During the Winter 2014 term, I worked with Peter Van Beek to develop a new strategy to improve DSLR autofocusing algorithms using machine learning.

DSLR cameras have a moving lens, and autofocus is about finding the lens position that maximizes image contrast. One way that cameras achieve autofocus is through passive autofocus, where the camera computes a contrast measure at every movement step of the lens. This reduces to a one-dimensional search problem over the lens positions. However, the physical constraints in this situation make the problem challenging. The lens can only move in fixed step sizes (1 and 8 steps for many DSLRs) and direction changes are expensive as they can cause a mechanical backlash. Finally, the focus measure can be very noisy in the presence of camera shake of low-light situations.

In our research, we use supervised machine learning to determine what action the camera should take for the lens. We gathered data by taking a picture for each lens position in a few dozen scenes. By calculating the focus measure for each picture, we could find the peaks (lens position where an object is in best focus) and generate labels. We trained a C4.5 decision tree classifier to output, at every search step, whether we should continue searching in the same direction, backtrack and search in the opposite direction, or stop searching because we passed a peak.

Using this approach, we obtained 98.5% accuracy, over the 91.5% accuracy of previous techniques that used handcrafted rules, using only 20% more steps. That is, in 98.5% of simulations, our algorithm allowed the lens to find the correct peak. The result is even more drastic in low-light conditions, that are known to be notoriously difficult to focus in, where our accuracy was 94.0% compared to 70.3% of other techniques.

Our approach works as the decision tree classifier makes decisions a lot better than handcrafted rules. To achieve this, we used a number of tricks to tune the classifier and avoid overfitting. We added some small amount of random noise to focus measure calculations to better simulate non-ideal, real-world conditions where objects in the scene can be moving and the camera can be shaking. We balanced the number of features of each label in favor of taking more steps to reduce false negative rates. We also have much better resilience to difficult scenes that have a lot of noise and multiple objects at different focal distances, thanks to the backtracking component of our algorithm.

We also made sure that the work done could be used by other people by making it sufficiently general. Whereas some past work assumed certain features of the camera such as the size of a large lens movement step and the number of step sizes, our machine learning features were crafted in such a way to be independent of such settings.