Table of Contents

1. Note on Deep Learning Projects

As discussed in the lecture, deep learning projects should have a computer vision focus, i.e. Deep Vision rather than pure Deep Learning. Obtain feedback from LIC and tutors if in doubt.

2. Special Projects

Contact the Lecturer in Charge directly if you are interested in the projects in this section.

2.1. Image processing to quantify growth rate - a lizard evolution and ecology project

Rising global temperatures poses a serious threat to animals, particularly of ectothermic species. Accumulating evidence suggests that early-life environment experienced by embryos can alter the physiology, which can have cascading effects on growth rate, postpartum. These developmentally plastic responses can in some cases better match an individual to the environment it will have to survive in. This project will specifically investigate how 'hot' and 'cold' nest temperatures affect lizard hatching growth rates. Images of newly hatched lizards are taken every 21 days, head and body dimensions need to be extracted from these images and over time. Our goal is to estimate a growth rate curve for each lizard born from these different temperatures. The project allows us to answer big questions about whether animals can respond well to environment changes such as global warming.

2.2. Image processing to measure dimensions of turtle eggs

There are around 1000 photos of eggs of a predatory ground beetle Pterostichus oblongopunctatus. The photos were taken under a microscope and have same resolution and magnification, thus measurements in pixels are comparable between the photos. The aim is to measure the eggs: at least length, width, but perimeter, area, volume are also desirable (the last 2 can be estimated from the first two, but there might be better ways of estimating these from the photos). The method can be potentially transferable to eggs of other species of animals with similar eggs (especially birds).

3. Successful Projects from Past Years

3.1. Hypercube: An Augmented Reality Assistant to Solving the Rubik's Cube

A realtime augmented reality application for guiding a user at any level of expertise through steps to solve a Rubik's Cube, providing responsive hints based on how the user is holding the cube. The goal of this project is to design and implement a realtime augmented reality (AR) guide for solving the Rubiks Cube. The emphasis will be placed on building an application which is robust and responsive. The user should be able to handle the cube naturally without being too constrained to hold it still for a long time or in a specific angle. This will mean that the system should be quick to respond to changes in the cube in terms of simple orientation as well as rotation of faces. The interface of the application will consist entirely of easily readable augmented reality based directions and smooth interactions with the user based on the current orientation of cube.

3.2. Spherical Panorama Generation and Vanishing Point alignment with Transitional Effects

This project aims to produce 360°x180° spherical panorama images through the use of our own image stitching algorithms to combine individual images (either as a set of still images or as the frames of a single video) into one spherical image in a standard spherical image format. A system will then detect vanishing points in multiple such panoramas (which are assumed to be taken from locations connected by a straight line) to detect the direction of motion connecting each subsequent panorama. A 3D viewer will then utilise the found direction of motion to align each panorama in order to allow for simulated movement between each panorama.

3.3. Visual Fall Detection using Aggregated Vision and Machine Learning Models

Our aim is to create a visual based system that can recognise when a person has fallen down, making it possible for automated alerts to proper personnel. The system is designed to run on cheap hardware, so only a single colour camera or a kinect camera will be used (with a reasonable resolution). Processing will have to occur in real-time to be able to handle events as they come. Furthermore, the system should be able to gracefully handle fall-like events like tying one's shoelaces, while ignoring ‘safe-zones’ like beds and couches. It should also recognize the difference between people and people-like entities, such as pets or mats, and ignore these non-persons.

3.4. Hand-held Video Enhancement with Stabilization and Mosaicing

Compact/cell phone cameras will continue to have two major limitations in the foreseeable future – i.e. lack of hardware based image stabilization, and small sensor size which result in video capture with unintentional shakes, noise, and low- resolution. Yet due to their convenience and ubiquity, they often capture unique and important footages. We design and implement a solution to address the two major problems above. The shakes are removed with feature detection and motion estimation. This is followed by mosaicing, which fills ‘gaps’ or undefined regions in the stabilized video. Furthermore, super- resolution techniques are used to improve the resolution and remove noise by interpolating sub-pixels from multiple low- resolution images.

3.5. Video Google: A Text Retrieval Approach to Object Matching in Videos

This project is mainly focuses on object and scene retrieval i.e. tries to find all the occurrence of the user query object within all frames of a video. The query object is presented by scale and viewpoint invariant features, so it is resistant to viewpoint, illumination and partial occlusion. The idea of retrieval is coming from google text retrieval algorithm where google search documents and webpages for a particular word.Since the frames and shots descriptors are pre-computed, retrieval and frame matching is done with no delay.

3.6. Natural landmark localisation for RoboCup

This paper outlines a fast and effective method for mapping natural landmarks to improve localisation during RoboCup soccer matches. The method uses modified 1D SURF features extracted from the row of pixels on the robot's horizon. These features are then matched across images using a dynamic programming algorithm that ensures that features can only be matched if they appear in the same order in each image. We demonstrate that this approach is robust to lighting changes, occlusion and small changes in viewing angle, and computationally cheap enough to run on the Nao humanoid robot in real time. As such we are confident that with further work this approach can be used as the basis of an enhanced localisation system for playing robot soccer.

3.7. Unsupervised 3D Object Reconstruction and Visualisation

In this project, an unsupervised 3D object reconstruction technique using various computer vision algorithms are implemented and visualised. For key point and feature detections, Scale-invariant feature transform (SIFT) algorithm and Speeded Up Robust Feature (SURF) algorithm are benchmarked and used. Those key points are matched by using Cross-checking algorithm and kd-Tree based search algorithm. With this matched key points, the fundamental matrix can be calculated and the point cloud and cameras can be extracted from a list of images. In order to improve the quality of the point cloud, various techniques such as outlier rejection and bundle adjustment algorithms are applied. Then the point cloud and cameras are visualised by using OpenGL and the Qt framework.

3.8. 360◦ × 180◦ Panorama

We present our experience of implementing a panorama stitcher with efficient bundle adjustment using the Levenberg-Marquart optimisation method and many optimisations, including implementation details and evaluation. While the performance is impressive, reliability is low, though we believe further work could improve this. We also present some modifications to homography estimation using RANSAC in order to increase accuracy, and provide an evaluation.

3.9. An augmented reality system

The goal of this project was to create an augmented reality system that uses a deck of playing cards to augment the environment. This system used openCV to track and solve the motion of the cards and openGL to draw the cube onto the cards. In openCV an optical flow algorithm is used to track the motion of the card. We use a Lucas- Kanade pyramid to interpolate the points between frames. The Lucas-Kanade method retrieves the key points of the object. These are then passed into a function that calculates the pose between the object and the camera. After the pose is retrieved this information can then be given to the openGL side of the program for the cube to be drawn. Initially a SIFT matcher was implemented in an attempt to match and augment multiple cards. However, throughout the testing phase of the project, we realised that there was a major shortcoming with using this method. It ran at an incredibly slow frame rate. Thus an optical flow method was quickly adopted. Using this method, a system that could track a card quickly and accurately and also be able to augment a cube on top of this card was developed.

3.10. 3D Reconstruction with Handheld Camera

Pictures and videos are 2D representation of 3D world, when viewing a collection of pictures or video, the human brain could extract information from each eye and construct 3D shape of the shown object. In theory, a computer can do something similar to what human brain that can do via logical calculations. Due to this, it is possible to reconstruct a 3D world with 2D collection of pictures.

3.11. Generic Pedestrian Detector Computer Vision

This report illustrates the implementation of a generic pedestrian detector for the COMP9517 project. It provides an overview of the project's goals, relevant literature, design considerations and concluding remarks. It outlines issues that were encountered in the implementation and the decomposition of the project. Future extensions and suggestions are also explored in the report thus promoting further development and research in the field of computer vision.

3.12. Removing Background Objects from Videos

This project investigates using seam carving as a method to remove elements from the background of a scene in a video. A number of problems with doing so are encountered, and solutions to them are suggested and investigated.

3.13. A Webcam Based Touch Screen Approach Using Parallel Plane Screen Projections

This project presents a novel way to ascertain finger location with respect to a screen by using two overlapping webcams with projections taken parallel to the plane of the screen.

This method is theoretically less susceptible to occlusion and can determine touch events far easier than previous webcam based touch approaches. The results obtained are compared with (reference) and reference and found to be quite satisfactory. The proposed approach resulted in an average location deviation of 27.6 pixels, correctness w.r.t touch of 95.3%, completeness of 100% and was easily performed processing at an average of 27.9 frames per second per webcam feed at a resolution of 640 x 480 pixels.

3.14. Lane Detection and Tracking in Grassy Areas

This project proposes an algorithm to extract these white line obstacles from a video / image sequence stream and project them into a 3D coordinate system so they can be added to an occupancy grid and used for path planning by the robot. This can be fused to generate a consistent map, as well as for localization / estimation of movement between frames.

3.15. LingAR – Augmented Reality Translation

This report outlines the implementation and evaluation of an augmented reality translation system. It aims to justify decisions made in segmentation, algorithm selection, and limitation of goals in order to produce a successful static image translator. Further, it discusses how results are analyzed and how the system could be further expanded to a live video translation system.

3.16. Representation of Intensity Variation Inside Prostate on DCE-MRI

Dynamic contrast enhanced MR imaging (DCE- MRI) or Perfusion MRI has shown to be a promising imaging technique for identifying abnormalities in soft tissue cells in human body, specially in oncological imaging. Hence utilization of DCE-MRI in prostate cancer recognition has been discussed in recent studies widely. In this project a methodology to segment the prostate gland in 3D DCE-MRI, and a representation of intensity variation inside the prostate over time after the contrast enhancement agent is injected are presented. Expectation Maximization (EM) algorithm and Graph-Cuts algorithm alongside with shape based knowledge are used for 3D prostate segmentation. Then an intensity variation analysis is presented within the segmented prostate boundaries.

3.17. Classification of three-dimensional medical images using two-dimensional and three dimensional scale-invariant feature transform

The aim of this project is to develop a framework for the use of 3D medical image volumes (scans) to classify diseases. For development purposes, we used brain scans acquired from patients with Alzheimer's disease (AD) and from age-matched controls. Training was performed on 20 AD and 20 normal scans. Our method is based on an article by Toews, et al [1]. Scale invariant feature transform (SIFT) features, either 2D or 3D, are generated from pre-processed brain magnetic resonance images (MRI). Feature reduction is performed with unsupervised clustering by location and appearance, using the SIFT feature descriptors. Clusters are then assigned likelihood ratios based on the relative proportions of AD and normal scans occurring in each cluster, completing the learned model. The model is then used to classify new scans with a naïve Bayes approach. Newly extracted SIFT features are assigned to clusters if they fall within a threshold distance, and the likelihood of a scan belonging to a class is determined by the product of likelihood ratios of those features that match. The model was tested using 2D features in three planes. Area under receiver-operating-characteristic curve of 0.84 was attained. Peak classification accuracy of 0.83, sensitivity of 0.95 and specificity of 0.70 were attained. This degree of accuracy is comparable with published studies. There is potential for improvements in speed and accuracy given further work.

3.18. 3D Pose Estimation of known textured objects using Monocular Images

Pose estimation is the elementary problem in a number of applications, particularly where the position and orientation of an object is required. The aim of this project is to estimate the six degrees of freedom that describe the 3D pose of an object given a data stream from a single monocular camera. Fundamental in this process is a model of an object of interest from which feature matching between this object model and the data stream can be used to firstly recognise pixels as belonging to an object. Secondly, an iterative algorithm is used to estimate the pose. This paper will show that given good feature detection and matching methods and iterative algorithm the pose of a known object can be accurately and reliably estimated.

3.19. Advertising Elimination via Logo Detection in Everyday scenes

Detection of arbitrary logos in everyday scenes is an interesting problem involving computer vision. This report examines the use of a Bag-of-features combined with a Support Vector Machine to classify and localise logos within the FLICKR32 dataset. We discuss the tuning of key parameters of both the BOF and SVM elements and characterise the performance of the approach, finally outlining potential pitfalls and suggesting directions for future investigation.

3.20. Car Logo Recognition Using HOG Descriptors and linear SVM Classifier

Several image processing methods are implemented to detect the license plate location in this project. In order to discover ROI of car logo, a Sliding Window approach is used to find ROI of car logo which is combined with non-maximum suppression technique. The HOG feature descriptor could perform better for logo recognition. Training a linear SVM exemplar classifier for car logo recognition can improve the accuracy.

3.21. 3D Reconstruction of objects from a collection of 2D images from different angles

This project looks at creating a 3D model of an object in the form of a point cloud from a collection of 2D input images and the positions that they were taken from. This is based on the area of structure from motion and is related to stereoscopic mapping, 3D scanning and feature matching. It has been looked at in numerous ways previously, including for augmented reality and SLAM based mapping. This is a growing area of computer vision with numerous real world applications including manufacturing, component testing, medicine, 3D printing and robotics. While it is still a relatively recent area compared to most of computer vision, it has already matured to the point where there are usable solutions online and it is beginning to move towards integration in smartphones, such as Microsoft's Mobile Fusion. There are many approaches depending on how much is known of the scene, the cameras, the objects and the lighting. For example, depending on whether the scene has many features which look similar, areas which have a lack of keypoints, depending on whether camera positions and focal lengths are known and number of photos provided.

Most of these fall under one of two categories in terms of how they change the base algorithm, either to do with finding features differently (such as using curves) or to do with guessing camera parameters during construction. The scope of this project includes finding features of images using the SIFT detector, matching features, point cloud reconstruction and point cloud refinement. It does not include calculating the intrinsic and extrinsic matrices of the cameras.

The final result should be a program that given a set of images and the intrinsic and extrinsic matrices of the camera that took them, reconstruct a point cloud of the 3D object that the images are showing. The idea is for it to require no user input, simply given a collection of images and their intrinsic and extrinsic matrices, it should output a refined point cloud of the object.