Research Statement

Nathan Jacobs
November 23, 2021

People often think about computer vision from the perspective, "Given an image, understand something about it." However, this ignores contextual information that is often available with an image. My work explores ways to improve image-understanding systems by incorporating such context.

Theme 1: Relating Appearance, Time, and Location

The appearance of an outdoor scene changes for many reasons, including the time of day, the weather, and the turning of seasons. My view is that these changes should be explicitly modeled. I often refer to this as geo-temporal appearance modeling.

Much of my work has focused on developing algorithms that extract temporal patterns in sequences of images from static outdoor cameras. This work began when, as a graduate student, I helped build the *Archive of Many Outdoor Scenes* (AMOS) [1], which today contains over a billion images from thousands of outdoor webcams. I have used this imagery to: create models of scene appearance variations [1]; develop novel algorithms for camera localization [2] and video surveillance [3, 4]; evaluate novel methods for estimating and mapping weather conditions [5, 6, 7]; and measure leaf growth on trees [8]. I also developed algorithms that use cloud motion to estimate scene geometry and camera calibration [9, 10, 11, 12].

My recent work has explored how additional sources of data can help improve our geo-temporal appearance models. We have now used many different sources, including: ground-level panoramas; vector GIS data; overhead imagery (satellite or aerial); airborne and ground-based LiDAR, social media imagery and text; weather reports; and field audio recordings.

Cross-View Visual Appearance Mapping Social media imagery, which often comes geotagged and timestamped, provides billions of high-resolution samples of the world. We have developed approaches that predict features extracted from social media imagery using overhead imagery as input, what we call cross-view mapping [13]. We have used this idea for mapping scene categories [14, 15], scenicness or natural beauty [16], soundscapes [17], object distributions [18], and land use [19]. We have also used it to learn useful features for image localization [20] and remote sensing [21]. Our most recent work extends cross-view mapping to include dependence on when an image was captured [22]. We show how the trained model can be used for the image forensics task of metadata verification (i.e., does the scene appearance seem plausible for the purported location and time?) and shown how to improve the performance on this task using discriminative training [23].

Image Geo-Localization Knowing where an image was captured makes the information extracted from it more useful, but the location isn't always available. Given this, estimating the location from imagery has become an important task. My early work focused on estimating the location of a webcam using natural appearance variations across many images [2, 24]. We have also developed methods for single-image localization using either overhead [14, 15] or ground-level imagery [25] as reference data. We extended our work with overhead imagery to the problem of simultaneous orientation and location estimation [21], proposing a novel architecture that learns the spatial transformation between the overhead and ground-level viewpoint. A side benefit is that the network also learns to transfer semantics from ground-level imagery to overhead imagery. We also proposed a novel localization architecture [20] which enables simultaneously solving the problems of 1) estimating image capture location (with and without known capture time) and

2) image capture time (with and without known capture location). More recently, we showed how our dynamic visual appearance mapping method [22], which we also used for image forensics, results in significant improvements for localization. Most of this work has focused on coarse localization, but we have also begun developing novel approaches for fine-grained localization [26]. Our focus has been on developing non-blackbox methods for absolute pose regression, which have the speed/simplicity of neural network approaches but many of the interpretability advantages of classical approaches.

Camera Calibration Knowing the direction a given image pixel is viewing is also important, but this requires camera calibration, which is often not available. My early work focused on methods suitable for webcams, when it is not feasible to physically access or manipulate the camera. I proposed an innovative approach that uses the appearance of the sky on a clear day to determine the geo-orientation of the camera [27]. This was one of the first works to use a sky appearance model developed for graphics applications to aid in outdoor scene understanding, now common practice [28, 29, 30]. I proposed a method for estimating the absolute camera orientation and focal length that uses cloud motion [31]. I demonstrated how a solar refractive phenomenon (i.e., a rainbow) can be used to estimate the focal length of the camera and, given a video of a rainbow, provide a constraint on the geographic location of the camera [32].

More recently, we have also developed algorithms for (rough) single-image calibration, including estimating focal length [33] and horizon lines [34, 35]. As part of this, we introduced the *Horizon Lines in the Wild* dataset, which remains one of the standard evaluation datasets for single-image horizon-line estimation.

Data-Driven View Synthesis The problem of view synthesis, estimating the appearance of a scene from a new viewpoint, is a central task in geometric computer vision. Our first work on this problem was in predicting the appearance of a ground-level panorama from an overhead image [21], an extreme viewpoint shift. This led to work on synthesizing an overhead view from imagery captured by an autonomous vehicle [36]. Both of these problem settings have since received significant attention from the community. More recently, we are developing methods for less extreme viewpoint shifts, such as between pairs of panoramas. Our first work in this direction, Generative Appearance Flow [37], performs as well as state-of-the-art approaches but doesn't require known depth at training time. We have since extended this to support synthesizing across illumination conditions and seasons [38]. Our upcoming work will explore alternative architectures and fusing of multiple modalities.

Theme 2: Remote Sensing

I have also been working directly on problems in remote sensing. Our first work in this domain proposed a novel method for semantic segmentation of overhead imagery that incorporates features from ground-level imagery [39], addressing the task of fine-grained building labeling. We have a series of papers focused on understanding roadways, each of which proposes a novel neural network architecture. We estimated roadway safety, and many other safety-related attributes, from panoramas [40] and average roadway speeds both from overhead imagery [41] and from a fusion of overhead imagery and LiDAR [42]. In our most recent work, we estimate the location of roadways, their direction of travel, and how expected traffic speeds vary throughout the week [43]. We have also worked on environmental applications, such as detecting deforestation [44], counting tree crowns [45], and classifying sinkholes [46]. We are also exploring the use of deep learning for applications in astrophysics, where our first work shows the promise of this approach [47].

This research area has also motivated some novel network training strategies. We developed an approach for training a network to predict pixel-level labels using only polygon level annotations [48], detect building boundaries using only bounding-box annotations [49], learn to detect clouds with no manual annotations [50, 51], and recover occluded regions in airborne LiDAR scans [52].

Theme 3: Medical Imaging

While I have worked on medical imaging in various ways for many years [53, 54, 55, 56], I have recently been focused on developing neural network architectures for breast cancer classification. Our first work showed the potential for neural network-based techniques [57]. We have made numerous improvements and our current architecture is now able to use the full 2D and 3D mammograms [58], significantly improving performance over the previous method which subsampled the 3D volume.

A key challenge in adopting such techniques is whether or not the clinicians will trust the system. We evaluated claims that neural-network based approaches achieved human-level performance, and found they did not withstand scrutiny [59, 60], likely due to the extreme differences between the public datasets and images used in clinical practice. As part of our effort to address the issue, we recently proposed a novel method for network calibration [61] which addresses the problem of model overconfidence. This approach is potentially applicable to a wide range of image classification problems.

More recently, we have turned our focus to the problem of limited training datasets, which arises due to the cost of obtaining manual annotations. Our first work in this direction [62], adapts a self-training approach for the task of breast cancer localization. This means a clinician only needs to provide a categorical label for the image, not a polygon or pixel-level label. We have also been developing general-purposes methods for domain adaptation [63], which we plan to adapt to the medical domain to reduce the need to annotate images for every scanner type/setting.

Summary

My research, rooted in the fundamentals of computer vision, has been dramatically affected by recent advances in neural networks. Because of this, it is now feasible to make impactful contributions across many domains using essentially the same set of tools. The key to building useful systems is working closely with experts that understand the unique features of the domain, developing architectures that are tailored to the domain, and employing sound engineering practices. A key benefit in working in several domains is seeing what problems are shared. Many of my basic research contributions have been inspired by such observations. For additional information about these, or other projects not mentioned here, please feel free to reach out to me or consult my lab webpage.

References

[1] Nathan Jacobs, Nathaniel Roman, and Robert Pless. Consistent temporal variations in many outdoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–6, June 2007. Acceptance rate: 23.4%.

[2] Nathan Jacobs, Scott Satkin, Nathaniel Roman, Richard Speyer, and Robert Pless. Geolocating static cameras. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–6, October 2007. Acceptance rate: 23%.

[3] Mohammad T. Islam, Nathan Jacobs, Hui Wu, and Richard Souvenir. Images+Weather: Collection, validation, and refinement. In *IEEE CVPR Workshop on Ground Truth*, pages 1–7, 2013. Acceptance rate: 67%.

[4] Austin Abrams, Jim Tucek, Nathan Jacobs, and Robert Pless. LOST: Longterm observation of scenes (with tracks). In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 297–304, 2012. Acceptance rate: 44%.

[5] Nathan Jacobs, Walker Burgin, Richard Speyer, David Ross, and Robert Pless. Adventures in archiving and using three years of webcam images. In *IEEE CVPR Workshop on Internet Vision*, pages 39–46, June 2009.

[6] Calvin Murdock, Nathan Jacobs, and Robert Pless. Webcam2Satellite: Estimating cloud maps from webcam imagery. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 214–221, 2013. Acceptance rate: 40%.

[7] Calvin Murdock, Nathan Jacobs, and Robert Pless. Building dynamic cloud maps from the ground up. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–9, 2015. Acceptance rate: 30.3%.

[8] Nathan Jacobs, Walker Burgin, Nick Fridrich, Austin Abrams, Kylia Miskell, Bobby H. Braswell, Andrew D. Richardson, and Robert Pless. The global network of outdoor webcams: Properties and applications. In *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL)*, pages 111–120, November 2009. Acceptance rate: 20.9%.

[9] Nathan Jacobs, Brian Bies, and Robert Pless. Using cloud shadows to infer scene structure and camera calibration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1102–1109, June 2010. Acceptance rate: 4.5% (oral).

[10] Nathan Jacobs, Austin Abrams, and Robert Pless. Two cloud-based cues for estimating scene structure and camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(10):2526–2538, 2013. Impact factor: 5.694.

[11] Nathan Jacobs, Scott Workman, and Richard Souvenir. Scene geometry from several partly cloudy days. In *ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, pages 1–6, 2013.

[12] Scott Workman, Richard Souvenir, and Nathan Jacobs. Scene shape estimation from multiple partly cloudy days. *Computer Vision and Image Understanding (CVIU)*, pages 116–129, April 2015. Impact factor: 1.54.

[13] Nathan Jacobs, Scott Workman, and Menghua Zhai. Crossview convolutional networks. In *IEEE Applied Imagery and Pattern Recognition (AIPR)*, 2016. (oral).

[14] Scott Workman and Nathan Jacobs. On the location dependence of convolutional neural network features. In *IEEE/ISPRS Workshop: Looking from above: When Earth observation meets vision (EARTHVISION)*, pages 1–9, 2015. Acceptance rate: 30%.

[15] Scott Workman, Richard Souvenir, and Nathan Jacobs. Wide-area image geolocalization with aerial reference imagery. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–9, 2015. Acceptance rate: 30.3%.

[16] Scott Workman, Richard Souvenir, and Nathan Jacobs. Understanding and mapping natural beauty. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. Acceptance rate: 28.9%.

[17] Tawfiq Salem, Menghua Zhai, Scott Workman, and Nathan Jacobs. A multimodal approach to mapping soundscapes. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018.

[18] Connor Greenwell, Scott Workman, and Nathan Jacobs. What goes where: Predicting object distributions from above. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018.

[19] Connor Greenwell, Scott Workman, and Nathan Jacobs. Implicit land use mapping using social media imagery. In *IEEE Applied Imagery and Pattern Recognition (AIPR)*, 2019. (oral).

[20] Menghua Zhai, Tawfiq Salem, Connor Greenwell, Scott Workman, Robert Pless, and Nathan Jacobs. Learning geo-temporal image features. In *British Machine Vision Conference (BMVC)*, 2018. Acceptance rate: 29.5%.

[21] Menghua Zhai, Zach Bessinger, Scott Workman, and Nathan Jacobs. Predicting ground-level scene layout from aerial imagery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. Acceptance rate: 29.2%.

[22] Tawfiq Salem, Scott Workman, and Nathan Jacobs. Learning a dynamic map of visual appearance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. Acceptance rate: 25%.

[23] Rafael Padilha, Tawfiq Salem, Scott Workman, Fernanda A. Andaló, Anderson Rocha, and Nathan Jacobs. Content-based detection of temporal metadata manipulation. *arXiv*, preprint 2103.04736 [cs.CV], 2021.

[24] Nathan Jacobs, Kylia Miskell, and Robert Pless. Webcam geo-localization using aggregate light levels. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 132–138, 2011. (oral).

[25] Nam Vo, Nathan Jacobs, and James Hays. Revisiting IM2GPS in the deep learning era. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. Acceptance rate: 28.9%.

[26] Hunter Blanton, Scott Workman, and Nathan Jacobs. A structure-aware method for direct pose estimation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, January 2022.

[27] Nathan Jacobs, Nathaniel Roman, and Robert Pless. Toward fully automatic geo-location and geo-orientation of static outdoor cameras. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 1–6, January 2008. Acceptance rate: 33.3%.

[28] Jean-François Lalonde, Srinivasa G. Narasimhan, and Alexei A. Efros. What do the sun and the sky tell us about the camera? *International Journal on Computer Vision*, 88(1):24–51, May 2010.

[29] Kenji Inose, Shota Shimizu, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi. Refining outdoor photometric stereo based on sky model. *Information and Media Technologies*, 8(4):1095–1099, 2013.

[30] Jiyoung Jung, Joon-Young Lee, and In So Kweon. One-day outdoor photometric stereo via skylight estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4521–4529, 2015.

[31] Nathan Jacobs, Mohammad T. Islam, and Scott Workman. Cloud motion as a calibration cue. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1344–1351, 2013. Acceptance rate: 26.2%.

[32] Scott Workman, Radu Paul Mihail, and Nathan Jacobs. A Pot of Gold: Rainbows as a calibration cue. In *European Conference on Computer Vision (ECCV)*, pages 820–835, 2014. Acceptance rate: 25%.

[33] Scott Workman, Connor Greenwell, Menghua Zhai, Ryan Baltenberger, and Nathan Jacobs. DeepFocal: A method for direct focal length estimation. In *IEEE International Conference on Image Processing (ICIP)*, 2015. Acceptance rate: 45% (overall).

[34] Scott Workman, Menghua Zhai, and Nathan Jacobs. Horizon lines in the wild. In *British Machine Vision Conference (BMVC)*, 2016. Acceptance rate: 39.4%.

[35] Menghua Zhai, Scott Workman, and Nathan Jacobs. Detecting vanishing points using global image context in a non-Manhattan world. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. Acceptance rate: 29.9%.

[36] Samuel Schulter, Menghua Zhai, Nathan Jacobs, and Manmohan Chandraker. Learning to look around objects for top-view representations of outdoor scenes. In *European Conference on Computer Vision (ECCV)*, 2018. Acceptance rate: 31.8%.

[37] M. Usman Rafique, Hunter Blanton, Noah Snavely, and Nathan Jacobs. Generative Appearance Flow: A hybrid approach for outdoor view synthesis. In *British Machine Vision Conference (BMVC)*, 9 2020.

[38] Muhammad Usman Rafique, Yu Zhang, Benjamin Brodie, and Nathan Jacobs. Unifying guided and unguided outdoor image synthesis. In *New Trends in Image Restoration and Enhancement (IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops)*, pages 776–785, June 2021.

[39] Scott Workman, Menghua Zhai, David Crandall, and Nathan Jacobs. A unified model for near and remote sensing. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. Acceptance rate: 28.9%.

[40] Weilian Song, Scott Workman, Armin Hadzic, Reginald Souleyrette, Eric Green, Mei Chen, Xu Zhang, and Nathan Jacobs. FARSA: Fully automated roadway safety assessment. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018.

[41] Weilian Song, Tawfiq Salem, Hunter Blanton, and Nathan Jacobs. Remote estimation of free-flow speeds. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019. (oral).

[42] Armin Hadzic, Hunter Blanton, Weilian Song, Mei Chen, Scott Workman, and Nathan Jacobs. RasterNet: Modeling free-flow speed using lidar and overhead imagery. In *EARTHVISION: Large Scale Computer Vision for Remote Sensing Imagery*, 2020. Acceptance rate: 26%.

[43] Scott Workman and Nathan Jacobs. Dynamic traffic modeling from overhead imagery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. Acceptance rate: 5.7% (oral).

[44] Raian V. Maretto, Leila M. G. Fonseca, Nathan B. Jacobs, Thales S. Körting, Hugo N. Bendini, and Leandro L. Parente. Spatio-temporal deep learning approach to map deforestation in Amazon rainforest. *IEEE Geoscience and Remote Sensing Letters*, 18(5):771–775, April 2020. Impact factor: 3.534.

[45] Hamid Hamraz, Nathan B. Jacobs, Marco A. Contreras, and Chase H. Clark. Deep Learning for Conifer/Deciduous Classification of Airborne LiDAR 3D Point Clouds Representing Individual Trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:219–230, 2019. Impact factor: 6.946.

[46] Junfeng Zhu, Adam Nolte, Nathan Jacobs, and Ming Ye. Machine learning in identifying karst sinkholes from LiDAR-derived topographic depressions in the Bluegrass region of Kentucky. *Journal of Hydrology*, September 2020. Impact factor: 4.405.

[47] Yuanyuan Su, Yu Zhang, Gongbo Liang, John ZuHone, David Barnes, Nathan Jacobs, Michelle Ntampaka, William Forman, Paul Nulsen, Ralph Kraft, and Christine Jones. A deep learning view of the census of galaxy clusters in IllustrisTNG. *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 2020. Impact factor: 5.356.

[48] Nathan Jacobs, Adam Kraft, M. Usman Rafique, and Ranti Dev Sharma. A weakly supervised approach for estimating spatial density functions from high-resolution satellite imagery. In *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL)*, 2018. Acceptance rate: 22.5% (oral).

[49] M. Usman Rafique and Nathan Jacobs. Weakly supervised building segmentation from aerial images. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019.

[50] M. Usman Rafique, Hunter Blanton, and Nathan Jacobs. Weakly supervised fusion of multiple overhead images. In *IEEE/ISPRS Workshop: Large Scale Computer Vision for Remote Sensing (EARTHVISION)*, 2019. Acceptance rate: 23.5%.

[51] Scott Workman, M. Usman Rafique, Hunter Blanton, Connor Greenwell, and Nathan Jacobs. Single image cloud detection via multi-image fusion. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020. (oral).

[52] Hunter Blanton, Sean Grate, and Nathan Jacobs. Surface modeling for airborne LiDAR. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020. (oral).

[53] Michael Dixon, Nathan Jacobs, and Robert Pless. Finding minimal parameterizations of cylindrical image manifolds. In *IEEE CVPR Workshop on Perceptual Organization in Computer Vision (POCV)*, pages 1–8, June 2006.

[54] Radu Paul Mihail, Gustav Blomquist, and Nathan Jacobs. A CRF approach to fitting a generalized hand skeleton model. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 409–416, 2014. Acceptance rate: 40%.

[55] Radu Paul Mihail, Nathan Jacobs, Judy Goldsmith, and Kristine Lohr. Using visual analytics to inform rheumatoid arthritis patient choices. In Christian Sebastian Loh, Yanyan Sheng, and Dirk Ifenthaler, editors, *Serious Games Analytics*, Advances in Game-Based Learning, pages 211–231. Springer International Publishing, 2015.

[56] Radu Paul Mihail, Gongbo Liang, and Nathan Jacobs. Automatic hand skeletal shape estimation from radiographs. *IEEE Transactions on NanoBioscience*, 2019. Impact factor: 1.927.

[57] Xiaofei Zhang, Yi Zhang, Erik Han, Nathan Jacobs, Qiong Han, Xiaoqin Wang, and Jinze Liu. Whole mammogram image classification with convolutional neural networks. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017. Acceptance rate: 19%.

[58] Gongbo Liang, Nathan Jacobs, Jinze Liu, Kyle Luo, Wendi Owen, and Xiaoqin Wang. Translational relevance of performance of deep learning models on mammograms. In *SBI/ACR Breast Imaging Symposium*, 2019.

[59] Gongbo Liang, Xiaoqin Wang, and Nathan Jacobs. Evaluating the publicly available mammography datasets for deep learning model training. In *SBI/ACR Breast Imaging Symposium*, 2018.

[60] Xiaoqin Wang, Gongbo Liang, Yu Zhang, Hunter Blanton, Zachary Bessinger, and Nathan Jacobs. Inconsistent performance of deep learning models on mammogram classification. *Journal of the American College of Radiology*, 2020. Impact factor: 3.785.

[61] Gongbo Liang, Yu Zhang, Xiaoqing Wang, and Nathan Jacobs. Improved trainable calibration method for neural networks. In *British Machine Vision Conference (BMVC)*, 9 2020.

[62] Gongbo Liang, Xiaoqin Wang, Yu Zhang, and Nathan Jacobs. Weakly-supervised self-training for breast cancer localization. In *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020. (oral).

[63] Yu Zhang, Gongbo Liang, and Nathan Jacobs. Dynamic feature alignment for semi-supervised domain adaptation. In *British Machine Vision Conference (BMVC)*, November 2021.