Avi Singh

Curriculum Vitae

+91-8853544535
avisingh599@gmail.com
https://avisingh599.github.io

Research Interests

Computer Vision, Machine Learning, Robotics

Education

2012–2016 **Bachelor of Technology** *Indian Institute of Technology* Kanpur *GPA 9.4/10.*
Major: Electrical Engineering, Minor: Computer Science (Artifical Intelligence)

Publications

arXiv link **Recurrent Neural Networks for Driver Activity Anticipation via Sensory-Fusion Architecture**
Ashesh Jain, **Avi Singh**, Hema Koppula, Shane Soh, Ashutosh Saxena.
under review at ICRA 2016

extended abstract **Brain4Cars: Sensory-Fusion Recurrent Neural Networks for Driver Activity Anticipation**
Ashesh Jain, Shane Soh, Bharad Raghvan, **Avi Singh**, Hema Kopulla, Ashutosh Saxena.
Full Oral at BayLearn Symposium, Menlo Park, CA, USA, October 2015

Research Experience

May-July 2015 **Research Intern,** Cornell University
project page **Brain4Cars**: Anticipating Maneuvers via Learning Temporal Driving Models
under **Prof. Ashutosh Saxena**, Department of Computer Science.
Brain4Cars addresses the problem of anticipating driver maneuvers several seconds before they happen. It fuses the information from driver-facing and road-facing cameras with data from other sensors to make its predictions. These predictions can then be passed to driver assistance systems that can warn the driver if the maneuver is deemed to be dangerous. My contributions to the project are listed below:

- The KLT face tracker in the project was replaced with a facial landmark localization pipeline based on Constrained Local Neural Fields. This provided robust tracking and allowed the computation of head pose, which then served as a strong feature for maneuver anticipation.
- Implemented Gaussian Mixture Model-based initialization, and LBFGS optimization for training Autoregressive Input Output Hidden Markov Models (AIOHMM), which are a modification of HMMs and used for anticipation in Brain4Cars.
- The performance of the AIOHMM-based anticipation system improved from a Precision/Recall of 77.4/71.2 to 86.7/78.2.
- Further testing on a Long Short Term Memory (LSTM) network (replacing the AIOHMM) increased the performance to a Precision/Recall of 90.5/87.4.

July-Dec 2014 **Undergraduate Researcher,** IIT-KANPUR

Github-1 **Visual Odometry for Ground Vehicles**

Github-2 under Prof. KS Venkatesh, Department of Electrical Engineering.

Visual Odometry is the problem of estimating the trajectory and pose of a vehicle using a video stream from a camera (or a stereo pair) that is rigidly attached to the vehicle. Two **stereo** approaches were implemented and evaluated on the KITTI odometry benchmark:

- Jiang2014: Model Based ICP: 3D points triangulated from stereo data, inliers detected via the use Iterative Closest Point Algorithm that uses a 1-DOF motion model for initial estimate. Efficient PnP Algorithm is then used on the selected inliers to obtain the final rotation and translation
- Howard2008: Inlier detection using an assumption of scene rigidity. Problem reduced to finding the maximum clique in a graph, solved using a heuristic. Levenberg-Marquardt used for minimizing the reprojection error on the selected inliers.

A **monocular** visual odometry approach was also implemented:

- Utilizes Nister's five point algorithm for essential matrix computation.

May-July 2014 **Research Intern,** IIT-KANPUR

Scene Flow Estimation from RGB-D data

under Prof. KS Venkatesh, Department of Electrical Engineering.

Scene Flow is an extension of the classical Optical Flow problem to RGB-D data.

- Implemented an approach based on the principal of 'Global Minimum Energy Solution', which is an extension of the Horn-Schunk method for Optical Flow.
- The second approach implemented is an extension of the Lucas-Kanade method for Optical Flow, and makes use of the 'Total Least Squares Solution' principle.
- Captured RGBD Data from a Microsoft Kinect using OpenNI and OpenCV libraries in C++, and qualitatively evaluated the results.

Selected Projects

October 2015 **Deep Learning for Visual Question Answering** *For CS671*

Github under Prof. Amitabh Mukherjee, Department of Computer Science, IIT-Kanpur.

I developed Neural Network-based models to tackle the problem of answering open-ended natural language questions about images. This project was judged as the **best project** in the course (out of 20 projects).

- The image is passed through a **Convolutional Neural Network** (pre-trained on the ImageNet dataset), and the activations from the last hidden layer are extracted. This 4096-dimensional vector serves as a fixed-length representation of the image.
- Every word in the question is converted to its word vector using the Glove Word Embeddings, and these embeddings are sequentially passed to a Long Short Term Memory Network (LSTM). The output of the LSTM after all the words have been passed is used as an embedding for the question.
- The image embedding and the question embedding are concatenated and passed through two fully connected layers with 50% dropout. The entire network (except the CNN) is trained end-to-end.
- The VQA dataset was used with about 370K questions for training and 240K questions for testing. The system was able to achieve an accuracy of 53.34% on the test-dev split of the VQA dataset.

April 2015 **Hidden CRFs for Human Activity Classification from RGB-D videos** *For CS679*
under Prof. Vinay Namboodiri, Department of Computer Science, IIT-Kanpur.

As part of a course project, we worked (in a team of two members) to tackle the problem of recognizing human activities from RGB-D videos.

- Hidden Conditional Random Fields (hCRFs) were selected for modeling the problem, due to the advantage that they offered over HMMs, MEMMs, and regular CRFs.
- Pose features were extracted from a skeleton detection pipeline.
- A novel normalization operation was implemented for reducing the effect of variance in body sizes of different subjects.
- An accuracy of 71% was achieved on a reduced version of MSR Daily Activity 3D Dataset (6-class classification).

Jan 2015 **DAAnT - Computer Vision for Monitoring Oral Health** MIT REDX CAMP
under Dr. Hyunsung Park, Postdoc at Camera Culture Group, MIT Media Lab.

We developed Computer Vision algorithms for early detection of dental problems using images obtained from an intraoral camera. My contribution in the eight-member team was primarily in the following areas:

- Stitching Images: Used Affine SIFT to overcome the difficulties in feature matching due to changes in perspective, and then computed Homography with RANSAC to stitch the images.
- Segmenting Every tooth: Marker-controlled watershed transform was used to segment every tooth from the image. Both semi-automated and automated approaches were implemented and compared.

Dec 2013 **Landmark-based Robotic Localization from RGBD data**
under Mr. Arjun Bhasin, Project Engineer at Mechatronics Lab, IIT-Kanpur.

Robotic Localization is the problem of determining the pose and location of a robot, often using sensors installed on the robot.

- Geometric Triangulation was used to determine the pose (modeled as a set of random variables with Gaussian Distribution) of a robot, using bearing measurements of known landmarks.
- A Microsoft **Kinect** was used to identify the landmarks (using color histogram based models), **CAMshift** algorithm was used to track these landmarks, and bearing measurements were calculated using the depth data.
- An error model for the Kinect data was used along with the Error Propagation Law to arrive at the uncertainty in the final pose computed using the Geometric Localization Algorithm.

March 2014 **Hilbert Transform on FPGA/Verilog** *For Techkriti 2014*.
Github Hilbert Transform is a mathematical operation used in Signal Processing. A hardware implementation can provide faster computations as compared to a generic CPU implementation.

- Implemented a 32-point Discrete Hilbert Transform in **Verilog**, making use of the **Fast Fourier Transform** in the intermediate stages.
- Won **2nd position** in FPGA Design, Techkriti 2014.

2012-14 **Hardware Hacks at Electronics Club** *Just-for-fun projects*.

- **Cashless Campus**: Developed an **arduino-based** point-of-sale device, with **biometeric (fingerprint) authentication** . An Arduino Mega was interfaced with an ethernet shield (with Wiznet51000 chip for UDC/TCP stacks), an LCD, a touchscreen, a thermal printer, and an SD card via **SPI and UART**.[Github]
- **SNAKE64**: Implemented the classic 'Snakes' game on a self-fabricated LED matrix of size 8x8, and wrote an original C code for ATmega8 to drive the matrix.
- **Laser Tag**: An infrared-based gun was implemented using IR LEDs and 38KHz modulated wave was generated on an ATmega32. The receiving unit employed a TSOP to detect bursts of infrared sent from the gun.

Awards

2014 Awarded an **A* grade**, for exceptional performance in Undergraduate Project - 1.

2012-2013 **Academic Excellence Award**, IIT Kanpur (Awarded to 60 students out of 840).

May 2012 Secured All India Rank **387** in IIT-JEE 2012 out of **0.5m** candidates.

May 2012 Secured All India Rank **345** in AIEEE 2012 out of **1.2m** candidates.

Jan 2012 **Top 1%** in National Physics Olympiad 2012

Relevant Coursework

Statistical AI Machine Learning for Computer Vision, Learning with Kernels*, Natural Language Processing*, Online Learning and Optimization**, Probabilistic Mobile Robotics**, Convex Optimization**, Artificial Intelligence**

Mathematics Linear Algebra, Probability and Statistics, Multivariate Calculus, Ordinary/Partial Differential Equations, Complex Analysis

Algorithms Algorithms-II*, Data Structures and Algorithms, Intro to Computing

Signal Processing Signals and Systems, Digital Signal Processing, Communication/Information Theory, Principles of Communication

* denotes course completed in Fall 2015 ** denotes course to be done in Spring 2016

Technical Skills

Languages C, C++, Python, MATLAB

Libraries Keras, OpenCV, ROS

OS GNU/Linux (Ubuntu), Microsoft Windows

Other Git, LaTeX

Activities

2014-2015 **Electronics Club Coordinator, Science and Technology Council, IIT-Kanpur**.

- Floated, mentored and ensured the completion of nine summer projects including a **3D Laser Scanner**, **A Video Surveillance Robot**, Conway's Game of Life simulation using FPGAs, Fast Fourier Transform on FPGA, An accelerometer based fitness and sleep tracker with accompanying Android App, a Surveillance system with **face recognition**, and a Laser Tag system.
- Led a team of 16 secretaries and handling a budget of Rs.76,000 to organize lectures, workshops, competitions, and another Rs.4,74,000 for funding projects and for participation in external events.
- Lectures attended by 400+ people, workshops attended by 200+ people, and participation of 100+ people in Takneek (intra-IIT Kanpur technical festival) Electronics competitions.

2014- **Blogger**.

I love explaining things in a simple manner, and I've written several blog posts that introduce research problems (and ways to tackle them) to beginners. Some of my posts have been published on websites like **KDNuggets** and **LearnOpenCV**. Some recent posts:

- Deep Learning for Visual Question Answering link
- Monocular Visual Odometry in OpenCV link
- Visual Odometry from Scratch - A tutorial for beginners link

2015- **Member, Special Interest Group in Machine Learning (SIGML), IIT-Kanpur**.
I have given a talk about my work, and attended lectures given by others in the various sub-fields of Machine Learning.

2013 - 2014 **Coordinator, ECDC, Techkriti 2014**.
Designed and verified the problem statement for the event Electromania. Prepared sample codes and tutorials for the participants.

2013 - 2014 **Secretary, Electronics Club**.
Assisted in organisation of lectures, workshops, tutorials, and maintenance of club.

2013- **Student Guide, Counselling Service**.
Assisted in the organisation of various counselling service activities such as the Orientation Program and helped a group of six freshers in settling in the new college environment.