AMME5710 Computer Vision and Image Processing

Example Major Projects, 2025

Project A: Board Game Tracking

Although traditional board games such as Chess and Go can be easily played on a computer or in an online environment, people still enjoy playing the old-fashioned way on a physical board. Playing on a physical board however makes it more difficult to “transcribe” games (i.e. create a record of exactly what moves were made when), or to adjudicate a game (i.e. determine if/when a player has made an illegal move or broken the rules in one way or another).

In this project, you would develop a computer vision system to track the exact moves made by players in a board game of your choice (i.e. Chess, Checkers, Go, Chinese Checkers, Monopoly). The system could use a video stream coming from a webcam, mounted in a way that it can view the board, but is not a hassle to setup, and does not impede the player’s ability to see and access the board and pieces. Such a system could, for example, provide data on the current position of pieces, track movements from one play to the next, detect when illegal moves have been made. Some issues that should be considered:

- How robust is the system to occlusions (i.e. from a player’s hand when they move a piece)?
- Can the system identify/classify and track individual types of pieces? (i.e. a rook vs. a pawn)
- How much cumbersome/clumsy calibration is needed before using the system? How much of this can be automated or designed such that the system is not particularly sensitive to where the camera is placed? Ideally, the system should be able to be mounted in a fairly haphazard way (as long as the board can be seen). Imagine a use case for this system where it is setup by someone who knows nothing about computer vision or cameras.

Project B: Human hand gesture recognition and sign-language translator

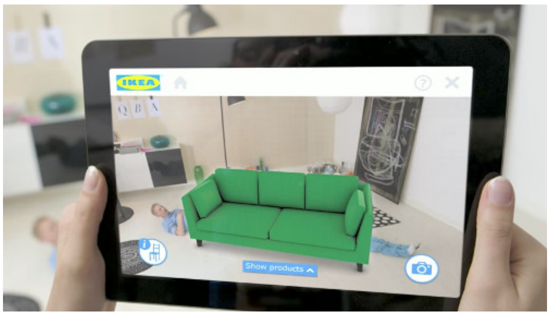


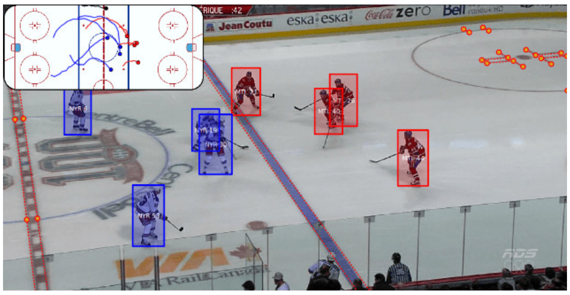
More than 1% of the world's population are unable to hear – many of these people use sign language as their primary means of communication. There are many different sign languages: in Australia, the Auslan system is used (http://www.auslan.org.au/).

In this project you would develop a system for recognising Auslan hand gestures from a video stream that can translate or recognise basic sign language letters, numbers or words. Such a system would potentially be able to recognise/detect the position of someone's hands in a video stream, normalise the orientation of the hands and potentially the fingers, and to recognise certain gestures based on this data. The Auslan alphabet A-Z involves mostly static hand gestures (excepting the letter J). Recognition of these gestures therefore could potentially involve a single image or video frame. Most other sign language words involve movement in the gesture, and a more advanced system for translating words would need to account for the motion of multiple frames in a video sequence. Some issues that should be considered:

- How does the person signing have to be oriented with respect to the camera position?
- Is the system robust to different clothing that the person might be wearing or backgrounds (i.e. does the person need to be positioned in front of a specific background)?
- How slowly must the gesture be performed?

Project C: Augmented Reality Systems



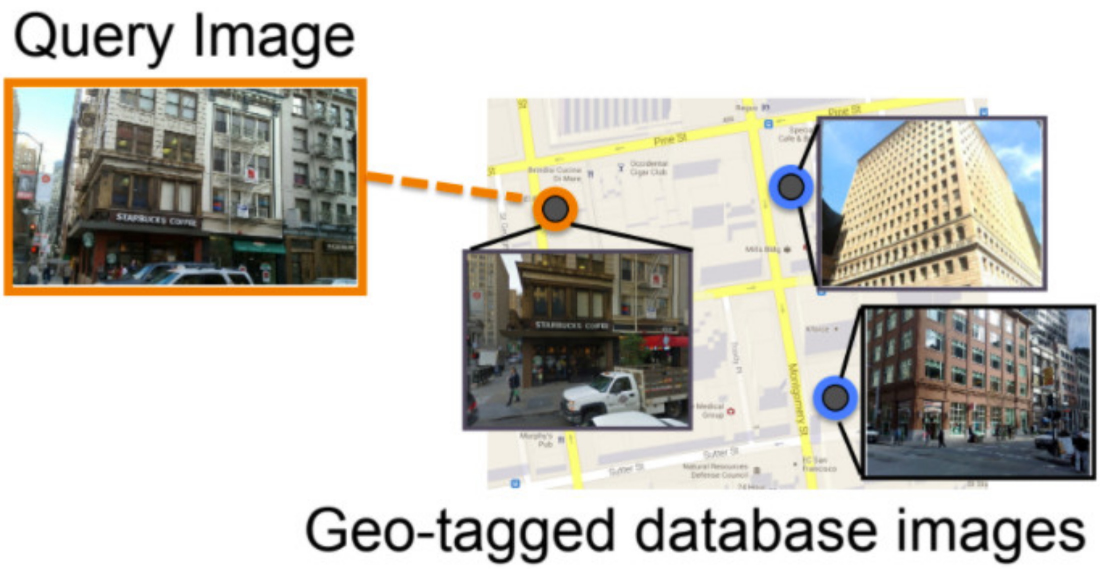


Augmented reality systems involve the superposition of virtual or computer-generated data onto real-time imagery to create a composite scene composed of real and virtual objects. Most augmented reality systems therefore need to be able to estimate the 3D position and orientation of a camera with respect to a world reference frame from imagery data itself and simultaneously project a virtual object's 3D location into image coordinates based on the object and camera motion. Systems may use either a target of known geometry to estimate the pose of the camera or use "marker-less" methods, for which techniques in structure-from-motion can be used to estimate the motion of the camera with respect to the world.

This project would involve developing an augmented reality system that can project a 3D object into a live video stream or sequence of images, such that the object appears to exist in the real world. One example application is to generate an Ikea furniture visualiser: a phone, tablet or other mobile device is held in front of an inside area in an empty room, and the user is able to visualise the appearance of a piece of furniture, without having to physically move it there. There are also many other potential applications in enhancing video, for example, from sporting matches: superimposing on sports video a player's stats right next to that player, indicating "hotspots" for the location of the ball throughout the match etc. Some issues that should be considered:

- Most of these capabilities rely on being able to extract information about the field from the video stream and being able to compute the spatial relationships between the camera and the scene. How would this be done?
- Does the method rely on any targets or other markers to be present in the image?
- How do you deal with the interaction between real-world and virtual objects? Which object is occluding which?

Project D: Virtual Guide/Visual Place Recognition



GPS and other satellite navigation systems are built into almost all new mobile devices, such that finding your way around a new neighbourhood is easy. But GPS tends to fail in the middle of a city or within indoor areas where signals don't reach. In this situation, most people look for recognisable landmarks that they can see and can be used to help figure out where they are.

In this project, you would develop a system for visual place recognition that can be used for example from a phone camera (or webcam attached to a laptop) that can work out a user's location on a map using an image of the immediate surroundings. One way to do this is to use a database that contains geotagged images of an environment, and to use techniques from feature matching and stereo vision to estimate the location from which the current photo was taken based on how it matches to the database. Some issues that should be considered:

- How do you account for changes in the perspective between query photos and the image database?
- How do you account for differences in lighting and appearance that occur, for example due to the time of day (shadows etc.) or day vs. night-time navigation?