

Program: **B. Tech.** Discipline: **CSE**

Semester **VI** Academic Year **2020**

**End Semester** **Assignment**

Course Code: **CS4111** Course Title: **Computer Vision**

Date of Submission: **16/07/2020** Time: **8 PM** Max. Marks: **30**

**Instructions**:

1. **Answer all questions. All questions carry equal marks.**
2. **You can write your answers in this document only by creating space after the questions. Please do not submit PDF of your assignment. Submit only in Doc format.**
3. **Write your answers in bold fonts with a line spacing of 1.15.**
4. **This is an open book assignment. However, you should write your answers in your own words.**
5. **Do not copy the answers either from internet or from each other. If the answers are found to be copied, you would simply be awarded zero.**
6. **Do not share your answers with others. If the answers are found to be same no investigations would be done as to who has copied from whom. Both will be awarded zero.**
7. **If you wish to support your answers by diagrams/figures, either draw them using ‘insert shapes’ option in MS Word or you can draw using pen/pencil on a piece of paper, convert it into an image by taking snapshot with your mobile and paste in between your answers. Do not copy paste from internet/textbook/notes.**
8. **Although there is no word limit for your answers, but try to be logical and to the point while expressing your answers.**
9. **You have to submit your assignment on Moodle by 16th July (Thursday) before 8 PM. No further extension will be given to you to submit.**
10. **No assignment will be accepted on mail.**
11. **It is advised to all of you to finish and submit your assignment well before time in order to avoid any technical glitches at last minute. I repeat, no assignment will be accepted on mail.**
12. **There is no makeup for this assignment.**
13. Explain the process of image formation in a digital camera. What are the factors that play an important role in the image formation process?

**Ans)Image formation encompasses the radiometry and geometric processes by which 2D images of 3D objects are formed. An illuminated object will scatter light towards a lens and the lens will collect and focus the light to create an image. Light on an imaging sensor is usually picked up by an active sensing area.**

**In a digital camera , a CCD(charge-coupled device) can be used for image formation. CCD is an image sensor, and it senses the image intensity values and converts it into an electric signal. The CCD is actually in the shape of an array. It is like a matrix with each cell in the matrix containing a sensor that senses the intensity of a photon. Each sensor of the CCD array itself is an analog sensor. When light strikes the chip , it is held as a small electrical charge in each photo sensor. The response of each sensor is directly equal to the amount of light that strikes on the surface of the sensor. In the transfer phase, the charges are transferred in a kind of bucket brigade until they are passed through the amplifiers, which amplify the signal and then pass it to an analog-to-digital converter.**

**We can also use CMOS(complementary metal-oxide semiconductor) image sensors which are cheaper than CCD and consume lower power for working. It has a very similar working principle to a CCD sensor.**

**Factors that play an important role in image formation process:-**

* **Shutter speed directly controls the amount of light reaching the sensor, and hence determines if images are under or over exposed. Higher shutter speed is usually good for fast moving object but the image will turn out darker, lower speed is good for still objects and gives out a brighter image.**
* **Sampling pitch is the physical spacing between adjacent sensor cells on an imaging chip. Increasing the sampling frequency results in a smaller sampling and pixel pitch, which improves the resolution of the digital image. Decreasing the sampling frequency results in a larger sampling and pixel pitch and decreased resolution.**
* **The fill factor is the active sensing area size as a fraction of available sensing area. Fill factor is the ratio of photodiode area to total pixel area,but the use of microlenses increases the effective fill factor. Higher fill factors result in more light capture and less aliasing.**
* **Larger chip size is better, since each sensor cell can be photo-sensitive.**
* **Analog Gain is a setting that controls the amplification of the signal from the camera sensor. Different environments require adjustments for better quality.**
* **Noise is present from various sources. We need to remove them (as much as possible) to get better results. We can use suitable filters for this job.**
* **Final step is analog to digital conversion. There are various methods with different levels of quality.**
* **Resolution .The higher the better.**

1. Differentiate between Chain code shape representation and 4 and 8-neighbourhood shape representation techniques. According to you, which method is suitable for what kind of images and why?

**Ans) A chain code describes an object by a sequence of unit-size line segments with a given orientation. It encodes a list of connected points lying on a grid using a three-bit code corresponding to the eight cardinal directions (N, NE, E, SE, S, SW, W,NW) between a point and its successor. While this representation is more compact than the original edge list it is not very suitable for further processing.**

**In an 8 neighbour configuration, the spatial distances between the central pixel and its neighbours are not all equal, but in the 4 neighbour configuration they are equal.**

**For images which have large continuous “chains” of the same colour in the image chain code is better, otherwise 4 or 8 neighbour representations are better usually. Neighbour representation is also better for morphological operations too.Chain code is best for monochrome images.**

1. Explain any one technique of extracting shape from shadow in your own words. What are the challenges associated with this technique and how can they be overcome?

**Ans)The surface normal changes across the object in the shaded image, the apparent brightness changes as a function of the angle between the local surface orientation and the incident illumination. This allows us to get a sense of depth in the image. But this is easy for humans only , for a computer this is very hard.**

**Most shape from shading algorithms make large assumptions like:-**

**1)The surface under consideration has a uniform value proportion of the incident light or radiation that is reflected by a surface**

**2)Uniform reflectance of the surface.**

**3)The light source directions are either known or can be calibrated by the use of a reference object.**

**But in real life most objects don’t have uniform albedo and reflectance. Hence it needs to be combined with some other technique or extended in some way to make it useful. One such method is Photometric stereo in which we use multiple light sources that can be selectively turned on and off. For each light source, we have a different reflectance map. Given the corresponding intensities at a pixel, we can recover both an unknown albedo and a surface orientation estimate .From this the surface normals or gradients have been recovered at each pixel, they can be integrated into a depth map using a variant of regularized surface fitting. This although slightly computationally expensive is very effective. For more complicated shapes or specular surfaces more number of light sources may be required.**

1. Explain any one application of Kalman filter in your own words.

**Ans) An application of Kalman Filter is location determination. For example application, consider the problem of determining the precise location of a car. The car can be equipped with a GPS unit that provides an estimate of the position within a few meters. The GPS estimate is likely to be noisy; readings imprecise or fluctuate rapidly due to various factors, though remaining within a few meters of the real position. In addition, since the car is expected to follow the laws of physics, its position can also be estimated by integrating its velocity over time, determined by keeping track of wheel revolutions and the angle of the steering wheel. Typically, it will provide a very smooth estimate of the car's position, but it will drift over time as small errors accumulate. Here we can use Kalman Filter.**

**First we take the old position and modify it according to the physical laws of motion . Not only will a new position estimate be calculated, but also a new covariance will be calculated as well. Next we take a measurement of the car's position from the GPS . Along with this measurement comes some amount of uncertainty, and its covariance relative to that of the prediction from the Physical laws determines how much the new measurement will affect the updated prediction. Using both GPS and physics , the Kalman filter approximates the location of the car in a way that is affected very little by the noisy readings of GPS and imprecise velocity readings.Alternatively, we can take a series of images from a satellite (for example) as a set of sequential observations and merge them using Kalman filtering so as to get a continuous understanding of its position**

1. Suppose there is a dataset which contains images (containing multiple objects) which are categorized into different classes. The categorization has been done on the basis of presence of a particular object (e.g. if a football is present anywhere in the image, it is categorized as football image). Suppose an image is given as a query. The objective is to classify the image in the correct category. Explain your own idea of how you will classify the image in the correct category. Also, justify why the idea proposed by you can produce the best accuracy. (**Note-** Don’t just mention the concepts you will use to classify the image. You must mention full technique- starting from feature extraction followed by construction of feature vector and then recognition/classification methodology)

Ans) **Faster R-CNN is the best method and I'll explain with reasons.**

**In R-CNN the CNN acts as a feature extractor and the output dense layer consists of the features extracted from the image and the extracted features are used as an input to an SVM to classify the presence of the object within that candidate region proposal. To get the candidate region we use selective search algorithm steps to which are:-**

**1. We first generate initial sub-segmentation, and generate many candidate regions in the image.**

**2. Use greedy algorithms (or any algorithm) to recursively combine similar regions into larger ones .**

**3. Use the generated regions to produce the final candidate region proposals.These then become a part of the candidate region.**

**But this part of R-CNN has a drawback of being slow hence we use Faster R-CNN in which a separate neural network is used to predict the candidate region proposals.**

**Steps of Faster R-CNN:-**

**1)Faster R-CNN uses a system of ‘anchors’ or boxes, allowing the operator to define the possible regions that will be fed into the Region Prediction Network. Anchors will slide over the image and skip few pixels at a time allowing more speed.**

**Region Prediction Network- The RPN predicts the possibility of an anchor being background or foreground and refines the anchor or bounding box. The training data of the RPN is the anchors and a set of ground-truth boxes. Anchors that have a higher overlap with ground-truth boxes should be labeled as foreground, while others should be labeled as background. Finally, the output is fed into a Softmax or logistic regression activation function, to predict the labels for each anchor.**

**ROI pooling - The RPN provided proposed regions with different sizes. Each of these is a CNN feature map of a different size. Now the algorithm applies Region of Interest (RoI) pooling to reduce all the feature maps to the same size. It then takes the feature map for each region proposal, flattens it, and passes it through two fully-connected layers with ReLU activation. It then generates a prediction for each of the objects using 2 different fully connected layers.**

**The algorithm has 70% mAp. This technique is very useful in the context of Transfer Learning, especially for training a classifier on a small dataset using the weights of a network trained on a bigger dataset.**

**Problems with R-CNN:-**

**1)It still takes a huge amount of time to train the network as you would have to classify 2000 region proposals per image.**

**2)It cannot be implemented real time as it takes around 47 seconds for each test image.**

**3)The selective search algorithm is a fixed algorithm. Therefore, no learning is happening at that stage. This could lead to the generation of bad candidate region proposals.**

**ALL 3 of the problems of R-CNN are solved with Faster R-CNN. Hence Faster R-CNN is one of the best image recognition algorithms.**