework of egocentric vision.

[6] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6(2):161–180, 2010.

[7] J. Bradwell and K. Michael. Security workshop brings' sousveillance'under the microscope. 2012.

[8] A. Cavoukian. Privacy by design. *Gov't of Ontario, Inf. and Priv. Commissioner office*, 2013.

[9] A. Cavoukian, S. Shapiro, and R. J. Cronk. Privacy engineering: Proactively embedding privacy, by design. *Gov't of Ontario, Inf. and Priv. Commissioner office*, 2014.

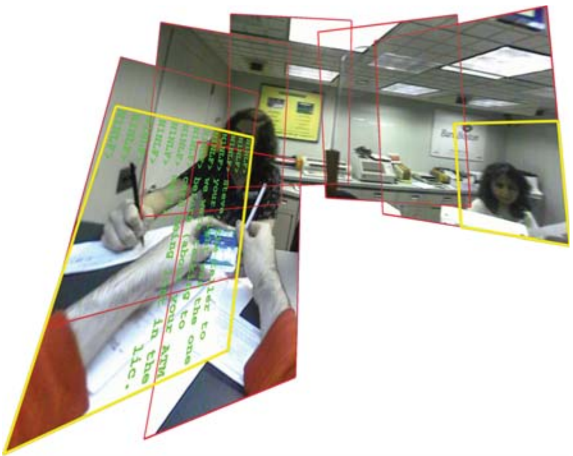

Figure 5. Realtime egography (first-person computer vision) and communications system for use in everyday life. S. Mann, Sep 5, 1996

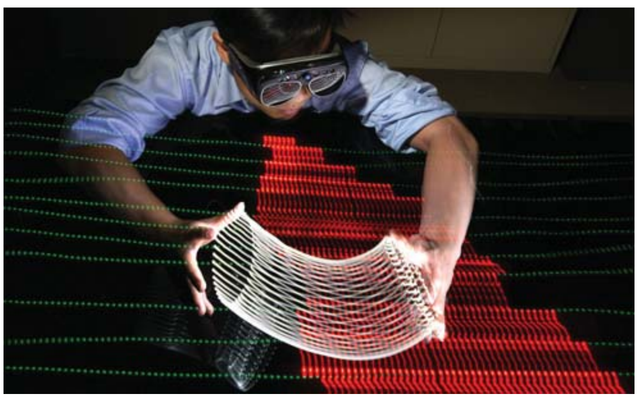

Figure 6. Gesture-based interaction using Metaview Spaceglasses. The veillance flux (sightfield) from the Spaceglass is indicated in red. The lightfield (from which the veillance flux is sampled), is indicated in green. The virtual object being manipulated is shown in white.

[10] A. Fathi. *Learning Descriptive Models of Objects and Activities from Egocentric Video*. PhD thesis, Georgia Institute of Technology, 2013.

[11] A. Fathi, A. Farhadi, and J. M. Rehg. Understanding egocentric activities. In *ICCV*, 2011.

[12] A. Fathi, Y. Li, and J. M. Rehg. Learning to recognize daily actions using gaze. In *Computer Vision–ECCV 2012*, pages 314–327. Springer, 2012.

[13] A. Fathi and J. M. Rehg. Social interactions: A first-person perspective. In *CVPR*, 2012.

[14] A. Fathi, X. Ren, and J. M. Rehg. Learning to recognize objects in egocentric activities. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*, pages 3281–3288. IEEE, 2011.

[15] G. Fletcher, M. Griffiths, and M. Kutar. A day in the digital life: a preliminary sousveillance study. *SSRN, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1923629*, September 7, 2011.

[16] A. Flint, I. Reid, and D. Murray. Learning texton models for real-time scene context. In *WECV*, 2009.

[17] H. Fujiyoshi, Y. Goto, and M. Kimura. Inside-out camera for acquiring 3d gaze points. In *WECV*, 2012.

[18] J. Geary. *The body electric: An anatomy of the new bionic senses*. Rutgers University Press, 2002.

Figure 7. Implantable Egographic System: First-person perspective from camera implant. Inventor S. Mann (July 19, 2000). Collaborators R. Spence, D. Desjardins, and others.

Quantigraphic camera promises HDR eyesight from Father of AR

Chris Davies, Sep 12th 2012 Discuss [1]

Traditional welding helmets use a sheet of smoked glass for the eyepiece, cutting down on the dangerous glare from the welding process itself, but also reducing overall visibility. The HDRrchitecture system, instead, processes images coming from one or more cameras, rendering a Full HD, 30fps stream with the brighter elements stripped out but the core details retained, all in real-time.

Figure 8. Egographic welding helmet: First-person point-of-eye (for each eye) system provides AR (Augmediated Reality) to see a dynamic range of more than 100 million-to-one, while simultaneously providing computer-generated overlays along with advice streaming live from remote experts. S. Mann, invention 1992, picture taken 2011.

[19] B. Greschbach, G. Kreitz, and S. Buchegger. The devil is in the metadatanew privacy challenges in decentralised online social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 333–339. IEEE, 2012.

[20] M. Havlena, A. Ess, W. Moreau, A. Torii, M. Jancosek, T. Pajdla, and L. Van Gool. Awear 2.0 system: Omni-directional audio-visual data acquisition and processing. In *WECV*, 2009.

[21] J. Healey and R. W. Picard. Startlecam: A cybernetic wearable camera. In *ISWC*, 1998.

[22] R. Janzen and S. Mann. Vixels, veillons, veillance flux: An extromissive information-bearing formulation of sensing, to measure surveillance and sousveillance. *Proc. IEEE CCECE2014*, page (to appear), 2014.

[23] Y. Kameda and Y. Ohta. Image retrieval of first-person vision for pedestrian navigation in urban area. In *ICPR*, 2010.

[24] T. Kanade and M. Hebert. First-person vision. *Proceedings of the IEEE*, 100(8):2442–2453, 2012.

[25] H. Kang, A. A. Efros, M. Hebert, and T. Kanade. Image matching in large scale indoor environment. In *WECV*, 2009.

[26] K. M. Kitani, T. Okabe, Y. Sato, and A. Sugimoto. Fast unsupervised ego-action learning for first-person sports videos. In *CVPR*, 2011.

[27] M. Kolsch and M. Turk. Fast 2d hand tracking with flocks of features and multi-cue integration. In *Workshop on Real-Time Vision for Human-Computer Interaction*, 2004.

[28] M. Kourogi and T. Kurata. Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera. In *ISMAR*, 2003.

[29] T. Kurata, T. Okuma, M. Kourogi, and K. Sakaue. The hand mouse: Gmm hand-color classification and mean shift tracking. In *Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 2001.

[30] M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41(25):3559–3565, 2001.

[31] Y. J. Lee, J. Ghosh, and K. Grauman. Discovering important people and objects for egocentric video summarization. In *CVPR*, 2012.

[32] C. Li and K. M. Kitani. Model recommendation with virtual probes for egocentric hand detection. In *ICCV*, 2013.

[33] C. Li and K. M. Kitani. Pixel-level hand detection in ego-centric videos. In *CVPR*, 2013.

[34] Y. Li, A. Fathi, and J. M. Rehg. Learning to predict gaze in egocentric video. In *ICCV*, 2013.

[35] Z. Lu and K. Grauman. Story-driven summarization for egocentric video. In *CVPR*, 2013.

[36] D. Lyon. *Surveillance Studies An Overview*. Polity Press, 2007.

[37] S. Mann. **Wearable, tetherless computer–mediated reality**: WearCam as a wearable face–recognizer, and other applications for the disabled. TR 361, M.I.T. Media Lab Perceptual Computing Section; Also appears in **AAAI Fall Symposium on Developing Assistive Technology for People with Disabilities**, 9-11 November 1996, MIT; http://wearcam.org/vmp.htm, Cambridge, Massachusetts, February 2 1996.

[38] S. Mann. An historical account of the 'WearComp' and 'WearCam' projects developed for 'personal imaging'. In *International Symposium on Wearable Computing*, pages 66–73, Cambridge, Massachusetts, October 13-14 1997. IEEE.

[39] S. Mann. Smart clothing: The wearable computer and wearcam. *Personal Technologies*, 1(1):21–27, 1997.

[40] S. Mann. Wearable computing: A first step toward personal imaging. *IEEE Computer*, 30(2):25–32, 1997.

[41] S. Mann. Humanistic intelligence/humanistic computing: 'wearcomp' as a new framework for intelligent signal processing. *Proceedings of the IEEE*, 86(11):2123–2151+cover, Nov 1998.

[42] 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 *ISWC*, 1998.

[43] S. Mann. Slip and fall detector, method of evidence collection, and notice server, for visually impaired persons, or the like, 2002. US Patent App. 20020198685.

[44] S. Mann. Sousveillance, not just surveillance, in response to terrorism. *Metal and Flesh*, 6(1):1–8, 2002.

[45] S. Mann. Sousveillance: inverse surveillance in multimedia imaging. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 620–627. ACM, 2004.

[46] S. Mann. Eye am a camera: Surveillance and sousveillance in the glassage. *TIME*, Nov. 02, 2012.

[47] S. Mann, R. Janzen, T. Ai, S. N. Yasrebi, J. Kawwa, and M. A. Ali. Toposculpting: Computational lightpainting and wearable computational photography for abakographic user interfaces. In *Proceedings of the IEEE CCECE*. IEEE, 2014.

[48] S. Mann, J. Nolan, and B. Wellman. Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments. *Surveillance & Society*, 1(3):331–355, 2003.

[49] W. Mayol and D. Murray. Wearable hand activity recognition for event summarization. In *ISWC*, pages 122–129, 2005.

[50] W. Mayol, B. Tordoff, and D. Murray. On the positioning of wearable optical devices. Technical report, Technical Report OUEL224101. Oxford University, 2001.

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

[52] K. Michael and M. Michael. Sousveillance and point of view technologies in law enforcement: An overview. 2012.

[53] H. Nakayama, T. Harada, and Y. Kuniyoshi. Ai goggles: real-time description and retrieval in the real world with online learning. In *Canadian Conference on Computer and Robot Vision*, 2009.

[54] K. Ogaki, K. M. Kitani, Y. Sugano, and Y. Sato. Coupling eye-motion and ego-motion features for first-person activity recognition. In *WECV*, 2012.

[55] G. O'Loughlin, S. J. Cullen, A. McGoldrick, S. O'Connor, R. Blain, S. O'Malley, and G. D. Warrington. Using a wearable camera to increase the accuracy of dietary analysis. *American journal of preventive medicine*, 44(3):297–301, 2013.

[56] T. Ozawa, K. M. Kitani, and H. Koike. Human-centric panoramic imaging stitching. In *Proceedings of the 3rd Augmented Human International Conference*, page 20. ACM, 2012.

[57] H. S. Park, E. Jain, and Y. Sheikh. 3d social saliency from head-mounted cameras. In *NIPS*, 2012.

[58] L. Piccardi, B. Noris, O. Barbey, A. Billard, G. Schiavone, F. Keller, and C. von Hofsten. Wearcam: A head mounted wireless camera for monitoring gaze attention and for the diagnosis of developmental disorders in young children. In *ROMAN*, 2007.

[59] B. R. Pires, M. Devyver, A. Tsukada, and T. Kanade. Unwrapping the eye for visible-spectrum gaze tracking on wearable devices. In *WACV*, 2013.

[60] H. Pirsiavash and D. Ramanan. Detecting activities of daily living in first-person camera views. In *CVPR*, 2012.

[61] Y. Poleg, C. Arora, and S. Peleg. Temporal segmentation of egocentric videos. In *CVPR*, 2014.

[62] J. Rehg, D. Ramanan, X. Ren, A. Fathi, and H. Pirsiavash. Second workshop on egocentric (first-person) vision, 2012.

[63] X. Ren and M. Philipose. Egocentric recognition of handled objects: Benchmark and analysis. In *WECV*, 2009.

[64] C. Reynolds. Negative sousveillance. *First International Conference of the International Association for Computing and Philosophy (IACAP11)*, pages 306 – 309, July 4 - 6, 2011, Aarhus, Denmark.

[65] A. Rituerto, A. C. Murillo, and J. J. Guerrero. Line image signature for scene understanding with a wearable vision system. In *SenseCam & Pervasive Imaging Conference*, 2013.

[66] M. S. Ryoo and L. Matthies. First-person activity recognition: What are they doing to me? In *CVPR*, 2013.

[67] T. Shiratori, H. S. Park, L. Sigal, Y. Sheikh, and J. K. Hodgins. Motion capture from body-mounted cameras. In *SIGGRAPH*, 2011.

[68] E. H. Spriggs, F. De La Torre, and M. Hebert. Temporal segmentation and activity classification from first-person sensing. In *WECV*, 2009.

[69] T. Starner, B. Schiele, and A. Pentland. Visual contextual awareness in wearable computing. In *ISWC*, 1998.

[70] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition using desk and wearable computer based video. *PAMI*, 20(12):1371–1375, 1998.

[71] K. D. Stephan, K. Michael, M. G. Michael, L. Jacob, and E. P. Anesta. Social implications of technology: The past, the present, and the future. *Proceedings of the IEEE*, 100:1752–1781, 2012.

[72] L. Sun, U. Klank, and M. Beetz. Eyewatchme3d hand and object tracking for inside out activity analysis. In *WECV*, 2009.

[73] M. Sun, J. D. Fernstrom, W. Jia, S. A. Hackworth, N. Yao, Y. Li, C. Li, M. H. Fernstrom, and R. J. Sclabassi. A wearable electronic system for objective dietary assessment. *Jour. of the Amer. Dietetic Assoc.*, 110(1):45, 2010.

[74] S. Sundaram and W. W. M. Cuevas. High level activity recognition using low resolution wearable vision. In *WECV*, 2009.

[75] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin. Context-based vision system for place and object recognition. In *ICCV*, 2003.

[76] A. Tsukada, M. Shino, M. Devyver, and T. Kanade. Illumination-free gaze estimation method for first-person vision wearable device. In *Workshop on Computer Vision in Vehicle Technology*, 2011.

[77] R. Vertegaal and J. S. Shell. Attentive user interfaces: the surveillance and sousveillance of gaze-aware objects. *Social Science Information*, 47(3):275–298, 2008.

[78] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto, and K. Hiraki. Can saliency map models predict human egocentric visual attention? In *ACCV Workshop*, 2010.

[79] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto, and K. Hiraki. Attention prediction in egocentric video using motion and visual saliency. In *Advances in Image and Video Technology*, pages 277–288. Springer, 2012.

[80] Z. Ye, Y. Li, A. Fathi, Y. Han, A. Rozga, G. D. Abowd, and J. M. Rehg. Detecting eye contact using wearable eye-tracking glasses. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 699–704. ACM, 2012.

[81] A. Yeole, S. C. V. Sriraman, P. Shinde, and P. Sadhwani. Virtual keyboard. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 3:518–522.