https://doi.org/10.2352/ISSN.2470-1173.2019.7.IRIACV-462
© 2019, Society for Imaging Science and Technology

Application of Semantic Segmentation for an Autonomous Rail Tamping Assistance System

Gerald Zauner; University of Applied Sciences Upper Austria, Wels / Austria and Plasser & Theurer Linz / Austria
Tobias Mueller, Andreas Theiss, Martin Buerger, Florian Auer; Plasser & Theurer Linz / Austria

Abstract

Safe and comfortable travel on the train is only possible on tracks that are in the correct geometric position. For this reason, track tamping machines are used worldwide that carry out this important track maintenance task. Turnout-ta.mping refers to a complex procedure for the improvement and stabilization of the track situation in turnouts, which is carried out usually by experienced operators. This application paper describes the current state of development of a 3D laser line scanner-based sensor system for a new turnout-tamping assistance system, which is able to support and relieve the operator in complex tamping areas. A central task in this context is digital image processing, which carries out so-called semantic segmentation (based on deep learning algorithms) on the basis of 3D scanner data in order to detect essential and critical rail areas fully automatically.

Introduction

Tamping Process

When a train drives along the railway, it generates enormous forces. The entire track consisting of rails, sleepers and ballast is an elastic system that deforms and then returns to its original position. In the end, this high load leads to a deterioration of the track geometry. This can lead to anomalies, because of which the ideal geometry of the track can no longer be guaranteed. In these areas, for example, temporary speed restrictions must be imposed. To avoid such a situation, tracks have to be maintained at regular intervals. This ensures that the ideal geometry of the track is restored. In this context, the so-called 'track tamping' represents the most common maintenance task on railway tracks.

Figure.1: A typical tamping machine during work.

Lining refers to correcting the horizontal and vertical alignment of the track, and lifting to the compaction and displacement of the substructure with complete removal of cavities under the sleepers. The combined lifting-lining unit works with a measuring system, gripping the track, raising the track to a predetermined height, correcting for vertical misalignment and simultaneously pivoting the track to correct horizontal alignment (= simultaneous leveling and alignment). Subsequently, the tamping units are lowered and the tamping tines dip into the ballast. The tamping unit vibrates to fluidize the ballast so that it can rearrange and settle in a dense matrix. Controlled vibration reduces the force required to insert the tamping tines into the ballast without damaging or crushing the ballast stones. A special cylinder arrangement additionally exerts a force on the tamping arms, which results in an additional movement of the tamping tines (squeezing). The tamping compactifies then the ballast below the sleeper, i.e. into the cavity created by the lifting process. Thereafter, the tamping machine moves forward to the next sleeper and the process is repeated. Finally, behind the tamping machine, the result is a track at the correct geometric level, on a homogeneous ballast bed and with restored elasticity (Fig. 2). [1]

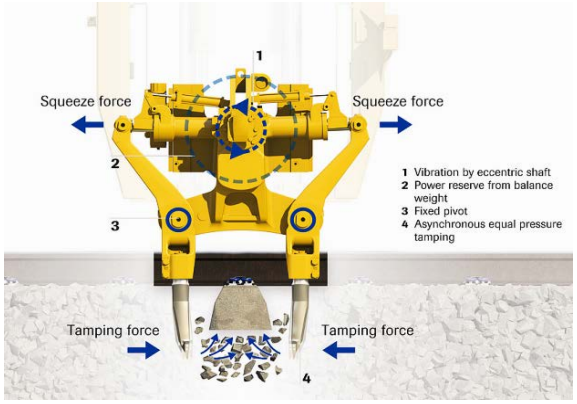


Figure 2: Tamping unit used for improving the rail track quality (base construction).

Autonomous Tamping

To date, many years of experience and skill of the operator are necessary to achieve a high-quality tamping result in turnouts. The tamping process, especially in turnout areas, is characterized by a large number of manual interventions. High working speeds, combined with the highest possible quality of work, are only achieved by an experienced or well-trained operating staff. For this reason, extensive knowledge in various specialist areas (track surveying, mechanical engineering, superstructure technology, regulations) for the ideal tamping process is necessary. Among the most important tasks of the two operators are: control of the tamping unit including rotation and pitch spreading or pivoting, control of the lifting and lining unit and operation of additional lifting in the area of the diverging track, etc.

The purpose of the turnout-tamping assistant is to develop an automatic assistance system comparable to level 3 of the SAE J3016

standard (which was originally defined to characterize the autonomous driving of road-bound motor vehicles). Generally, the focus is on the automated support of tamping in difficult environments such as switches and crossings (but not restricted to). At this level of automation, the system creates action recommendations that the operator can confirm prior to each action. The aim is to relieve the operator, to increase the working speed and to stabilize the quality of work at a consistently high level. Basically, the tamping assistance system is also suitable for higher degrees of autonomy. [2] [3]

Deep Learning in Rail Track Maintenance

To our knowledge, this is the first published attempt of a semantic segmentation based on deep learning with 3D laser scanner depth images in the field of automated rail tamping. The literature describes various deep learning approaches based on the analysis of color images (RGB images). In most cases, the methods described there treat applications in the field of rail inspection or rail track infrastructure management (e.g. [4] [5] [6] [7] [8] [9]).

Relevant object information from 3D scanner image data

The environment (i.e., mainly the superstructure directly in front of the tamping machine) is scanned with a 3D line scanner mounted on the tamping machine roof (Figure 3). The individual scans deliver accurate 3D data, which is computationally merged with information such as the position or timing of the recording. At the same time, the position of the machine and each individual unit are constantly updated. In this way, position data of all relevant machine parts are fused in a 3D overall model. This then forms the basis for every further control decision during operation.

Figure 3: Tamping machine with a roof-mounted 3D laser scanner.

The central task of digital image processing is now the exact detection of the relevant infrastructure components within the working area of the machine (such as ballast, sleepers, rails, fasteners, etc.) For example, the hydraulic tamping tines must only penetrate into suitable ballast areas, but they must not hit rails or sleepers (which could then be seriously damaged). On the other hand, non-critical areas (such as plants or the like) should not interrupt the work process by indicating a pseudo obstacle.

The decision was to apply state-of-the-art deep learning (DL) methods to this demanding image analysis problem, as it has proven to be extremely robust in many other areas (such as autonomous car driving) and far surpasses other traditional image-based segmentation approaches.

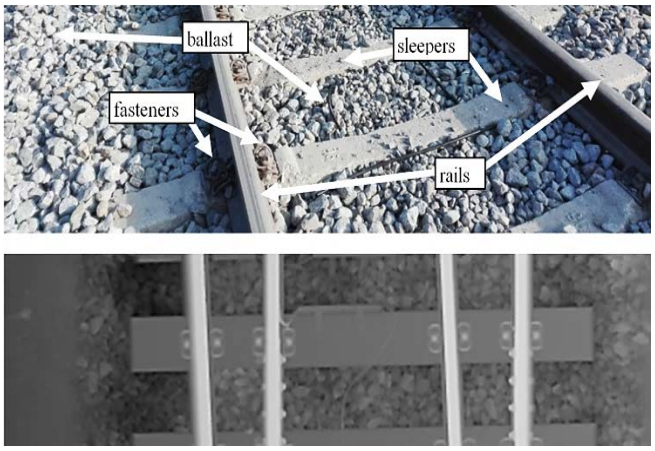


Figure 4: Picture of the relevant track area (top); typical 3D scan depth image (bottom).

The system should be flexible enough to be easily adapted to completely new scenarios. It should e.g. cope with unknown forms of vegetation or different track infrastructures (characterized by a wide range of variants worldwide) through a simple learning process that can be handled even by non-expert staff. Traditional methods can work quite well, but they often require a high level of expertise and very specific domain knowledge to create hand-crafted features. In contrast, DL approaches learn from the data itself, which means the expertise for feature engineering is replaced (partially or completely) by the DL and in some application cases can outperform humans and human-coded features. [11-14]

3D Depth Image Acquisition by Laser Scanner

The image analysis process described here uses only 3D depth images as input data (see Fig. 4, right). The different gray values correspond to different distances to the sensor (i.e., the brighter the image pixels, the closer). The advantage of using depth images is that the appearance of the objects does not vary (for example due to different light conditions), as would be the case with standard RGB cameras, although the 3D scanner system is of course more expensive compared to RGB cameras. However, since the scanner is also used for various other purposes during machine operation, this disadvantage is not significant in the concrete application.

The 3D depth images are provided by a rotating 3D laser scanner. The scanner itself delivers single line scans with millimeter depth accuracy, which are then continuously merged into a depth image with a typical resolution of approximately 4000 x 1000 pixels. The working speed during tamping is approx. 1000m/h, which leads to a lateral scan resolution of approx. 2mm. This is sufficient to create detailed scan images that also allow visualization of small objects (such as fasteners, etc.). The scanner head is mounted directly in front of the train whereas the actual tamping unit is located approximately in the middle of the machine. Thus, due to moving of the vehicle, there is a small time offset between the scanning of a certain region and the actual tamping process at this particular position, which provides a time window of about 10 seconds for all necessary data processing tasks. Additionally, the raw line scans have to be geometrically corrected as the scanning laser spot moves in a helix-like trajectory along the railway tracks (Fig. 5). This correction is of course speed-dependent.

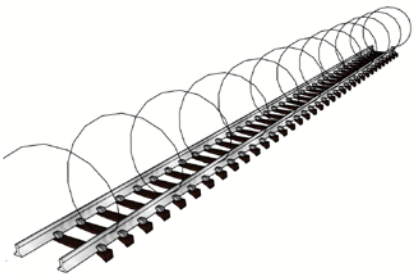


Figure 5: Helix-like trajectory of the laser scanning spot while driving.

Semantic Segmentation of 3D-Scanner Data

Fully Convolutional Neural Networks for Image Segmentation

An extension of image classification is the so called semantic segmentation. It generally plays a crucial role in computer vision and enables a computer to not only recognize objects in images, but also to locate them pixel-exactly. The recognition and exact delineation of objects in the image is achieved by the classification of each individual pixel, i.e. each pixel is assigned a defined object class (Fig. 6). Our original segmentation approach is based on a Fully Convolutional Network (FCN) [13], a popular algorithm for semantic segmentation.

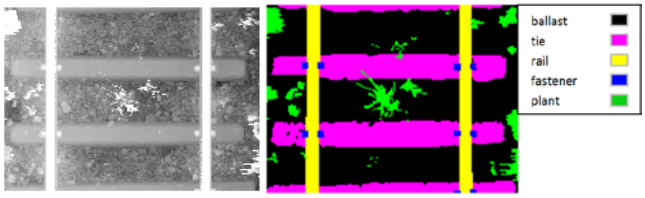


Figure 6: 3D scanner image (left). Desired result image (right) with pixel-exactly segmented areas representing the relevant image contents such as ballast, rails, plants, etc.

The idea behind an FCN is to extend a classic Convolutional Neural Network (CNN), which is commonly used for image classification, by replacing the fully connected layer with a *1x1xn* convolution layer and adding a convolution transpose layer (Fig. 7). In detail, this network model uses various convolution and pooling layers to analyze an image and reduce it to a fraction of its original size (usually 1/32). This compressed intermediate result contains a class prediction at this resolution level. Finally, it uses so-called convolution transpose layers to rescale the image to its original dimensions. By extending a CNN to an FCN, the classification part is replaced by a convolution part. When creating a larger image, the net now not only gives probabilities for each trained class, but a whole array of such probabilities. An element of this array then stands for the recognized class of a subarea of the input image. After scaling up this array to the original size of the input image using the Convolution-Transpose layer, one finally obtains a heat map of the local distribution of the detected classes, since each pixel in this result image corresponds to a classified feature vector.

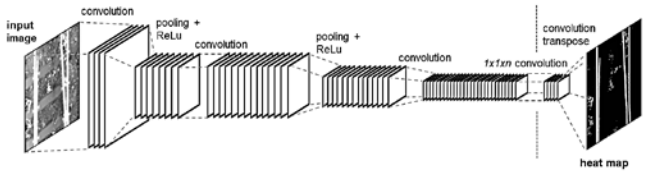


Figure 7: Basic structure of a fully convolutional neural network (FCN) for semantic image segmentation.

Image annotation

A general challenge in deep learning is the large amount of learning data needed to produce good results i.e. enough annotated images must be provided to train the network. Especially, in semantic segmentation the according effort is considerable, since a pixel-precise marking of the image objects is necessary. For this reason, a special annotation tool was developed which supports the user in this time consuming work by semi-automatic marking functions. As the depth images considered are very often characterized by homogeneous image areas (like rails, sleepers, etc.) and not so much by rich texture, the supporting tool offers intuitive region labelling functionalities with a dedicated focus on processing this specific type of images (e.g. watershed algorithms, super-pixel algorithms or contour tracking algorithms, Fig. 8).

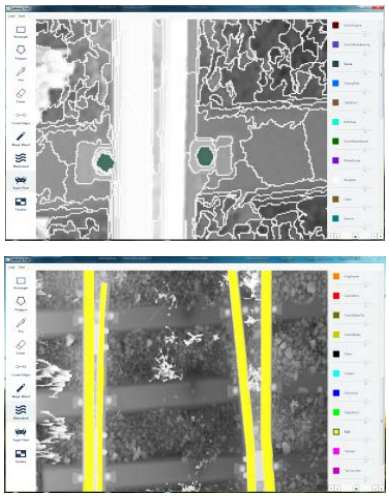


Figure 8: Image annotation software tool for generating ground truth segmentation images.

Our experiments have also shown that in addition to segmented full-frame images, small (jittered) image patches are also appropriate for training, since the FCN architecture allows input data in various image sizes. This makes it possible to generate a large number of additional training image patches from just a single acquired depth image (Fig. 9). One more challenge is the fact that we have to deal with a very imbalanced dataset. The images typically consist of large areas of ballast structures whereas only very few pixels represent objects like rail screws.

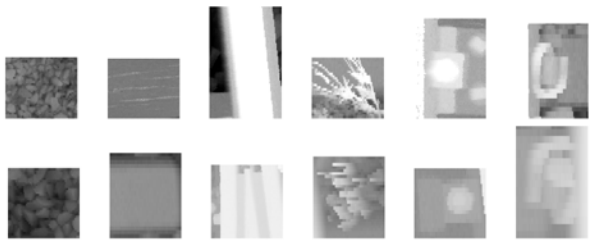


Figure 9: Typical image patches used for learning (ballast, sleeper, rail, plant, fasteners, clamps).

Besides real world data we also used artificially generated depth images from a virtual simulation environment (this simulator was originally intended for machine operator training purposes thus providing very realistic 3D scenarios), see Fig. 10. Thus, no manual labelling of the images was necessary and we were able to provide large quantities of images very quickly (and we could even vary image structures specifically, such as different gravel sizes, etc.). On the other hand, with the help of the simulator, we were also able to intensively test and improve the functionality of the entire assistance system (digital twin).

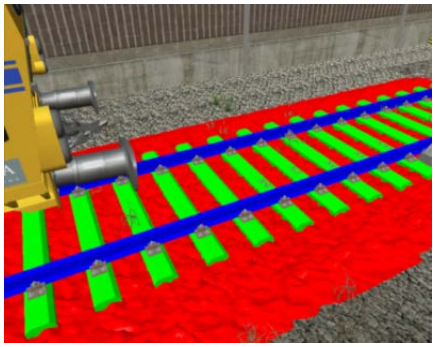


Figure 10: Automatically labelled images provided by a 3D simulator. This software was originally intended for machine operator training purposes thus providing very realistic scenarios.

Results

The hardware setup used is as follows:

- Intel® Xeon® CPU E5-1630 v4 @ 3.70GHz 3.70GHz
- 128GB RAM
- 2x GTX 1080Ti (not cooperative, but for simultaneous trainings)

We have experimented with different deep learning frameworks (like MatConvNet, DIGITS and TensorFlow) and various network architectures. However, the most time consuming part was the tuning of the hyper parameters (learning rate, batch-size, training epochs, etc.) to minimize overfitting and loss in the validation dataset. Fig. 11 shows the typical evolution of a segmentation result during learning (shown are the results at epochs 5, 20, 40, 100 and 500).

For training we have generated about 11.500 image patches in the size of 256x256 from about 350 labelled full-size depth images (500x4000). The training time was typically 22 minutes per epoch over about 50 epochs. The inference time in the application software is currently 0.8 sec for a 1000x4000 image (on a GTX 1080Ti graphics card). Increasing the number of training images clearly improved the segmentation results considerably. Another observation was that retrospectively adding a new class (that is visually very similar to an existing class) has no effect on the recognition performance of the original classes within an existing network.

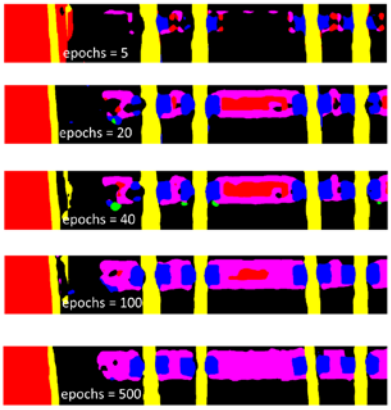


Figure 11: Improvement of the segmentation result during learning (ballast - black, track - yellow, screw - blue, ties - pink, red - unknown).

Finally, from the segmentation result relevant informations for the tamping process are derived, e.g. exact ballast areas (into which the tamping tines can penetrate) but also sleeper positions and orientations, which are important for correct control of the tamping units. Additionally, the beginning and ending of turnout sections are identified automatically. Also, special equipment along the rail track (like switch rods, etc.) can be identified robustly.

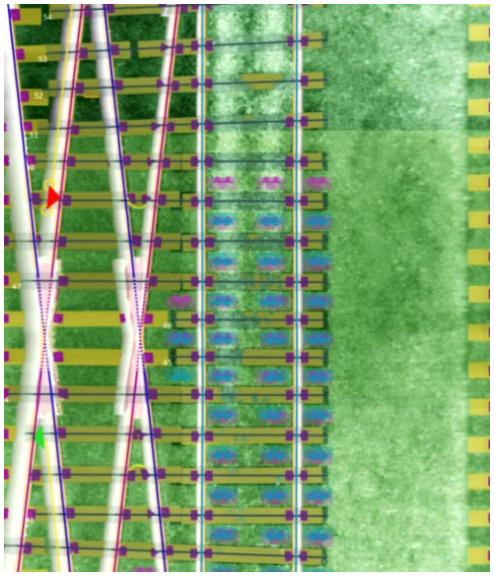


Figure 12: Typical result of the segmentation process. Different segmented classes such as ballast (green), sleepers (yellow), tracks (white), screws (purple) and plants (green) are shown as a half transparent overlay on the depth image. Also shown are automatically derived informations like sleeper orientations or the beginning and ending of turnout areas.

Summary and Outlook

In this article, an image processing system for a semi-autonomous tamping assistance system was presented. Based on deep learning algorithms, it is able to localize and classify relevant infrastructure in the working area of a tamping machine. A 3D laser scanner is used, which enables accurate and robust scanning of the machine environment, regardless of the particular lighting situation (day, night, fog, etc.). First tests in real operation confirm the excellent suitability of the method described. In conclusion, deep learning based semantic segmentation enables the practical realization of very robust machine vision solutions without which robust outdoor applications would not be possible especially under very harsh conditions as those described above. The algorithms used are constantly being improved - for example, a new generation of network architecture with an improved segmentation approach is currently being worked on which promises to further improve the detection properties even for very small or thin objects (such as cables, etc.). Active learning approaches will further improve the results.

Acknowledgements

We thank our colleagues from the department *Digital Track Systems* (Plasser & Theurer) for their support.

References

[1] Homepage Plasser & Theuer: https://www.plassertheurer.com/en/machines-systems/tamping.html

[2] F. Auer et al.: Smart Tamping – Anwendungsmöglichkeiten des Weichenstopf-Assistenzsystems, ZEVrail 2018, Nr. 6.

[3] M. Bürger: Hilfe beim Stopfen – Entwicklung eines Weichenstopf-Assistenzsystems, Eisenbahn-Ingenieur 2017, Nr. 6

[4] Xavier Gibert et al.: Deep Multi-task Learning for Railway Track Inspection, IEEE Transactions on Intelligent Transportation Systems, Volume: 18, Issue: 1, Jan. 2017, doi: 10.1109/TITS.2016.2568758

[5] Arun Kumar Singh et al.: Vision based rail track extraction and monitoring through drone imagery, KICS – ICT Express, Dec. 2017, doi: 10.1016/j.icte.2017.11.010

[6] Junwen Chen et al.: High-speed railway catenary components detection using the cascaded convolutional neural networks, 2017 IEEE International Conference on Imaging Systems and Techniques (IST); Electronic ISBN: 978-1-5386-1620-8

[7] Shahrzad Faghih-Roohi et al.: Deep convolutional neural networks for detection of rail surface defects, 2016 International Joint Conference on Neural Networks (IJCNN); Electronic ISBN: 978-1-5090-0620-5

[8] Ye T et al: Automatic Railway Traffic Object Detection System Using Feature Fusion Refine Neural Network under Shunting Mode, Sensors (Basel), 2018 Jun 12-18(6), doi: 10.3390/s18061916.

[9] Yan Li, Xiukun Wei: Pantograph Slide Plate Abrasion Detection Based on Deep Learning Network, 3rd International Conference on Electrical and Information Technologies for Rail Transportation (EITRT) 2017

[10] Bichen Wu, Alvin Wan, et al.: SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud, arxiv: 1710.07368

[11] Alex Krizhevsky et al.: ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25 (NIPS 2012)

[12] Karen Simonyan et al.: Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv technical report, 2014 (conference paper presented at ICLR 2015)

[13] Christian Szegedy et al.: Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Electronic ISBN: 978-1-4673-6964-0, doi: 10.1109/CVPR.2015.7298594

[14] Jonathan Long et al.: Fully Convolutional Networks for Semantic Segmentation, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Electronic ISBN: 978-1-4673-6964-0

Author Biography

Gerald Zauner received his PhD in applied physics from Vienna University of Technology (2005). Since then he has worked at the University of Applied Sciences Upper Austria (School of Engineering) where he is professor for signal processing (2015). He also works for Plasser & Theurer in the research department. His work has focused on industrial machine vision, optical metrology and non-destructive testing.

Tobias Müller studied automation engineering at the University of Applied Sciences Upper Austria. Since 2018, he works at Plasser & Theurer in the department 'Digital Track Systems' with focus on deep learning in image segmentation.

Andreas Theiss studied software engineering at University of Applied Sciences Hagenberg. From 2012 to 2016 he worked as software engineer with focus on time-critical applications. He joined Plasser & Theurer in 2016 where he is developing assistance systems for rail maintenance machines.

Martin Bürger studied computer science at the Johannes Kepler University Linz (1998). He was software project manager at Siemens AG (2000) and Software quality representative at Magna Powertrain Engineering Center (2009). Since 2012, he is with Plasser & Theurer in Research & Development and head of the department 'Digital Track Systems' (2018).

Florian Auer studied a degree course in Civil Engineering at Graz Technical University, specializing in transport engineering (PhD degree in 2010). Between 2002 and 2012 he worked in various positions at OEBB Infrastructure AG. He joined Plasser & Theurer in 2012. Since 2017, he has been Director of Innovation and Technology.

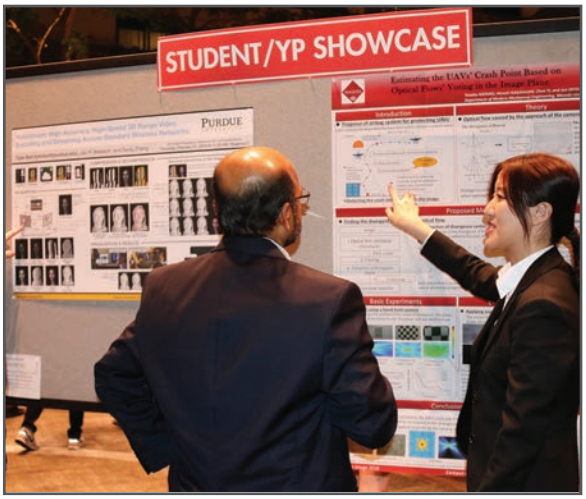
JOIN US AT THE NEXT EI!

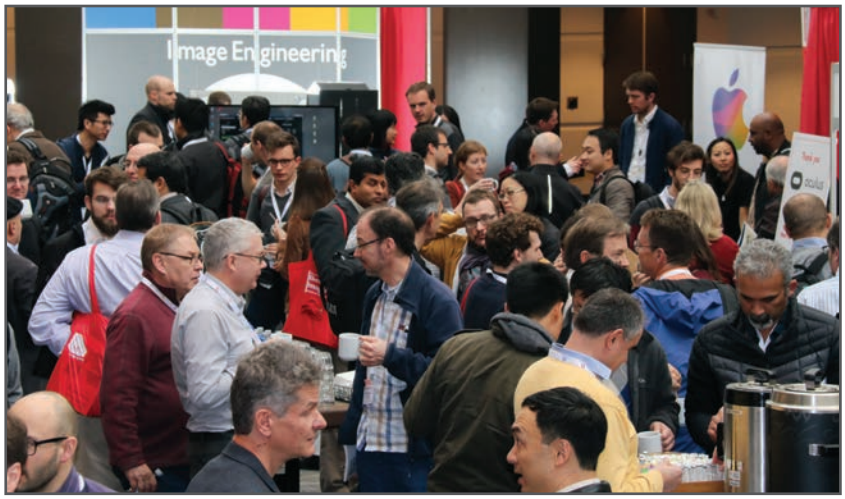
IS&T International Symposium on

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