https://doi.org/10.2352/ISSN.2470-1173.2019.7.IRIACV-459

People Recognition and Position Measurement in Workplace by Fisheye Camera

Haike Guan, Makoto Shinnishi;
Ricoh Institute of Information and Communication Technology, R&D Division, Ricoh Co. LTD.; Ebina, Kanagawa/Japan

Abstract

In workplace of factory or office, measuring positions of workers is important to make the workplace visible for improving working efficiency, avoiding miss operation and accidents. Monocular fisheye camera is used by conventional methods to find and track people in workplace. But it is difficult for these methods to get 3D position of workers. Stereo fisheye cameras were used to measure 3D position. But calibrating these fisheye cameras is a hard and time-consuming work.

We propose a new method to measure 3D position of people in workplace by only one fisheye camera. One 360-degree fisheye image is projected to a unit sphere and several perspective projection images are generated to correct fisheye distortion. People recognition is made in the corrected images by machine learning method. Recognition results are used to calculate 3D position of people in workplace. Only very few markers are set on floor of workplace to make calibration between perspective projection images and plane of workplace floor.

People recognition and position measurement experiments were made in workplace by Ricoh R 360-degree monocular fisheye cameras. People recognition rate of 92.5%, false positive rate of 0.2%, people position measurement accuracy of 6.6% were obtained. The evaluation results demonstrate effectiveness of our proposed method.

Introduction

In workplace of factory or office, measuring positions of workers is important to make the workplace visible for improving working efficiency, avoiding miss operation and accidents. Information of people's 3D position can also be used for robot navigation, cooperation between robot and people.

Monocular fisheye camera was used by conventional methods to find and track people in workplace. It is difficult for these methods to get 3D position of workers. Stereo fisheye cameras were used to measure 3D position of people. But calibrating these fisheye cameras is a hard and time-consuming work. Because of large fisheye distortion and lack of parallax, it is a challenge to get 3D position of people from only one 360-degree fisheye camera.

Objective of our research is to recognize people and measure his 3D position in workplace by one monocular fisheye camera. Because fisheye camera has a wide angle of view, a large factory can be covered by fewer 360-degree fisheye cameras. Setting the fisheye camera is much easier than setting stereo cameras.

We propose a new method to measure 3D positions of people in workplace by only one monocular fisheye camera. One 360-degree fisheye image is projected to a unit sphere and several perspective projection images are generated to correct fisheye distortion. People recognition are made in the corrected images by machine learning method. Recognition results are used to calculate 3D position of people in workplace. Only very few markers were set on floor of workplace to make calibration between perspective projection images and plane of workplace floor. People recognition and 3D people position measurement experiments made in workplace by Ricoh R 360-degree monocular fisheye cameras demonstrated effectiveness of our proposed method.

Related Work

Many methods have been proposed to recognize pedestrian, for example histograms of oriented gradients features were used for pedestrian recognition [1][2]. Deep learning methods have been used for people recognition, which can get much higher recognition accuracy [3]-[6]. These methods are used for small angle of view cameras without fisheye distortion. Calibration is also easy for the cameras. Stereo camera was used to recognize people and measure distance to get 3-dimension position [7][8]. But calibrating the stereo cameras is a hard and time-consuming work, especially for stereo fisheye cameras [9]. Fisheye surveillance camera was used for recognizing and tracking people. Although fisheye distortion is corrected at first, then people is recognized [10]-[14], it is difficult for these methods to correct 360-degree fisheye camera's distortion as shown in Figure 1. Recognizing pedestrian directly from distorted fisheye image was propose [15]. But collecting learning image data is a hard work. Because fisheye distortion is different at different position in fisheye image, many recognition models should be made for one fisheye image. The people recognition models are made for a specific camera. To apply to a different fisheye camera, we should collect learning image data and train models again. It is difficult for conventional methods to recognize people and get his 3D position from one 360-degree fisheye camera.

Proposed Method

We propose a new method to measure 3D positions of people in workplace by only one monocular fisheye camera. As shown in Figure 3, flow chart of the proposed method, one 360-degree fisheye image is used to recognize people and get his 3-dimension position. The fisheye image is projected to a unit sphere and several perspective projection images are generated to correct fisheye distortion.

Figure 1. *360-degree fisheye image*

Figure 2. *Fisheye distortion corrected image*

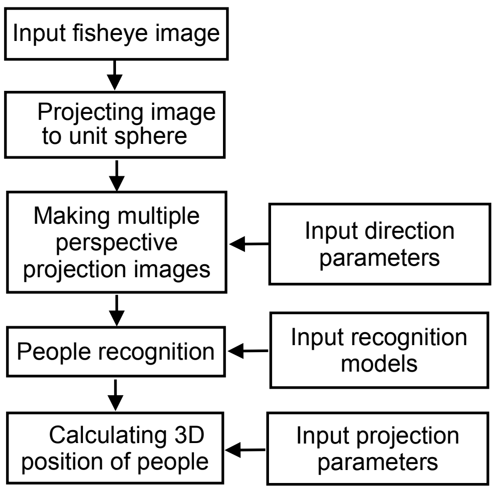


Figure 3. *Flow chart of proposed method*

People recognition are made in the corrected images by machine learning method. Recognition results are used to calculate 3D position of people in workplace. Only very few markers were set on floor of workplace to make calibration between perspective projection images and plane of workplace floor. Detail of the proposed method is described in the following sections.

Fisheye Distortion Correction

In our proposed method, several projection transforms are used to correct fisheye distortion. Figure 1 illustrates an equirectangular projection image example taken by a 360-degree fisheye camera. There is a large fisheye distortion in this image. Figure 2 shows a fisheye distortion corrected image made by our method.

Equirectangular projection

Equirectangular projection is a type of projection for mapping a portion of the surface of a sphere to a flat image. The horizontal coordinate is simply longitude, and the vertical coordinate is simply latitude. Equirectangular images cover 360° horizontally and 180° vertically. Equirectangular projection is the most often used format for 360-degree fisheye camera. To correct distortion, the equirectangular image is projected to a unit sphere.

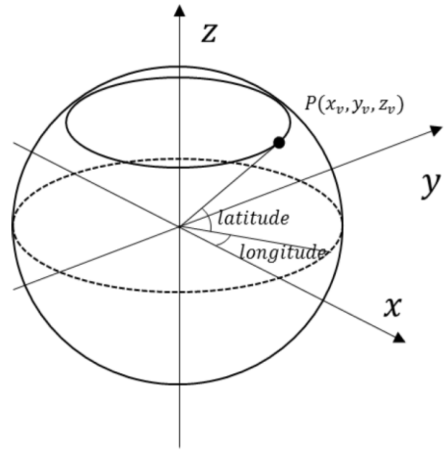


Figure 4. *Unit sphere used for image projection transform*

Unit Sphere Projection

To correct fisheye distortion, equirectangular projection image is projected to a unit sphere as shown in Figure 4. Radius of the sphere is 1. Light input to the fisheye lens which passes point P (x_v, y_v, z_v) on the unit sphere and center of the sphere O. Angle between input light ray and equatorial plane in vertical direction is *latitude* and horizontal direction angle is *longitude*.

Relation between coordinate (x_v, y_v, z_v) and input angles of light ray (*latitude, longitude*) is shown in Equation (1) to Equation (3). 1 pixel on equirectangular projection image has coordinate of (*latitude, longitude*). This image pixel is projected to the point on the unit sphere which has coordinate of (x_v, y_v, z_v).

$$x_v = \cos(latitude)\cos(longitude) \tag{1}$$

$$y_v = \cos(latitude)\sin(longitude) \tag{2}$$

$$z_v = \sin(latitude) \tag{3}$$

Unit Sphere Rotation

As shown in Figure 1, 360-degree fisheye image covers large area of workplace. To cover the area by fisheye distortion corrected images, multiple images are made by using unit sphere of Figure 4. Each distortion corrected image is made at specific direction. The unit sphere is rotated to several directions to make the distortion corrected images. Angle of (α, β, γ) are the rotation angles by which the unit sphere is rotated. α, β and γ are rotation angles around x, y and z axis. Coordinate (x, y, z) of point on the sphere is transfer to (x', y', z') by Equation (7). Coordinate transforms are made by Equation (4) to Equation (6) separately. Combined rotation by angles (α, β, γ) is made by Equation (7). By rotating the unit sphere to a specific direction by angles (α, β, γ), fisheye distortion corrected image can be made. Rotating the unit sphere several times, multi fisheye distortion corrected images can be made to cover whole area of workplace.

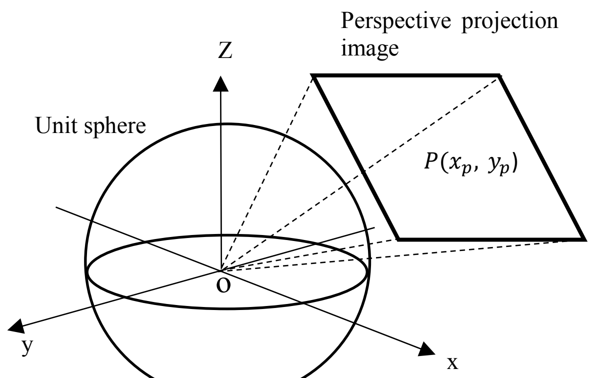


***Figure 5.** Making perspective projection image from unit sphere*

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\alpha & -sin\alpha \\ 0 & sin\alpha & cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_x(\alpha) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (4)$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_y(\beta) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} cos\gamma & -sin\gamma & 0 \\ sin\gamma & cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_z(\gamma) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (6)$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R_x(\alpha) R_y(\beta) R_z(\gamma) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (7)$$

Perspective Projection

As shown in Figure 5, perspective image is made by projecting image from the unit sphere to the perspective image plane. One point of $P(x_p, y_p)$ on the perspective image plane has relation with input light ray angles (*latitude, longitude*), which has relation expressed by Equation (8) and Equation (9). The perspective image plane is tangent plane to the unit sphere.

$$longitude = \arctan(\frac{y_p}{x_p}) \quad (8)$$

$$latitude = \arctan(\frac{z_p}{\sqrt{x_p{}^2 + y_p{}^2}}) \quad (9)$$

Making Distortion Correction Image

The fisheye distortion correction image is made by several projections described above. Intensity of image pixel on the perspective image plane at position $P(x_p, y_p)$ is calculated from equirectangular projection image pixel of coordinate (*latitude, longitude*). At first, coordinates (x_p, y_p) is mapping to a point of

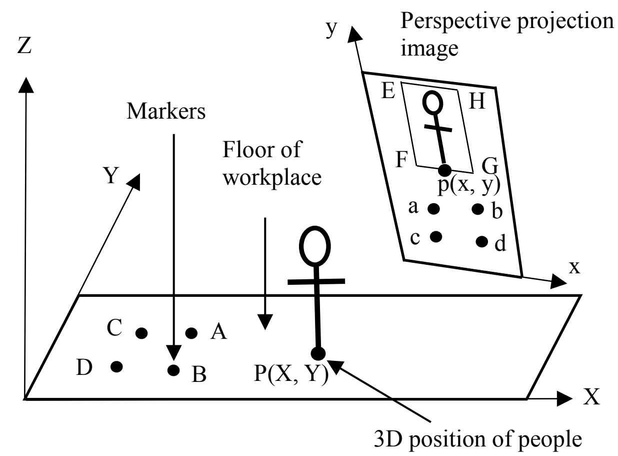


***Figure 6.** 3-dimension position measurement of worker*

(x', y', z') on the unit sphere by Equation (1) to Equation (3), Equations (8) and Equation (9). Then coordinates of (x, y, z) on the sphere before rotation of the unit sphere is calculated by Equation (4) to Equation (7). Rotation angle of (α, β, γ) are parameters set before calculation. At last, coordinates of (*latitude, longitude*) on equirectangular projection image are calculated by point of (x, y, z) on the unit sphere, using Equation (1) to Equation (3). Coordinates of (*latitude, longitude*) may not be integer. Bi-linear interpolation is used to calculate pixel intensity on equirectangular projection images. To cover whole work area, several perspective images are made.

3 Dimension Position Measurement

In our proposed method, as shown in Figure 6, 3-dimension position of worker on floor plane of workplace is measured. Position at point P(X, Y), the point on the floor plane, is used as position of a worker in workplace. It is supposed that a worker stands on the floor of workplace. Position of his foot is used as his 3D position in the workplace. In the perspective projection image, the worker's 3D position is imaged as point p(x, y). By calibrating perspective plane with 3D plane of workplace floor, we can calculate 3D position P(X, Y) by point p(x, y). Because both floor plane of workplace and perspective projection image are planes, which can be expressed by perspective projection of Equation (10). Parameters of m_{ij} can be calculated by calibration. By setting several markers on the floor of workplace plane, the parameters m_{ij} can be calibrated. 4 points can be used for calculated transform parameters between one perspective image and the floor plane of the workplace. A, B, C, D illustrate markers on the floor plane. Points a, b, c and d on perspective image plane are images of points A, B, C and D. Because positions of A, B, C, D, a, b, c, d are known value, 8 equations can be made to get 8 unknown value of m_{ij}. We project several perspective images to cover whole workplace. Each perspective images are calibrated by marker on the floor plane. Then we can calculate 3D position of a worker by his position on the perspective image.

$$\begin{pmatrix} X \\ Y \\ u \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (10)$$

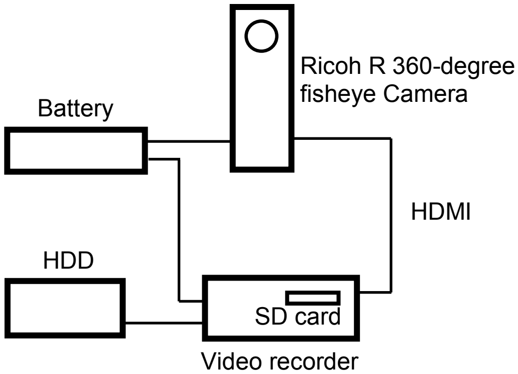


Figure 7. *Experiment setup*

People recognition and Position Measurement

As shown in Figure 6, people recognition is made on perspective projection image plane. Histograms of Oriented Gradients(HOG) feature and Support Vector Machine(SVM) machine learning are used to recognize people in our proposed method [1]. A pre-trained model was used for people recognition. People recognition results is expressed by a rectangle of EFGH. Position at the middle of line FG, bottom line of the result rectangle, is used as position of the worker on perspective image. By using the calibrated parameters of m_{ij} and Equation (10), 3D position P(X, Y) can be calculated by position p(x, y).

Experiment Results

Ricoh R 360-degree fisheye cameras as shown in Figure 8 are used to make people recognition and 3-dimension position measurement experiments. Experiment setup is illustrated in Figure 7. Video stream is outputted from Ricoh R camera to video recorder by HDMI cable. Video data is recorded in the recorder by SD card or USB Hard disk (HDD). Power supply to the camera and video recorder is given by a mobile battery. The Ricoh R fisheye cameras are attached to wall or pillar in the workplace.

Output fisheye image of Ricoh R is equirectangular projection image which has size of 1920 by 1080 pixels. One example of the fisheye camera is shown in Figure 1. 3 direction perspective projection images were made to correct fisheye distortion. Angle of view 30-degree, 90-degree and 30-degree were used for making left, front and right direction perspective projection images separately. Size of each perspective image is 640x480 pixels. Figure 9 is people recognition results. Figure 9 (a) is left direction perspective image and people recognition results. People recognition results are expressed by green rectangle. Figure 9 (b) is right direction perspective image and people recognition results. In Figure 10 red points are foot points of people in perspective image, which are used for 3D position measurement.

Recognition evaluation results of recognition rate and false positive rate are shown in Table 1. 33,298 images were used for evaluation. Peoples' 3D position measurement results are shown in Figure 11. Red points are 3D position of people, which are results of 290 frames. X and Y are distance in longitude direction and vertical direction by unit of millimeter. Distance measurement accuracy was 6.6%.

Figure 8. *Ricoh R 360-degree fisheye camera*

(a)

(b)

Figure 9. *People recognition results on perspective projection images (a)Left direction people recognition results, (b)Right direction people recognition results.*

Table 1 People recognition evaluation results

	People recognition results
Recognition rate (%)	92.5
False positive rate (%)	0.2

Figure 10. *People recognition results example. Red points illustrate people position in perspective projection image*

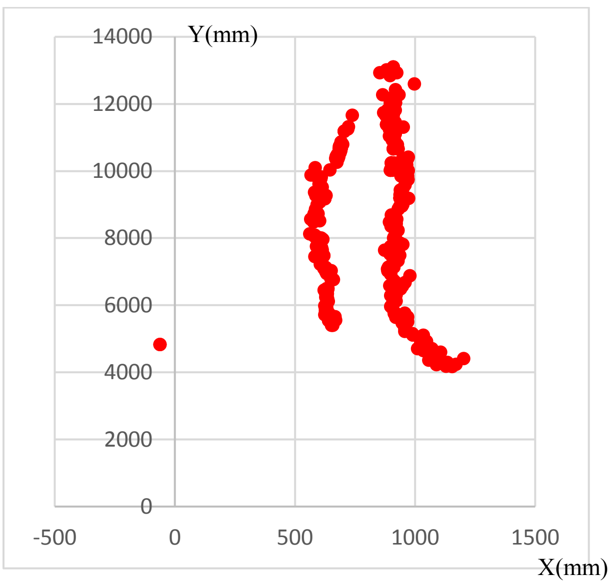


Figure 11. *3D position of people measurement results. People's 3D position calculated from 290 fisheye image frames are printed at the same graph by red points.*

Conclusion

We propose a new method to recognize people and measure 3D position of the people in workplace by only one fisheye camera. One 360-degree fisheye image is projected to a unit sphere and several perspective projection images are generated to correct fisheye distortion. People recognition is made in the corrected images by machine learning method. Recognition results are used to calculate 3D position of people in workplace. Only very few markers are set on floor of workplace to make calibration between perspective projection images and plane of workplace floor.

People recognition and position measurement experiments were made in workplace by Ricoh R 360-degree monocular fisheye cameras. People recognition rate of 92.5%, false positive rate of 0.2%, people position measurement accuracy of 6.6% were obtained. The evaluation results demonstrate effectiveness of our proposed method.

References

[1] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893, 2005.

[2] P. Dollar, C. Wojek, B. Schiele, P. Perona, "Pedestrian Detection: A Benchmark" 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp.304-311, 2009

[3] A. Angelova, A. Krizhevsky, V. Vanhoucke, A. Ogale, D. Ferguson, "Real-Time Pedestrian Detection with Deep Network Cascades", Proceedings of the British Machine Vision Conference (BMVC), pp. 32.1-32.12. BMVA Press, 2015.

[4] D. Ribeiro, A. Mateus, P. Miraldo, and J. C. Nascimento, "A Real-Time Deep Learning Pedestrian Detector for Robot Navigation", 2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), pp. 165-171, 2017

[5] Y. Tian, P. Luo, X. Wang, X. Tang , "Deep Learning Strong Parts for Pedestrian Detection", 2015 IEEE International Conference on Computer Vision (ICCV), pp.1904-1912, 2015

[6] D.Xu, W. Ouyang, E. Ricci, X. Wang, N. Sebe, "Learning Cross-Modal Deep Representations for Robust Pedestrian Detection", IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 5363-5371, 2017

[7] S. Nedevschi, S. Bota, C. Tomiuc, "Stereo-Based Pedestrian Detection for Collision-Avoidance Applications", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 10, NO. 3, pp.380-391, 2009

[8] C. G. Keller, M. Enzweiler, D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection", 2011 IEEE Intelligent Vehicles Symposium (IV), pp.691-696, 2011

[9] J. Ahn, "Calibration of stereo fish-eye camera for Driver Assistance Systems", 25th International Conference of Image and Vision Computing New Zealand, pp. 1-5, 2010

[10] K. V. Beeck, T. Goedeme, "Real-time accurate pedestrian detection and tracking in challenging surveillance videos", Proceedings of the 10th International Conference on Computer Vision Theory and Applications, Vol.3, pp.325-334, 2015

[11] K. V. Beeck, T. Goedeme, T. Tuytelaars, "A warping window approach to real-time vision-based pedestrian detection in a truck's blind spot zone", Proceedings of the ninth international conference on informatics in control, automation and robotics, Vol.2 pp. 561-568, 2012

[12] S. Silberstein, D. Levi, V. Kogan and R. Gazit, "Vision-Based Pedestrian Detection for Rear-View Cameras", 2014 IEEE Intelligent Vehicles Symposium, pp. 853-860, 2014

[13] M. T. Bui, V. Fremont, D. Boukerroui, P. Letort, "People Detection in Heavy Machines Applications", International Conference on Cybernetics and Intelligent System & Robotics, Automation and Mechatronics, pp.18-23, 2013

[14] Z. Zhang, P. L. Venetianer, A. J. Lipton, "A Robust Human Detection and Tracking System Using a Human-Model-Based Camera Calibration", The Eighth International Workshop on Visual Surveillance, pp.1643-1653, 2008

[15] M. Demirkus, L. Wang, M. Eschey, H. Kaestle and F. Galasso, “People Detection in Fish-eye Top-views”, In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017), pp.141-148, 2017

Author Biography

Haike Guan received his PhD in information processing from Tokyo Institute of Technology (1995). Since then he has worked in Ricoh Company LTD. and currently is a senior specialist researcher in Ricoh Institute of Information and Communication Technology, R&D Division. His work has focused on research and development in image processing, pattern recognition, computer vision, human activity recognition and related fields.

Makoto Shinnishi received his MS in Media and Governance from Keio University (2000). Since then he has worked in NTT (Nippon Telegraph and Telephone Corporation) as a researcher. He joined Ricoh Company LTD. in 2003 and currently is a specialist researcher. His work has focused on research and development in IoT, human computer interaction and related fields.

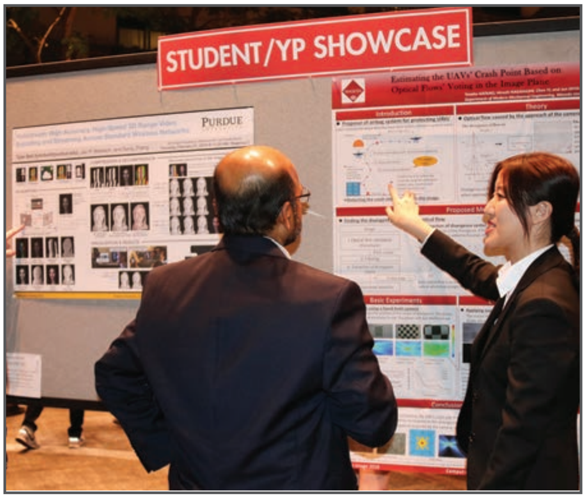
JOIN US AT THE NEXT EI!

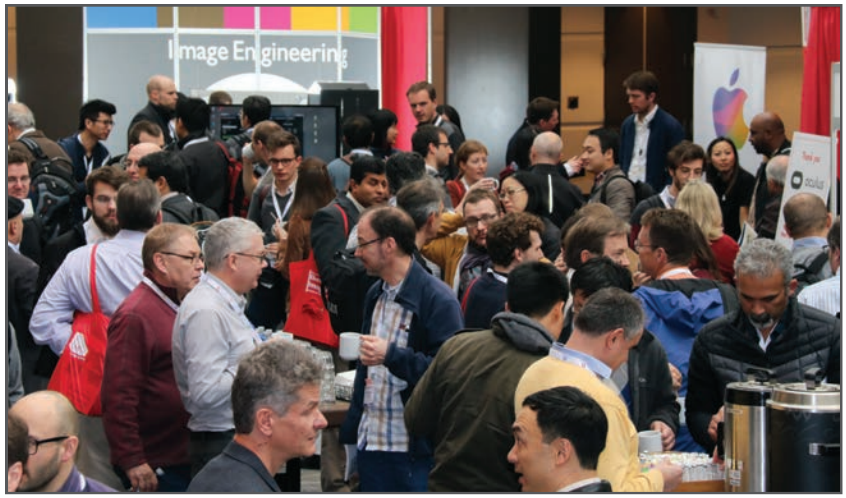
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org