CIS 510 Final Project

Predicting Atom Positions Using Machine Learning Algorithms

Project Team:
Adam Martini
Wes Erickson
Ran Tian

April 23, 2014

Executive Summary

The Steck Lab is investigating the motion of a single atom in a Magneto Optical Trap (MOT). The atom is imaged using a pair of aspheric lenses and a CCD camera. Given an 2D image from the CCD, the lab wants to predict the probability distribution of the atom's location within the MOT.

1 Project Description

To tackle this problem we will run simulations of the CCD output for a given atom location by solving the linear wave equation using a Partial Differential Equation (PDE) Solver. The output images will be labeled by atom location and used as training set for machine learning algorithms. The algorithms can then be used to generate a probabilistic distribution for the atom's location for a new CCD image.

2 Highlevel Architecture

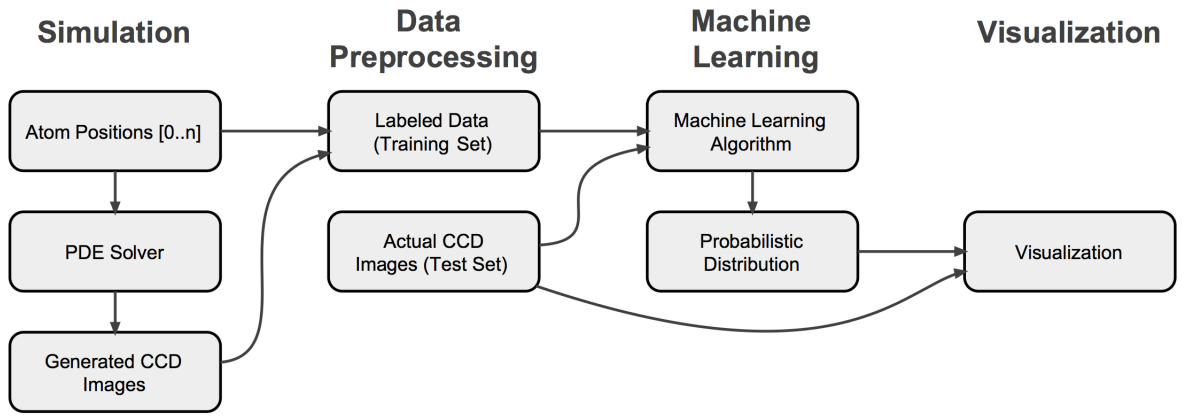

Figure 1: High Level Data Flow Architecture Diagram

Figure 1 shows a block level diagram of the project data flow. There are three main components to the solution.

Simulation

Machine Learning Test multiple machine learning algorithms to find and optimize a strong predictive model.

Visualization

2.1 Simulation

2.2 Machine Learning

We will train and test multiple machine learning algorithms on the simulated images by splitting the training data into a training and validation set. The validation set will be used to determine the best algorithm for this application and then to tune its parameters. The final classifier will be used to generate probabilistic predictions for the test set of actual CCD images.

2.3 Visualization

2.4 Data Format

The data will be formatted in the standard Visualization Tool Kit (VTK) format. We will use a regular regtagonal structured grid.

3 Parallel Plan

The each stage of the project data flow includes several places to introduce and explore parallelism.

Simulation PETSc...

Machine Learning There is at least one parallel machine learning library written in c++ that we can use as a starting point. To get specialized functionality or algorithms that perform well for image recognition, we may need to extend existing libraries or code our own algorithms. It is unclear whether a distributed program will be necessary to train the models given the number of features and size of the dataset.

Visualization Visit, paraview...

4 Project Schedule

Week	Deliverable
5	PETSc PDE solver installed and running data runs, identify parallel classifiers and get them running on acciss.
6	Get classifiers working on acciss and extend/create code based on prediction goals.
7	Process initial Petsc ouput to identify best classifiers.
8	
9	
10	