Computer Vision 1

October 25th, 2019, 13.00-16.00

Question 1: Reflection Models

To understand the image formation process, a simple Lambertian reflection model is given by:

$$R = \cos\theta \sum_{\lambda=380}^{780} e(\lambda)\rho(\lambda)f_R(\lambda),\ G = \cos\theta \sum_{\lambda=380}^{780} e(\lambda)\rho(\lambda)f_G(\lambda),\ B = \cos\theta \sum_{\lambda=380}^{780} e(\lambda)\rho(\lambda)f_B(\lambda) \tag{1}$$

where R, G and B are the pixel values, $e(\lambda)$ is the light source, $\rho(\lambda)$ the surface reflectance and $f_R(\lambda), f_G(\lambda), f_B(\lambda)$ are the R, G, B color filters. Further, $\cos\theta = \vec{n} \cdot \vec{s}$ is the angle between the surface normal $\vec{n}$ and light source direction $\vec{s}$.

- (a) What is the color of the reflected light for a white object (i.e. $\rho(\lambda) = 1$)? (1 pt)
- (b) Is the reflected light independent of the viewpoint of the camera? What about the direction of the light source? Please explain. (1 pt)
- (c) What are the conditions to obtain the maximum intensity of the reflected light? (1 pt)
- (d) To correlate the R, G and B values with a standard human observer (human perception), which kind of filter responses would you choose? Why? (1 pt)
- (e) Sketch the spectral power distribution of $e(\lambda)$ for an orange and a purple light source. (1 pt)

Assume narrow-band filters i.e. $f_R(\lambda_{650}) = 1$, $f_G(\lambda_{550}) = 1$, $f_B(\lambda_{400}) = 1$ and 0 elsewhere.

- (f) Prove that $L = \frac{R}{G}$ (at a pixel) is independent of the object geometry and the direction of the light source. (1 pt)
- (g) A simple color invariant is given by R_{x_1}/R_{x_2}, where R_{x_1} and R_{x_2} are the measured red quantities at neighboring positions x_1 and x_2 of a flat surface. It is assumed that the light source is constant over neighboring locations. Show for a flat surface that this color invariant only depends on the surface albedo. (1 pt)

From now on, assume a white object i.e. $\rho(\lambda) = 1$ and $\theta = 0°$.

- (h) For this (narrow-band) reflection model, and assuming a flat, white object (i.e. $\rho(\lambda) = 1$ with $\theta = 0°$), a light source A is given for which the color values are 1 for $400 - 580$ nm and 0.5 elsewhere. Compute R, G and B for A denoted by R_A, G_A and B_A. (1 pt)
- (i) Consider another light source B where color values are 1 for $530 - 780$ nm and 0 elsewhere. Compute R_B, G_B and B_B. (1 pt)
- (j) Compute the intensity values for A and B. Do A and B differ in intensity? Explain numerically. (1 pt)
- (k) Compute the chromaticity values for A and B. Do A and B differ in saturation or hue? Explain in words. (1 pt)

Question 2: Filters and Image Features

Edges and corners are important information cues in images. They are useful for different computer vision tasks such as object recognition and tracking. Below are two types of filters.

3x3 pyramid filter:

$$f = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

3x3 edge filter:

$$g = \begin{array}{|c|c|c|}\hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$

- (a) Calculate the convolution of f and g i.e. $f * g$ where $*$ is the convolution operator. (1 pt)
- (b) What is the difference when an image is convolved by g or by $f * g$? Under which circumstances is $f * g$ preferred over g? (1 pt)
- (c) Is the resulting image different when (1) the image is convolved first by f and then (afterwards) by g, or (2) by convolving it by $g * f$? (1 pt)
- (d) What kind of filter is $f * g$? A low pass, high pass or band pass filter? Why? (1 pt)
- (e) Show that filter g is separable into two (2x2) filters. (2 pts)

Edge classification can be used to detect and classify transitions based on their physical nature. One possible transition type is a shadow one. An example is shown in Table 1 where the R, G and B values of a small image patch P are given containing an intensity (shadow) transition.

$R = G = B =$

10	10	10	10	10
10	10	10	10	10
10	10	20	20	20
10	10	20	20	20
10	10	20	20	20

Table 1: The R, G and B values of a small image patch P containing an intensity (shadow) transition.

- (f) Compute the derivatives f_x and f_y for image patch P (only) for the color channel R using the following simple derivative filters $h_x = \begin{pmatrix} -1 & \underline{1} \end{pmatrix}$ in the x-direction and $h_y = \begin{pmatrix} \underline{1} \\ -1 \end{pmatrix}$ in the y-direction. All elements exceeding the image patches are mirrored. The elements outside the derivative filters are all zero. (1 pt)
- (g) Compute the normalized red ($r = \frac{R}{R+G+B}$) response of the image. Then, compute the derivatives f_x and f_y of this normalized red map. (1 pt)
- (h) Using the results of the two previous questions, how can you classify the transition to be of an intensity (shadow) type? (1 pt)
- (i) Compute the autocorrelation matrix $M_R = \begin{pmatrix} \sum f_x^2 & \sum f_x f_y \\ \sum f_x f_y & \sum f_y^2 \end{pmatrix}$ for image patch P (only) for the red channel R. Then, compute the eigenvalues of M_R. How can they be used to detect a corner? (2 pts)
- (j) Compute the eigenvectors of M_R. What do these eigenvectors mean? (2 pts)

Question 3: Image Classification, Detection and Performance

BoW and ConvNets are useful for image classification and object recognition.

- (a) Consider the task of labeling an image as "containing a car". Is this task known as detection or classification task in object recognition? Explain the difference. (1 pt)
- (b) Summarize the computational steps required to perform image retrieval using Bag-of-Visual Words. (1 pt)
- (c) To compute the codebook for BoW, k-means clustering can be used. If we initialize the k-means clustering algorithm with the same number of clusters but different starting positions for the centers, does the algorithm always converge to the same solution (i.e. codebook)? Why? (1 pt)
- (d)Which of the following sampling strategy is preferred for the BoW representation for image classification/retrieval: (1) dense sampling, or (2) interest point sampling? (1 pt)

- (f) For ConvNets, describe the differences between Conv layer, ReLu, and max pooling layer. (1 pt)
- (g) Assume the input of a particular Conv Layer is 100x100x50. 60 filters with a receptive field of 3x3 are learned. What is the total number of parameters to be learned in this layer? (1 pt)
- (h) The next Conv Layer learns also 100 filters with a receptive field of 5x5. How many parameters are learned in this layer? (1 pt)
- (i) For image classification, consider the following ranking:

Image id	1	2	3	4	5	6	7
Ground truth label	0	1	1	0	1	0	1
Score	0.1	0.9	0.3	0.5	0.8	0.7	0.4

Compute Average Precision for the scoring of the classification system above. (1 pt)

Question 4: Deep Video

You want to design a video classifier using RNN that is able to classify each frame of a video. Suppose each video has 4 frames with 224x224 resolution and the RNN has one layer, no-bias and 20 hidden neurons.

- (a) Design and draw your proposed architecture. (2 pts)
- (b) To determine the end of the generated caption, what is a possible mechanism for a RNN? (1 pt)
- (c) How many parameters does the proposed RNN has in total? (1 pt)
- (d) How can you reduce the number of parameters? (1 pt)
- (e) Discuss the differences and similarities of RNNs and 3D convs for videos. (1 pt)