**First Principles of Computer Vision Course**

**Module 1:** Camera and Imaging

**Week 1 Lecture Notes**

**GOAL:** To build machines that can see.

* Methods to recover information about the physical world from images.
* Our vision system is qualitative not quantitative and is not good at making precise measurements about our physical world. A computer vision system can be designed to surpass the capability of human vision and extract information about the world that we cannot.
* Without light, there can be no vision.

**How vision works:**

1. The scene reflects some light it receives towards the camera. The camera plays the role of the human eye.
2. The camera receives light from a 3D scene to form a 2D image.
3. The 2D image is passed to vision software that seeks to analyse the image and come up with a symbolic description of the 3D scene.

* Image is an array of pixels. **Pixel** is short for “picture element”. In an image, a pixel records the brightness and colour of the corresponding point in the scene. A pixel can be richer in the info it records and can also record the distance (depth) of the corresponding scene point.
* **Optical Character Recognition (OCR):** OCR can be used in traffic systems to identify vehicles that violate traffic rules (license plates).
* **Iris Recognition:** Biometric identification of a person based on the intricate patterns in a person’s irises unique to them. Can determine identity with very high confidence. Example of machines seeing things that the human eye cannot see.
* **Face detection:** One of the most successful applications of vision.
* **Intelligent marketing:** Vending machine in Shinagawa Station Tokyo can estimate the rough age and gender of a person through face detection and recommend products to purchase.
* **Tracking people moving through space:** In surveillance, a person can be tracked moving through a crowd, even if they get obstructed by other objects or people in a scene. If a person leaves the field of view of one camera, they can be handed off to another camera that continues to track them.
* **Optical mouse:** The computer vision system here tracks the movement of the pattern or texture of the surface on which the mouse sits. This is done using a low-resolution camera with a very high frame rate, allowing the mouse to precisely estimate its motion with respect to the surface it sits on. This info is used by computers to control the position of the cursor.

**How humans see:**

1. Images of the world are captured by our two eyes. Early visual processing takes place in the eye itself (info recorded by the retina is processed by retinal cells to reduce the info that needs to be transmitted to the brain).
2. The info travels via the optic nerve to the lateral geniculate nucleus where it is relayed to the visual cortex (part of the brain in the back that does the visual processing).
3. There are different parts of the cortex that perform different functions (e.g. perception of shape, colour, motion, etc.).
4. Don’t know enough about the visual cortex to replicate it electronically! So we reinvent. No computer vision system has been created that is as versatile as the human one (however for quantitative tasks, there are many computer vision systems in use today that exhibit higher precision and reliability than human eyes).

* **Image sensors:** convert the optical image formed by a lens into a digital image.
* **Imaging revolution contributor:** convert optical images to electrical signals.
* **Binary image:** simplest type of image which is two-valued and obtained by thresholding a captured image. You end up with white (or 1) for object or black (or 0) for background. Used for factory automation and are easy to store and process.
* **Information theory:** edges in an image are of great value (edge and corner detection; detecting intensity changes in image). When an edge detector is applied to an image, you get a lot of edges but these may not be related to each other. To extract objects from an image, need to go from edges to boundaries. Develop a variety of algorithms that group edges to extract boundaries.
* **Scale Invariant Feature Transform (SIFT):** A feature detector which can detect interesting blobs in an image. SIFT features can be used to robustly detect and recognise planar objects in an image, even when scaled, rotated or occluded by other objects.
* Panorama view on smart phones is a technique of feature detection where an algorithm takes a set of overlapping images of a scene and stitches these images together to make one seamless scene.

**Algorithms that recover the 3D structure of a scene from one or more images:**

* **Radiometry:** measuring light.
* **Photometric Stereo:** if a few images of an object are taken under different known lighting conditions, we can compute the surface normal at each point on the object. For continuous surfaces, the measured surface normals can be used to reconstruct the shape of the object.
* **Looking at shape from shading:** More challenging problem where seek to extract 3D shape of a surface from a single shaded image.
* Can recover 3D structure from a scene using both focus and defocus (like a camera).
* **Active Illumination Models:** When controlling for the illumination of objects that are being imaged, can develop ways to recover the shape of objects.
* How does a position in the point of an image (measured in pixels in 2D) relate to the position of the corresponding point in the 3D world? We need to know the parameters of image formation. The process of estimating these parameters is called camera calibration. A single image of a 3D object of known dimensions can be used to figure out the camera parameters.
* **Binocular Stereo:** computing depth using two views (because we use two eyes to perceive the 3D structure around us). Two eyes capture two slightly different views in front of us. The small differences in these two views are used to estimate the depth of each point in the scene.
* Objects are not static but in motion all the time. How does the motion of a point in 3D relate to its motion in the image; concept to be covered. The visible motion of a point in an image is called **Optical Flow**.
* Even without knowing the motion of the camera, the 3D structure of the scene can be captured. In addition to this, we can capture the motion of the camera; this technique is called “structure from motion”.
* **Image Segmentation:** Algorithms that can take an image and segment it into clearly defined regions, where each region more or less corresponds to a single physical object. Group pixels with similar visual characteristics to create regions. Segmentation is an ill-defined task as what exactly constitutes an object often depends on the context.
* **Recognition:** In this topic, the first approach to be covered is appearance matching. Here, many images of an object under different poses and illumination are captured. Dimension reduction is used to compute a compact appearance model of the object from its images. When the object appears in a new image, it is segmented and recognized using the appearance model. The model also reveals the pose and lighting of object.
* **Artificial neural networks for solving complex recognition tasks:** Here, how a neural network can be efficiently trained with back propagation is demonstrated.