# Bringing Vision-Based Measurements into our Daily Life: A Grand Challenge for Computer Vision Systems

*Jacob Scharcanski\**

*Programa de Pós-Graduação em Engenharia Elétrica, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil*

Bringing computer vision into our daily life has been challenging researchers in industry and in academia over the past decades. However, the continuous development of cameras and computing systems turned computer vision-based measurements into a viable option, allowing new solutions to known problems. In this context, computer vision is a generic tool that can be used to measure and monitor phenomena in wide range of fields. The idea of using vision-based measurements is appealing, since these measurements can be fast and easy to obtain and handle by humans. On the other hand, these vision-based measurements need to be at least as precise and accurate as those obtained by more traditional measuring systems. This article discusses the perspectives and challenges of introducing vision-based measurements into our daily life.

Keywords: computer vision, measurements, instrumentation, computer vision, artificial intelligence, intelligent computing, sensor fusion

## 1. INTRODUCTION

Bringing computer vision into our daily life has been challenging researchers in industry (Google, 2015) and in academia (Shirmohammadi and Ferrero, 2014; Fernandez-Caballero, 2015) over the past decades. The continuous development of cameras and computing devices to improve the ability to capture, interpret, and understand images turned computer vision-based instruments into a viable option (Google, 2016), providing new solutions to known problems. In this context, computer vision becomes a generic tool that can be used to measure and monitor phenomena in wide range of fields.

However, considering measurements used on a daily basis, such as measuring the volume of a room or trying to find out if a piece of furniture fits in a room, even such *simple* tasks involve complex computational procedures (e.g., using visual–inertial odometry to provide an accurate three-dimensional reference system for indoors scenes, or motion tracking to capture three-dimensional scene representations using sequences of two-dimensional images of uncontrolled environments) (Google, 2015). Despite the computational complexity involved in such tasks, new computer vision developments potentially can help facing such challenging tasks at manageable costs (Google, 2016). For example, computer vision systems could help improving the quality of the images taken in the presence of reflecting or occluding elements, such as windows and fences (Xue et al., 2015), compensating the viewpoint from which the image was captured in three-dimensional visual data processing (Satkin and Hebert, 2013), localizing the user in the surrounding environment, so that indoor and outdoor scenes can be represented in referenced spaces (Bettadapura et al., 2015), or yet

capturing automatically the relationships among visual elements in real-world scenarios (Choi et al., 2015).

The idea of using vision-based measurements in daily life is appealing, since these measurements can be obtained fast and easily by humans. On the other hand, these vision-based measurements need to be at least as precise and accurate as those produced by more traditional measuring systems. To avoid any confusion with colloquial terms, accuracy evaluates how close the measurement of a quantity is to its true value [Bureau International des Poids et Mesures (BIPM), 2008], and precision reflects the degree to which repeated measurements show the same results under the same conditions. For example, a vision-based measurement system that produces a systematic error would be consistent but inaccurate in practice. An important challenge is to develop vision-based measurement systems that are both accurate and precise under the conditions found in practice.

## 2. RELIABLE MEASUREMENTS *VIA* COMPUTER VISION

Several technological developments rely on the acquisition and processing of images and videos to extract and enhance information (Scharcanski et al., 1993; Cavalcanti et al., 2010), in areas ranging from engineering (Scharcanski and Dodson, 1966, 2000; Scharcanski, 2005) and medicine (Wong et al., 2011) to biometric authentication (Behaine and Scharcanski, 2012; Verdoolaege et al., 2014). In particular, the growing processing capability and the improvement of the camera quality of consumer electronic devices (Google, 2016), such as smartphones and laptops, allow to obtain a wide range of vision-based measurements in a variety of areas, such as monitoring food intake, assisting the visual impaired, biometric user authentication, and in several other medicine and engineering applications.

Monitoring and controlling the nutritional intake in daily life can be challenging for chronically ill people and for healthy individuals as well (e.g., monitoring what we eat in restaurants). Computer vision systems can be trained to recognize the contents of meals, estimate the food portions, and predict their nutritional contents, such as calories (Myers et al., 2015). Furthermore, computer vision systems provide the adequate tools to capture visual cues during food preparation and help control of the nutritional intake (Malmaud et al., 2015).

Computer vision-based measurements have the untapped potential for helping persons with blindness or visual impairments to interpret and explore the visual world (Manduchi and Coughlan, 2012). This type of assistive technology has several roles to play in our society. For example, the properties of the immediate surroundings could be captured by computer vision systems, helping to measure distances and avoiding obstacles, handling apertures, such as doors, and maintaining a desired trajectory while walking. By measuring and identifying the position of a person within his/her environment, a route to a destination could be planned and tracked. Detecting printed materials could make textual information accessible to the visually impaired by using computer vision-based translators (e.g., assisting a blind person while shopping at a supermarket). Mobile computer

vision devices are promising to deploy these assistive technologies and make them viable for real life applications (Manduchi and Coughlan, 2012; Google, 2016).

The utilization of biometrics has increased specially in modalities, such as face recognition, fingerprint recognition, iris recognition, hand gesture recognition, and palmprint recognition. It is desirable for vision systems to perform biometric recognition covertly, based on contactless data acquisition, and to provide reliable recognition of human beings (Scharcanski et al., 2014). This is especially true considering the applications involving uncontrolled scenarios (e.g., airports, public areas, etc.). Therefore, the development of biometric recognition systems for unconstrained conditions is of current interest, and advances in aspects, such as dealing with varying illumination sources, variations in poses and distances, or blurred and low quality data during image acquisition have been made recently (Scharcanski et al., 2014). However, advances in other areas are also important to allow vision systems operate autonomously and covertly and perform reliable recognition of human beings, such as (a) less controlled/covert data acquisition frameworks, (b) segmentation of poor quality biometric data, (c) biometric data quality assessment, (d) contactless biometric recognition strategies, (e) biometric recognition robustness to data resolution, illumination, distance, pose, motion, and occlusions, and (f) multimodal biometrics and fusion at different levels.

Vision-based measurements relying on cameras available *via* consumer electronics devices, such as smartphones and laptops, can have a great impact in areas, such as personalized health and medicine. For example, assistive vision systems based on RGB-D cameras (e.g., Kinect) can improve the recovery times of physiotherapy patients by increasing their engagement in the rehabilitation sessions [e.g., through the integration with 3D serious games (Duarte et al., 2014)], while providing feedback about their progress in the rehabilitation program. Also, image-based measurements can play a decisive role in early diagnosis of melanoma, which is one of the most rapidly increasing cancers in the world, with an incidence of 76,690 in the United States in 2013 (Scharcanski and Celebi, 2014). Early diagnosis is very important, since melanoma can be cured with a simple excision if detected early. Recently, macroscopy (i.e., skin lesion imaging with standard cameras) has proved valuable in evaluating morphological structures in pigmented lesions (Cavalcanti et al., 2010, 2013; Wong et al., 2011; Cavalcanti and Scharcanski, 2013; Scharcanski and Celebi, 2014). Although computerized techniques are not intended to provide a definitive diagnosis, these imaging technologies are accessible and can be useful in early melanoma detection (especially for patients with multiple atypical nevi) and can serve as an adjunct to physicians.

Computer vision systems can help overcoming common limitations of the human visual system in daily tasks (e.g., in repetitive tasks that are tiresome for humans, or in tasks that require fast response, or yet in tasks that require visual discrimination beyond the human visual system capacity). Therefore, vision-based measurements can provide an alternative to the measurement methods used traditionally in different areas, such as medicine, traffic monitoring, and industry. For example, infrared image data captured by non-mydriatic digital

retinography systems are used to diagnose and treat diabetic macular edema (DME) cases using lasers. However, sequences of infrared eye fundus images are inherently noisy and show frequent background illumination changes, increasing the risk of DME laser treatments for the patients. However, highly accurate vision-based measurements (with a precision level of micrometers) can be used, for example, to detect the DME lesion spots automatically and accurately, or for detecting important anatomical retinal structures automatically (e.g., the optic disk or the macula), or yet for detecting fast even small retinal motions before the main laser hits healthy regions of the retina and compensate for involuntary retinal motions of the patient, increasing the safety of retinal laser treatment systems for DME (Scharcanski et al., 2013; Welfer et al., 2013).

Detecting and counting vehicles in urban traffic videos can be a tedious task and/or may require fast human operator responses. However, such measurements can be performed effortlessly by vision systems at user-defined virtual loops with the accuracy and precision comparable to conventional traffic monitoring approaches (e.g., by detecting and counting vehicles at user-defined virtual loops in video sequences) (Barcellos et al., 2015). However, partial and total occlusions in vehicular tracking situations can cause vehicle tracking and counting errors, posing a challenge to human controllers. Nevertheless, vision-based measurement systems can track multiple vehicles simultaneously, detect when a vehicle total occlusion starts and ends, and resume the vehicle tracking process after vehicle disocclusions, minimizing such difficulties (Scharcanski et al., 2011). Furthermore, critical visual tasks, such as pedestrian detection, are increasingly important in vehicular technology (e.g., autonomous driving). The integration of deep learning with computer vision tools has the untapped potential for approaching such visual tasks, where reduced miss rates and real-time are very important (Angelova et al., 2015).

Estimating materials properties plays a central role in vision, robotics, and engineering, especially in tasks that require going beyond the limits of human visual perception (e.g., estimating material properties from videos by detecting material vibration modes that are visually imperceptible) (Davis et al., 2015). Currently, large amounts of visual data are acquired, providing important information about the materials produced by continuous manufacturing processes, and about the manufacturing processes themselves. However, it is often difficult to measure objectively the similarity among such industrial stochastic images, or to discriminate between texture images of stochastic materials with distinct properties. The degree of discrimination between materials required by such industrial processes often goes beyond the limits of human visual perception. However, vision-based measurements can be designed to analyze such stochastic processes continuously and effortlessly and to perform tasks that would be challenging for humans (e.g., real-time analysis of industrial stochastic textures in the non-woven textiles and paper industries) (Scharcanski and Dodson, 2000; Scharcanski, 2007).

## 3. CONCLUSION

Computer vision have the potential to make measurement systems easier to use by humans and to allow the development of new measurement tools in engineering, medicine, and other areas. Vision-based measurements have the potential for making measurements even in conditions that would be uncomfortable for humans, for example, in repetitive tasks, or that require fast responses, or yet that require visual discrimination beyond the human visual system capacity. The main challenges in this area involve integrating imaging devices and other sensing devices (e.g., odometry devices) with computer vision tools, aiming to develop measurement systems that are accurate, precise, and can be used easily in daily life.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES

Angelova, A., Krizhevsky, A., and Vanhoucke, V. (2015). "Pedestrian detection with a large-field-of-view deep network," in *IEEE International Conference on Robotics and Automation (Proceedings of ICRA 2015)*, May 26–30. Seattle.

Barcellos, P., Bouvie, C., Escouto, F., and Scharcanski, J. (2015). A novel video based system for detecting and counting vehicles at user-defined virtual loops. *Expert Syst. Appl.* 42, 1845–1856. doi:10.1016/j.eswa.2014.09.045

Behaine, C. A. R., and Scharcanski, J. (2012). Enhancing the performance of active shape models in face recognition applications. *IEEE Trans. Instrum. Meas.* 61, 2330–2333. doi:10.1109/TIM.2012.2188174

Bettadapura, V., Essa, I., and Pantofaru, C. (2015). "Egocentric field-of-view localization using first-person point-of-view devices," in *Proceedings of Winter Conference on Applications of Computer Vision (WACV)*, January 6–9. Waikoloa Beach.

Bureau International des Poids et Mesures (BIPM). (2008). *JCGM 200:2008 International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*. Sèvres.

Cavalcanti, P. G., and Scharcanski, J. (2013). A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images. *Comput. Methods Programs Biomed.* 112, 684–693. doi:10.1016/j.cmpb.2013.08.010

Cavalcanti, P. G., Scharcanski, J., and Baranoski, G. A. (2013). Two-stage approach for discriminating melanocytic skin lesions using standard cameras. *Expert Syst. Appl.* 40, 4054–4064. doi:10.1016/j.eswa.2013.01.002

Cavalcanti, P. G., Scharcanski, J., and Lopes, C. B. O. (2010). "Shading attenuation in human skin color images," in *Advances in Visual Computing, Vol. 6453, Lecture Notes in Computer Science*, eds Bebis G., Boyle R., Parvin B., Koracin D., Chung R., Hammoud R., et al. (Berlin: Springer), 190–198.

Choi, W., Chao, Y., Pantofaru, C., and Savarese, S. (2015). Indoor scene understanding with geometric and semantic contexts. *Int. J. Comput. Vis.* 112, 204–220. doi:10.1007/s11263-014-0779-4

Davis, A., Bouman, K. L., Chen, J. G., Rubinstein, M., Durand, F., and Freeman, W. T. (2015). "Visual vibrometry: estimating material properties from small motion in video," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2015)*, June 7–12. Boston.

Duarte, N., Postolache, O., and Scharcanski, J. (2014). "KSGphysio – kinect serious game for physiotherapy," in *2014 International Conference and Exposition on Electrical and Power Engineering (EPE 2014)*, October 16–18. Iasi.

Fernandez-Caballero, A. (2015). A grand challenge for vision systems: improving the quality of life and care of ageing adults. *Front. Robot. AI* 2:15. doi:10.3389/frobt.2015.00015

Google. (2015). *Project Tango*. Available at: https://developers.google.com/project-tango/

Google. (2016). *Computer Vision in Next-Generation Devices*. Available at: http://techcrunch.com/2016/01/27/google-and-movidius-partner-to-propel-computer-vision-in-next-generation-devices/

Malmaud, J., Huang, J., Rathod, V., Johnston, N., Rabinovich, A., and Murphy, K. (2015). "What's cookin? Interpreting cooking videos using text, speech and vision," in *North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)*, May 31–June 5. Denver.

Manduchi, R., and Coughlan, J. (2012). (Computer) vision without sight. *Commun. ACM* 55, 96–104. doi:10.1145/2063176.2063200

Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, S., et al. (2015). "Im2Calories: towards an automated mobile vision food diary," in *IEEE International Conference on Computer Vision 2015 (Proc. of ICCV 2015)*, December 13–16. Santiago.

Satkin, S., and Hebert, M. (2013). "DNN: viewpoint invariant 3D geometry matching for scene understanding," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2013)*, December 3–6. Sydney.

Scharcanski, J. (2005). Stochastic texture analysis for monitoring stochastic processes in industry. *Pattern Recognit. Lett.* 26, 1701–1709. doi:10.1016/j.patrec.2005.01.017

Scharcanski, J. (2007). A wavelet-based approach for analyzing industrial stochastic textures with applications. *IEEE Trans. Syst. Man Cybern A* 37, 10–22. doi:10.1109/TSMCA.2006.886345

Scharcanski, J., and Celebi, E. (eds) (2014). *Computer Vision Techniques for the Diagnosis of Skin Cancer, Ser. Lecture Notes in Bioengineering*. Berlin: Springer-Verlag.

Scharcanski, J., and Dodson, C. T. J. (1966). Texture analysis for estimating spatial variability and anisotropy in planar stochastic structures. *Opt. Eng.* 35, 2302–2309. doi:10.1117/1.600804

Scharcanski, J., and Dodson, C. T. J. (2000). Stochastic texture image estimators for local spatial anisotropy and its variability. *IEEE Trans. Instrum. Meas.* 49, 971–979. doi:10.1109/19.872916

Scharcanski, J., Oliveira, A., Cavalcanti, P., and Yari, Y. (2011). A particle-filtering approach for vehicular tracking adaptive to occlusions. *IEEE Trans. Veh. Technol.* 60, 381–389. doi:10.1109/TVT.2010.2099676

Scharcanski, J., Proença, H., and Du, E. (eds) (2014). *Signal and Image Processing for Biometrics, Ser. Lecture Notes in Electrical Engineering*. Berlin: Springer-Verlag.

Scharcanski, J., Schardosim, L. R., Santos, D., and Stuchi, A. (2013). Motion detection and compensation in infrared retinal image sequences. *Comput. Med. Imaging Graph.* 37, 377–385. doi:10.1016/j.compmedimag.2013.06.004

Scharcanski, J., Shen, H. C., and Silva, A. P. A. (1993). Colour quantisation for colour texture analysis. *IEE Proc. E Comput. Digit. Tech.* 140, 109–114.

Shirmohammadi, S., and Ferrero, A. (2014). Camera as the instrument: the rising trend of vision based measurement. *IEEE Instrum. Meas. Mag.* 17, 41–47. doi:10.1109/MIM.2014.6825388

Verdoolaege, G., Soldera, J., Macedo, T., and Scharcanski, J. (2014). "Data and information dimensionality in non-cooperative face recognition," in *Signal and Image Processing for Biometrics, Ser. Lecture Notes in Electrical Engineering*, Vol. 292, eds Scharcanski J., Proença H., and Du E. (Berlin: Springer-Verlag), 1–35.

Welfer, D., Scharcanski, J., and Marinho, D. R. (2013). A morphologic two-stage approach for automated optic disk detection in color eye fundus images. *Pattern Recognit. Lett.* 34, 476–485. doi:10.1016/j.patrec.2012.12.011

Wong, A., Scharcanski, J., and Fieguth, P. (2011). Automatic skin lesion segmentation via iterative stochastic region merging. *IEEE Trans. Inform. Technol. Biomed.* 15, 929–936. doi:10.1109/TITB.2011.2157829

Xue, T., Rubinstein, M., Liu, C., and Freeman, W. T. (2015). "A computational approach for obstruction-free photography," in *ACM Transactions on Graphics (Proc. of SIGGRAPH 2015)*, August 9–13. Vol. 34 (Los Angeles: ACM).

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.