Changil Kim | Research Statement

Over the past decade, we have witnessed the rapid spread of digital imaging devices and the popularity of internet-based media sharing platforms, which have produced an unprecedented quantity of images and videos easily accessible to researchers. The abundance of such visual data has become a cornerstone of the deep learning methods that are revolutionizing many fields in computing. Although data-driven approaches based on "big data" are becoming increasingly popular and advancing the state of the art for numerous hard problems, their application to problems of interest is often challenging when we are not equipped with suitable algorithms, statistical models, or data.

My work focuses on capturing, curating, and synthesizing data tailored to challenging problems in digital imaging and fabrication and applying data-driven approaches for those problems.

My research is based on the following three principles. First, algorithms should now be designed with the abundance of easily accessible data in mind. Second, challenging simulations and optimizations needed for modeling physical systems can be replaced with machine learning models without sacrificing accuracy. Third, ever more sophisticated graphics engines allow us to use them as synthesizers of large-scale labeled data that can be used to solve data-starved problems.

Designing algorithms in the era of data abundance

Existing problems were often formulated to work with noisy or incomplete input data since, until recently, data acquisition had been expensive and prone to noise. Although advancing sensor technologies lowered acquisition cost and increased data quality, existing algorithms were seldom ready to take advantage of this.

In fact, many algorithms were designed with data scarcity in mind. I argue that with these changing environments, we should now design algorithms with data abundance in mind.

For instance, my work on 3D reconstruction illustrates how I design new algorithms that use a large amount of data to solve hard problems and improve upon the state of the art. At the heart of 3D reconstruction is finding accurate correspondences between views of the same scene, which remains challenging despite decades-long research.

Figure 1: Carefully designed algorithms can capitalize on a larger amount of data to make depth estimation easier, more robust, and more accurate. Here, the data includes many images taken by a laterally moving camera. An image and computed depth thereof (middle) are shown with close-ups (top). The color and depth epipolar-plane image at a scanline (bottom) demonstrate a dense sampling of vantage points, where each 3D scene point appears as a streak whose slope depends on its depth.

I introduced a completely new approach to this problem using a much denser sampling of the vantage points and demonstrated unprecedented accuracy on large real-world scenes [1–3]. The new approach used an automated setup with a motorized stage to capture a large amount of highly detailed data of the scene (see Figure 1). This setup provided a near-continuum of viewpoint changes, which enabled the design of a novel algorithm to solve the correspondence problem with a dramatically increased accuracy. At the same time, the high compatibility of my algorithm with modern GPUs enabled pixel-level parallelization and fast look-up of interpolated color values, allowing the algorithm to achieve a high level of efficiency and scalability. The resulting 3D geometry revealed more and finer details (see Figure 2). The method achieved reconstruction quality often comparable to that of active techniques such as LIDAR while being much more accessible and affordable. Furthermore, it worked reliably for highly reflective surfaces and around object boundaries, domains where the LIDAR often failed, which allowed it to be used in production movies to scan glossy props and flora.

In collaboration with Disney Research, this work led to one best paper award and three patents in addition to multiple academic publications including an ACM SIGGRAPH paper. It has since been used in movie productions, earning me credits in two Disney movies.

Data-driven inverse modeling

Many real-world systems that researchers are interested in simulating are increasingly complex. Modeling the output performance of these systems as a function of input parameters is important for accurate simulation. Even more important and challenging is to invert the simulation, so that optimal input parameters can be sought from the desired performance, which is crucial to the realization of that performance. This process is particularly challenging for systems involving physical processes such as digital fabrication, due to the difficulty of accurately modeling various devices and materials used for those systems. Furthermore, inverting their simulation process is generally not possible analytically and is achieved through costly optimization (see Figure 3). Making this inverse process more tractable and even interactive is an active research area [4].

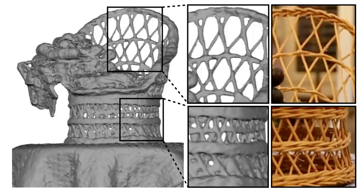

Figure 2: Shown here is the reconstructed 3D mesh of a basket with intricate details (left) from a video captured by a circularly moving camera. The insets (right) show the close-ups of meshes and roughly corresponding input views.

As an alternative to the conventional inverse modeling through optimization, I propose data-driven inverse modeling, which allows us to sidestep challenging simulation and optimization steps altogether without sacrificing accuracy. In addition, the inference of data-driven models is often instantaneous—ideal for interactive exploration of the performance space.

I led a project to physically reproduce fine-art paintings using multi-ink, multi-layer 3D printing and proposed accurate modeling of the bidirectional relation between a vertical layering of inks and its spectral reflectance [5]. The reflectance function of an ink layout involves complex subsurface scattering of various inks, interactions between ink layers, and the physical printing process, not to mention the spectral reflectance of each ink itself: inventing a physics-based model would be highly demanding. Furthermore, modeling the inverse process, i.e., predicting an optimal ink layout given a target spectrum to physically replicate, would require solving an extremely challenging combinatorial optimization problem.

Using a carefully designed dataset consisting of thousands of combinations of ink layouts, I modeled both the forward and inverse process using deep neural networks. The resulting forward model outperformed the existing physics-based models by a large margin. The inverse mapping was trained with the differentiable forward mapping as a building block, which measured the prediction accuracy of an inverse model. Our model was the first to realize the physical reproduction of *spectral* reflectance, which bears an advantage over previous color reproduction techniques in that our reproductions look faithful when compared to the original under varying illuminations (see Figure 4). The instantaneity of inference made the reproduction process highly efficient, in contrast to optimization-based approaches requiring costly optimization repeated for each instance of the input.

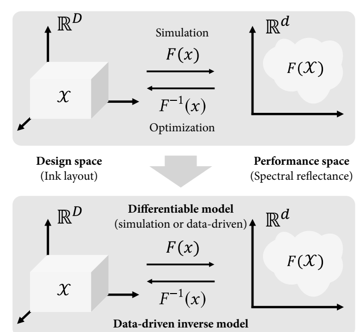

Figure 3: The simulation of a physical system is often inverted through costly optimization (top). I propose to learn the inverse process directly from data (bottom).

This work exemplifies the use of machine learning for challenging simulation and optimization problems. It has been covered in more than 30 media outlets since its presentation at ACM SIGGRAPH Asia this December and is drawing attention from art galleries and museums.

Learning by synthesis

Capturing the real world and obtaining application-specific annotations for learning systems are demanding in terms of both time and cost, and prohibitively so in some applications. Computer graphics is achieving ever more faithful simulations of the real world and suggests an alternative approach to this problem.

I see modern graphics engines as essential tools to create comprehensive and realistic labeled data that can train data-starved models, especially when the particular application does not allow for a straightforward way to collect needed data. I argue that in fact, it will be one of the major use cases of computer graphics in the future.

This possibility is especially promising in many areas desperate for sufficient data, e.g., computer vision, robotics, and physiological and neuroscientific analyses.

I applied this insight to the challenge of learning tiny motions, often imperceptible but present in everyday life—trembling muscles, respiratory motions, and peripheral pulses in a seemingly stationary face, for example [6]. Magnifying such motions allows us to see the world from a vastly different perspective, revealing issues usually hidden to the naked eye. The existing techniques used hand-engineered spatial and temporal filters to identify motions, which often fell short in cases where the particular assumptions they employed were not met.

Getting enough training data, however, posed significant challenges, since it was almost impossible to magnify such motions physically, let alone to have pairs of small motions and their magnified versions. With an image-based rendering pipeline, thousands of pairs of original and magnified motions can be photorealistically rendered with sufficient variations in the type of scenes and motions, without any human intervention. Equipped with synthetic data, I trained a deep convolutional autoencoder to learn the motion representation, which can be scaled and subsequently decoded for any factor at *inference* time. The quality of the result surpassed that of analytical models, hand-engineered and built on various assumptions, with better noise characteristics.

This work demonstrates that intricate machine learning models can be trained with synthetically generated data while attaining generalizability to the real-world input. Since its oral presentation at ECCV 2018, it has sparked various interdisciplinary research, especially for medical applications such as a noninvasive reading of vital signs or medical conditions.

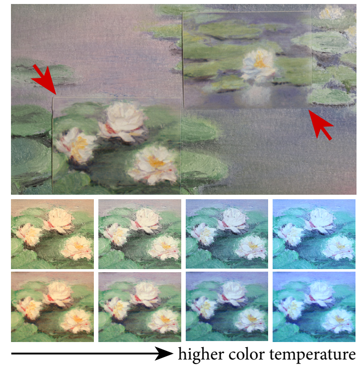

Figure 4: Our replicated patches are placed on top of the original painting, marked by red arrows (top). One of them is shown under varying illuminations (bottom), compared to the original with the matching illuminations (middle).

Future directions

So far in my research, I have demonstrated that the advancement of relevant hardware and computational tools facilitated the application of data-driven approaches, which enabled both accuracy and efficiency higher than what was possible with previous approaches. I aim to further tackle challenging problems in imaging and fabrication. Data-driven approaches owe their success to some extent to the fact that the information statistically distilled from a sufficient amount of data outsmarts our intuition and knowledge we use to solve the problem. However, this claim fundamentally depends on the availability and quality of data. Therefore, I plan to approach my goal from following three directions: using latest sensing technologies and mobile devices to mine large-scale data in multiple modalities; debiasing existing datasets and acquiring new ones in an unbiased, tractable way; and computationally synthesizing datasets for the areas where the data is inherently scarce. In what follows, I detail them with more concrete research agenda.

Embracing new sensing technologies. Rapidly developing sensor technologies offer a novel means of acquiring datasets, often of new kinds. On the one hand, ***I focus on accessible devices in widespread use.*** For example, modern smartphones and smartwatches are equipped with various sensors beyond cameras, starting from accelerometers, barometers, and gyroscopes to microphones, depth sensors, and health monitors. Previously, I used smartphones' built-in cameras and flashes to crowdsource a dataset of photograph pairs with ambient and flash illuminations [7], with the application of removing direct flash and hard shadow from flash-lit photographs in mind. I intend to apply similar approaches to other types of data. For example, realistically editing a range of photographic effects, such as depth of field, defocus blur, motion blur, and lighting, all requires the scene depth to be known. However, estimating depth from a single photograph of a general scene remains challenging and training a model that does so would require a large quantity of exemplary photo–depth pairs. While dedicated depth sensors have been used to collect 3D geometry data [8], their diversity or quantity does not match that of 2D image data. The depth sensors of smartphones can be used to crowdsource 3D data at a greater diversity and scale within a reasonable budget. These two examples illustrate only the tip of the iceberg. With the various sensors at millions of users' hands, fully utilizing the power of crowdsourcing to collect data will provide unique foundations for solving long-standing problems in respective research areas.

On the other hand, ***I embrace more dedicated sensors to scan the world in a wide variety of modalities.*** I am particularly interested in: tactile sensors, which are crucial to robotics and human-robot interaction, complementing machine vision; detectors which see outside the visible wavelength band, which paired with the data from visible detectors, can provide data essential to reduce the gap in the sensor technologies between the visible and the invisible; and a wide variety of wireless devices which use microwave signals that can join the Internet of Things.

Unbiased, intelligent data sampling. I aim to further my approaches of using data-driven methods to model physical processes in digital printing and fabrication, where it is essential to reduce the gap between what is *designed* and what is *printed*. In order to implement "WYSIWYG," any discrepancy between them arising in the fabrication process must be considered in the design stage. Apart from reflectance, I pay attention to *shape*: deviation from the target shape due to the specifics of a printing process or the level of detail achievable according to its physical limitation must be intuitively available to the designer, so as to be addressed and compensated in the design stage—potentially with interactive feedback. However, when modeling such complex physical systems, the parameter space we need to sample is often enormous. For example, in the multispectral reproduction, there would be around a billion combinations of all possible ink layouts. A set of rules based on physical limitations was used to reduce the number of data samples to a tractable number. This will be a recurring problem. Naïve sampling over all parameters evenly would be wasteful if possible, and in most cases is infeasible; an intelligent, more principled sampling strategy is crucial.

Furthermore, many existing datasets have biases [9]. One lesson I learned while working on multimodal learning of human faces and voices [10] was that the distributions of the subjects in existing large-scale datasets and the pool of crowdworkers were both very uneven regarding demography, familiarity with IT, and so on. This biased distribution has an immediate impact on the generalizability of the models trained with them. To summarize, ***in order to model complex systems requiring a large amount of data, which I am keen to delve into, it is of prime importance to sample the data in an unbiased yet tractable manner.***

I approach this problem from two fronts: from one, ***I will develop a better infrastructure for collecting data,*** such as improved interfaces or smarter sampling strategies, extending my previous research efforts in sampling and crowdsourcing interfaces [11,12]; from the other, ***I intend to develop improved learning algorithms*** that are robust to uneven data distributions and require minimal amounts of samples.

Analysis by synthesis. Modern computer graphics and computational simulation engines provide very sophisticated tools for generating a myriad of high-quality data samples required by ever more data-starved learning systems. As demonstrated by the work that uses video games to create labeled data [13] and my own work for motion magnification, the current graphics pipeline is already becoming sophisticated enough to create realistic content for some tasks. ***I predict this trend will accelerate and that realistic simulation engines will soon be widely used as data synthesizers.*** Besides, research into crowdsourcing has enabled more involved tasks to be carried out in a distributed manner by a large volume of crowdworkers. I proposed ensemble techniques to accommodate wildly varying skill sets exhibited by crowdworkers [12], a groundwork for more collaborative ensemble techniques for macrotasks. ***Together, the ability to synthesize a large amount of labeled data will open up a new horizon for data science, which I am eager to explore in depth.***

Crucially, a good simulator would allow us to recover the inverse model through *analysis by synthesis*, even if the simulator is a non-linear, non-differentiable black box [14]. The multispectral reproduction exemplifies the case of the generator being known and differentiable; analysis by synthesis complements such data-driven inverse modeling. Computer vision will certainly be the biggest beneficiary. However, other research areas in control, robotics, reinforcement learning, adversarial machine learning, and medical sciences, and experimental studies in physiology, cognitive science, and neuroscience will likely benefit from the versatile data synthesizer.

In sum, data-driven models provide advantages over previous approaches albeit requiring a significantly more amount of curated data. Modern technologies render it more and more viable to capture, curate, and synthesize required data, helping apply existing methods to a wider range of problems and lay the foundations for discovering novel approaches to tackling them. Although my research has focused on imaging and fabrication, I believe the principles I have presented apply to many other disciplines with practical applications therein. I am eager to collaborate with researchers in disciplines beyond computer graphics, such as computer vision, robotics, medical sciences, and cognitive science, to name a few.

References

[1] **Changil Kim**, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. "Scene reconstruction from high spatio-angular resolution light fields". *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 32(4), 73:1–73:12, 2013.

[2] Katja Wolff, **Changil Kim**, Henning Zimmer, Christopher Schroers, Mario Botsch, Olga Sorkine-Hornung, and Alexander Sorkine-Hornung. "Point cloud noise and outlier removal for image-based 3D reconstruction". *Proceedings of International Conference on 3D Vision*, 118–127, 2016.

[3] Kaan Yücer, **Changil Kim**, Alexander Sorkine-Hornung, and Olga Sorkine-Hornung. "Depth from gradients in dense light fields for object reconstruction". *Proceedings of International Conference on 3D Vision*, 249–257, 2016. *(**oral presentatin; best paper award**)*

[4] Adriana Schulz, Harrison Wang, Eitan Grinspun, Justin Solomon, and Wojciech Matusik. "Interactive exploration of design trade-offs". *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 37(4), 131:1–131:14, 2018.

[5] Liang Shi, Vahid Babaei, **Changil Kim**, Michael Foshey, Yuanming Hu, Pitchaya Sitthi-amorn, Szymon Rusinkiewicz, and Wojciech Matusik. "Deep multispectral painting reproduction via multi-layer, custom-ink printing". *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia)*, 37(6), 271:1–271:15, 2018.

[6] Tae-Hyun Oh*, Ronnachai Jaroensri*, **Changil Kim**, Mohamed Elgharib, Frédo Durand, William T. Freeman, and Wojciech Matusik. "Learning-based video motion magnification". *Proceedings of European Conference on Computer Vision*, 663–679, 2018. *(*equal contributions; **oral presentation**)*

[7] Yağız Aksoy, **Changil Kim**, Petr Kellnhofer, Sylvain Paris, Mohamed Elgharib, Marc Pollefeys, and Wojciech Matusik. "A dataset of flash and ambient illumination pairs from the crowd". *Proceedings of European Conference on Computer Vision*, 644–660, 2018.

[8] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas A. Funkhouser, and Matthias Nießner. "ScanNet: Richly-annotated 3D reconstructions of indoor scenes". *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2432–2443, 2017.

[9] Antonio Torralba and Alexei A. Efros. "Unbiased look at dataset bias". *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1521–1528, 2011.

[10] **Changil Kim**, Hijung Valentina Shin, Tae-Hyun Oh, Alexandre Kaspar, Mohamed Elgharib, and Wojciech Matusik. "On learning associations of faces and voices". *Proceedings of Asian Conference on Computer Vision*, 2018.

[11] **Changil Kim**, Kartic Subr, Kenny Mitchell, Alexander Sorkine-Hornung, and Markus Gross. "Online view sampling for estimating depth from light fields". *Proceedings of IEEE International Conference on Image Processing*, 1155–1159, 2015. ***(oral presentation; top 10%)***

[12] Alexandre Kaspar, Geneviève Patterson, **Changil Kim**, Yağız Aksoy, Wojciech Matusik, and Mohamed Elgharib. "Crowd-guided ensembles: How can we choreograph crowd workers for video segmentation?". *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*, 111:1–111:12, 2018.

[13] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. "Playing for data: Ground truth from computer games". *Proceedings of European Conference on Computer Vision*, 102–118, 2016.

[14] Vinod Nair, Joshua M. Susskind, and Geoffrey E. Hinton. "Analysis-by-synthesis by learning to invert generative black boxes". *Proceedings of International Conference on Artificial Neural Networks*, 971–981, 2008.