Computer Vision 1

March 26th, 2018, 09.00-12.00

Question 1: Reflection Models

The simplified Lambertian reflection model, using narrow-band filters, to define the R, G and B pixel values, is given by:

$$R = \cos\theta \ e \ \rho_R, \ G = \cos\theta \ e \ \rho_G, \ B = \cos\theta \ e \ \rho_B \tag{1}$$

where the intensity of a white light source is given by e. ρ_R, ρ_G and ρ_B are the amount of red, green and blue reflected by the surface (i.e. surface albedo). $\cos\theta = \vec{n} \cdot \vec{l}$ is the dot product of the two unit vectors $\vec{n}$ (i.e. surface normal) and $\vec{l}$ (i.e. direction of the light source).

- (a) What type of surfaces, matte and/or glossy, can be described by this Lambertian reflection model (eq. 1)? Why? (1 pts)

  only matte surfaces, there is no specular component to describe glossyness

- (b) Does the viewpoint influence the amount of the reflected light? (1 pts)

  No

- (c) When is the maximum intensity of the reflected light obtained? (1 pts)

  when the light source is perpendicular to the surface normal, yielding the highest dot produc

- (d) What determines the "shading" component in terms of intrinsic image decomposition? (1 pts)

  cos theta

- (e) Under which assumption(s), the light source direction can be calculated? (1 pts)

  if we know the pixel values for a given color, the surface reflectance, the surface normal and the light intensity

- (f) Consider a colored light source $\hat{e}$. Calculate the coefficients a, b, and c of the (diagonal) transformation matrix $\begin{bmatrix} R^e \\ G^e \\ B^e \end{bmatrix} = \begin{bmatrix} \mathrm{a} & 0 & 0 \\ 0 & \mathrm{b} & 0 \\ 0 & 0 & \mathrm{c} \end{bmatrix} \begin{bmatrix} R^{\hat{e}} \\ G^{\hat{e}} \\ B^{\hat{e}} \end{bmatrix}$ to convert pixel values $R^{\hat{e}}, G^{\hat{e}}, B^{\hat{e}}$ recorded under light source $\hat{e}$ to R^e, G^e, B^e values obtained under a canonical (white) light source e. (1 pts)

  a = R^e / R^e^
  b = G^e / G^e^
  c = B^e / B^e^

The simplified dichromatic reflection model, to define the R, G and B pixel values, is given by:

$$R = \cos\theta \ e \ \rho_R + e \ (\cos\phi)^s, \ G = \cos\theta \ e \ \rho_G + e \ (\cos\phi)^s, \ B = \cos\theta \ e \ \rho_B + e \ (\cos\phi)^s \tag{2}$$

where $\cos\phi = \vec{r} \cdot \vec{v}$ depends on ϕ which is the angle between the reflected light $\vec{r}$ and the viewer $\vec{v}$. Further, s is called the specular exponent.

- (g) Explain the mechanism of this new term $(\cos\phi)^s$. How is it used to model the glossy appearance of an object? (2 pts)

see sample exam 2019

- (h) What is approximately the shape of $(\cos\phi)^s$ for different values of s and what is the effect on the size of the specular highlights? (2 pts)

- (i) Show that the color of the highlights is dependent on the color of the light source. (1 pts)

Question 2: Photometric Invariants

Color invariants can be important to recognize objects under varying imaging conditions.

- (a) Considering the dichromatic reflection model (eq. 2), prove that $\frac{2R-B-G}{B-G}$ (at a pixel) is independent of the (intensity) light source, object geometry and the direction of the light source. (2 pts)

- (b) A simple color model is given by $R_{x_1}G_{x_2}/R_{x_2}G_{x_1}$, where R_{x_1}, R_{x_2}, G_{x_1}, G_{x_2} are the measured red and green quantities at neighboring positions x_1 and x_2. Considering the Lambertian reflection model (eq. 1), prove that $R_{x_1}G_{x_2}/R_{x_2}G_{x_1}$ is independent of the object geometry of a homogeneously colored object. (2 pts)

- (c) Would you choose $R_{x_1}G_{x_2}/R_{x_2}G_{x_1}$ or $\frac{2R-B-G}{B-G}$ to recognize the same object under (only) a change in illumination color? Why? (1 pts)

Color invariants become unstable for certain imaging conditions. One way to handle instabilities is by error propagation:

$$\sigma_q = \sqrt{(\frac{\partial_q}{\partial_u}\sigma_u)^2 + ... + (\frac{\partial_q}{\partial_w}\sigma_w)^2} \tag{3}$$

to compute a function $q(u, ..., w)$ for variables $u, ..., w$ and their corresponding uncertainties $\sigma_u, ..., \sigma_w$. Consider a pixel O having the following values $R = 20, G = 40, B = 60$ with $\sigma = 4$.

- (d) Calculate the uncertainty for color model $2R+4B$ at O. Is color model $2R + 4B$ stable? (1 pts)

- (e) Calculate the uncertainty for color model $\frac{R+B}{2G}$. Does this color model become unstable when the intensity is decreasing? Explain. (3 pts)

Question 3: Filters and Image Features

Edges and corners are important features from which image descriptors can be extracted.

- (a) What is the result of the convolution $\begin{bmatrix} \text{a} & \text{b} & \text{c} \\ \text{d} & \text{e} & \text{f} \\ \text{g} & \text{h} & \text{i} \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ for the element at coordinates $(2, 2)$ (that is, the central element) of the resulting image ? (1 pts)

we need to mirror our convolution filter in its centre before applying it (so mirror in x and y direction)
1 2 3
4 5 6
7 8 9
becomes
9 8 7
6 5 4
3 2 1
result = 9a + 8b + 7c + 6d + 5e + 4f + 3g + 2h + 1i

- (b) Given is the following filter $g = \left[\begin{array}{ccc} -1 & \underline{5} & -1 \end{array} \right]$. What is the effect when an image is convolved with g? (1 pts) sharpening, highlighted edges
- (c) Given is the following filter $h = \left[\begin{array}{ccccccc} 1 & 3 & 6 & \underline{7} & 6 & 3 & 1 \end{array} \right]$. What is the effect when an image is convolved with h? What is the name of this filter? (1 pts) this filter smoothens the image along the x axis. its a gaussian filter

h = [1, 1, 1] * [1, 1, 1] * [1, 1, 1]
= [1, 2, 3, 2, 1] * [1, 1, 1]
= [1, 3, 6, 7, 6, 3, 1]

- (d) Show that the filter h is separable into three identical box filters. (2 pts)

e): useful for gentle sharpening?

- (e) Consider the following filter $i = \left[\begin{array}{ccc} -1 & \underline{2} & -1 \end{array} \right]$. Under which circumstances is this a useful operation? (1 pts)
- (f) What is the result of convolving i with h i.e. $i * h$? Which filter, i or $i * h$, do you prefer in case of noisy images? Why? (2 pts)

[-1,-1,-1,2,2,2,-1,-1,-1]
h is preferable for noisy images *assuming* we want to denoise the image (since h blurs out the noise whereas i sharpens it)

Consider the following image patches:

$P =$

0	0	0	0	0
0	0	0	0	0
2	2	2	2	2
2	2	2	2	2
2	2	2	2	2

$Q =$

0	0	0	0	0
0	0	0	0	0
2	2	2	0	0
2	2	2	0	0
2	2	2	0	0

Intensity values of two small image patches P and Q.

- (g) Compute the derivatives f_x, f_y and the gradient magnitude of image patches P and Q using a simple derivative filter $h_x = \left(\begin{array}{cc} -1 & \underline{1} \end{array} \right)$ in the x-direction and $h_y = \left(\begin{array}{c} \underline{1} \\ -1 \end{array} \right)$ in the y-direction. All elements exceeding the image patches are mirrored. The elements outside the derivative filters are all zero. (2 pts)
- (h) Compute the autocorrelation matrix $M = \left(\begin{array}{cc} \sum f_x^2 & \sum f_x f_y \\ \sum f_x f_y & \sum f_y^2 \end{array} \right)$ for image patches P and Q. (1 pts)

(i): if lambda1 >> lambda 2 or vice versa, we likely have an edge.
If lambda1 = lambda 2 = small, the area is homogeneous
if lambda1 = lambda 2 = large indicates a corner

- (i) Compute the eigenvalues of M for image patch Q. How can these eigenvalues be used to determine a corner? (3 pts)
- (j) Compute the eigenvectors of M for image patch Q. What do these eigenvectors mean? (3 pts) (j) the eigenvectors will be parallel to the image gradients

Question 4: Object Classification, Detection and Performance

Deep learning and ConvNets are very useful for object recognition and detection. Consider the following four (simple) image patches of the letters I, L, O and X:

$I =$

0	1	0
0	1	0
0	1	0

$L =$

0	1	0
0	1	0
0	1	1

$O =$

0	1	0
1	0	1
0	1	0

$X =$

1	0	1
0	1	0
1	0	1

Intensity values of four image patches of the letters I, L, O and X.

- (a) After training a single layer neural network (multi-class), the following weight matrix $\vec{M} = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0 & 1 & 0 & 0 & 0.5 & 1 \\ 0 & 0 & 0 & 0.5 & 0 & 1 & 0 & 0.5 & 0 \\ 0.9 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0.9 \end{bmatrix}$ is obtained (i.e. final weight parameters). Bias $\vec{b}$ is not considered. Logit z for each class j is given by $z_j = \vec{M} \cdot \vec{x}$, where input $\vec{x}$ is an image (e.g. I, L, O and X) expressed in vector form (i.e. all pixel values are ordered from top-left to bottom right). Compute output y_i using a softmax layer i.e. $y_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$ for images I and X. What can you conclude about the prediction? (3 pts)

- (b) To tackle the image classification problem using fully connected neural networks is a daunting challenge. Why? (1 pts)

b) Lots more parameters for fc networks!

ConvNets have neurons arranged in three dimensions. So layers have a width, height and depth. Neurons are connected to a small, local region of the preceding layer.

- (c) Consider a 32x32x5 image and 10 filters with an extent of 5 (5x5 kernel). What is the depth of the 10 filters? If you use no padding and stride $s = 1$, what will be the output size of the activation map? (2 pts)

c) filter depth: 5
output size
= (h_i + 2 * p - k) / s + 1
= (32 + 2*0 - 5 / s + 1
= 27 / 1 + 1
= 28x28 output size
output volume =
28 x 28 x 10

- (d) What are the different layers in a standard CNN? What are the different layers in AlexNet? (1 pts)

- (e) What is max pooling and when is it useful? (1 pts)

e) max-pool: provides high-level
abstraction from the data
computationally efficient
as it downsamples,
meaning less parameters
in subsequent layers

Object detection is used to determine the location of objects in images.

- (f) What are the basic components of the pipeline used in window-based object detection? (2 pts)

- (g) What are the advantages and disadvantages of a sliding-window approach? (1 pts)

g) advantage: simple to implement
disadvantage: expensive;
lots of redundant computaiton

Consider two recognition systems: S_1 and S_2. The number of relevant images for search query Q_1 is 4 and is composed of the following set (A, B, C, D). The order of the 10 highest ranked images of systems S_1 and S_2 for query Q_1 is:

S_1	S_2
1. A	1. O
2. B	2. M
3. L	3. A
4. C	4. B
5. P	5. K
6. D	6. C
7. Q	7. D
8. O	8. M
9. S	9. S
10. E	10. L

- (h) Generate the precision-recall graph for the systems S_1 and S_2 for Q_1. What can you conclude from the precision-recall graphs? (2 pts)