ECE558 Homework 05 (100 points in total)

Due 12/05/2019

How to submit your solutions: put your report (word or pdf) and results images (.png) if had in a folder named [your_unityid]_hw05 (e.g., twu19_hw05), and then compress it as a zip file (e.g., twu19_hw05.zip). Submit the zip file through **moodle**.

If you miss the deadline and still have unused late days (we changed it to 0.5-day based metric, please check your email for the notice), please send your zip file to TAs and me.

Important Note: We will **NOT** accept any replacement of submission after deadline, even if you can show the time stamp of the replacement is earlier than the deadline. So, **please double-check if you submit correct files**.

You can still use your late days if had and needed.

Challenge: please try to finish the first version of this hw in 3 hours (as a final test for yourself), and then continue to work on it before deadline. Keep the first version for your own reference only and check how much you gain after we release the reference solution.

Problem 1. (10 points). Please select T/F (True or False) for the following statements.

- [F] if available hardware, or software routines, have only the capability to perform the DFT, we can use it to compute the inverse DFT using $f[x] = \frac{1}{M}\overline{DFT\{F[\mu]\}}$.
- [T] In either domain, time or frequency, the function is not periodic, then the argument in the other domain runs continuously. If in either domain, time or frequency, the function has a discrete argument, then the transformed function in the other domain is periodic.
- [F] Fourier spectrum carry much of the information about where discernable objects are located in an image.

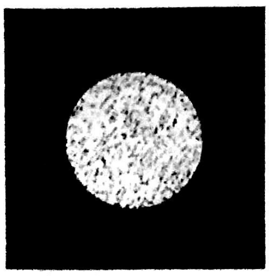
- [F] To construct a Gaussian image pyramid for a given image, we first down-sample it using bilinear interpolation method, and then apply a Gaussian filter to smooth potential artifacts introduced by the down-sampling.
- [T] Bag-of-SIFT descriptors is an effective model capturing rough layout/configuration information of scene or objects in an image due to the nice properties (scale and orientation invariance) of SIFT descriptors.
- [T] To improve accuracy performance in test and avoid overfitting, we should not always seek the classifier whose training error is zero (if possible).
- [F] To handle wraparound error in filtering in frequency domain, we need to pad a given input image before filtering. Either centered padding or left-top-based padding works.
- [T] Image restoration is usually posed as an objective process which utilizes a criterion of goodness that will yield an optimal estimate of the desired result.
- [F] An edge can be caused by depth discontinuity or surface color discontinuity. Corner locations are co-variant w.r.t. translation, rotation and scaling.
- [F] For the Canny edge detector, we will obtain more and stronger edges if we use larger Gaussian kernel size.

Problem 2. (10 points).

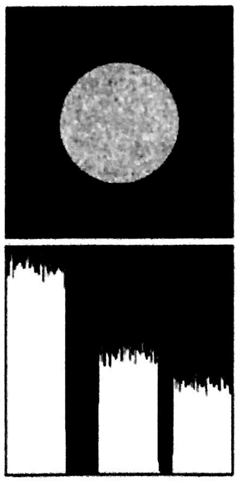
- Select the noise type from (a)~ (f) which describe the noise best in each figures:

(a) $p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\bar{z})^2/2\sigma^2}$, (b) $p(z) = \begin{cases} \frac{2}{b}(z-a)e^{-(z-a)^2/b} & \text{for } z \geq a \\ 0 & \text{for } z < a \end{cases}$, (c) $p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!} e^{-az} & \text{for } z \geq 0 \\ 0 & \text{for } z < 0 \end{cases}$

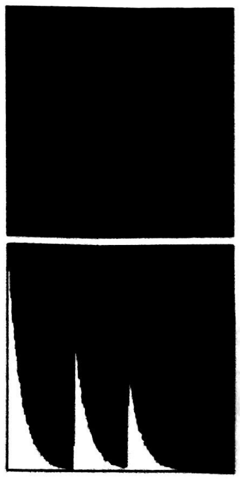
(d) $p(z) = \begin{cases} ae^{-az} & \text{for } z \geq 0 \\ 0 & \text{for } z < 0 \end{cases}$, (e) $p(z) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq z \leq b \\ 0 & \text{otherwise} \end{cases}$, (f) $p(z) = \begin{cases} P_a & \text{for } z = a \\ P_b & \text{for } z = b \\ 0 & \text{otherwise} \end{cases}$



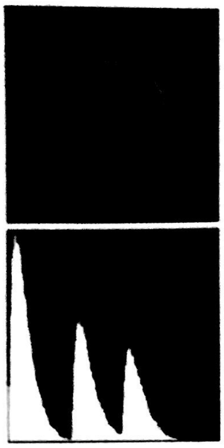
Noise type [f]



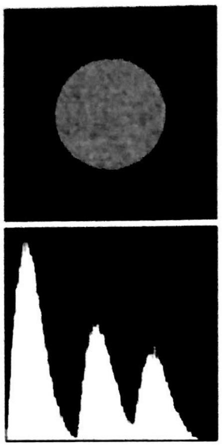
Noise type [e]



Noise type [d]



Noise type [c]



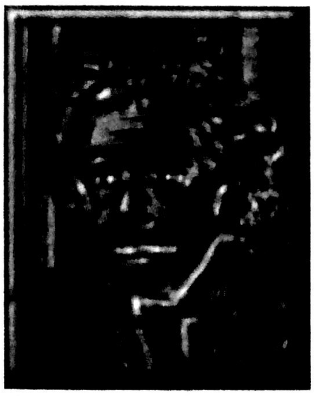
Noise type [b]

- To find $g[x,y]$  in a given image $f[x,y]$  or , select the best method from (a)~(d) which generate the result images $h[x,y]$. Let $\bar{g}$ be the mean value of $g[x,y]$.

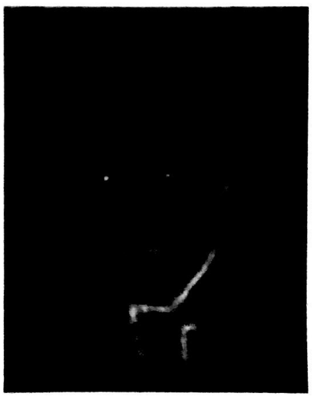
(a) $h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$, (b) $h[m,n]=\sum_{k,l}\{g[k,l]-\bar{g}\}f[m+k,n+l]$

(c) $h[m,n]=\sum_{k,l}(g[k,l]-f[m+k,n+l])^2$, (d) $h[m,n]=\dfrac{\sum_{k,l}(g[k,l]-\bar{g})(f[m-k,n-l]-\bar{f}_{m,n})}{\left(\sum_{k,l}(g[k,l]-\bar{g})^2\sum_{k,l}(f[m-k,n-l]-\bar{f}_{m,n})^2\right)^{0.5}}$

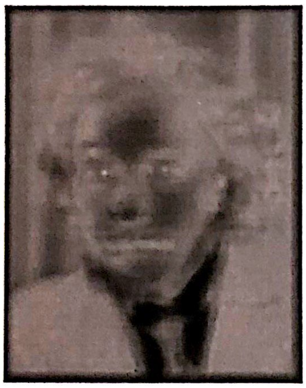
The result of [a]



The result of [d]



The result of [b]



The result of [c]

The result of [d]

Problem 3. (10 Points)

- Find one instance for each of the four types of edges in the given image: (a) illumination discontinuity, (b) depth discontinuity, (c) surface normal discontinuity, and (d) surface color discontinuity. Draw a small circle for each instance with the type (a)~(d) indicated in the circle.

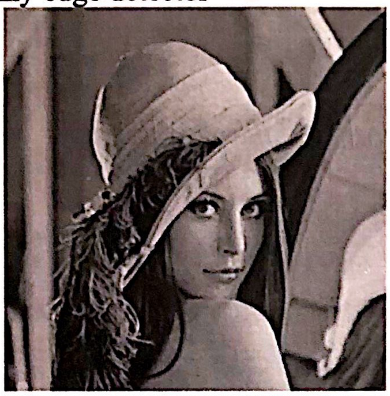
(a) illumination discontinuity

(b) depth discontinuity

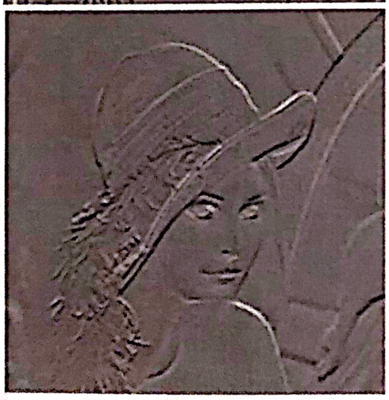
(c) surface normal discontinuity

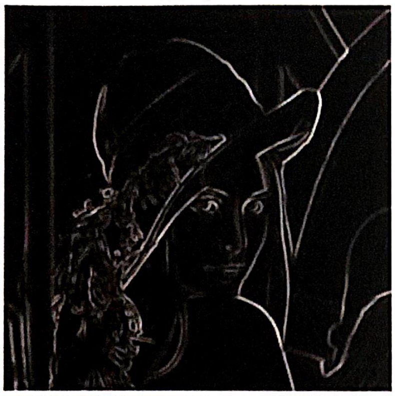
(d) surface color discontinuity

- Write the operation name (e.g., "x-derivative of Gaussian") for each of the steps of the Canny edge detector

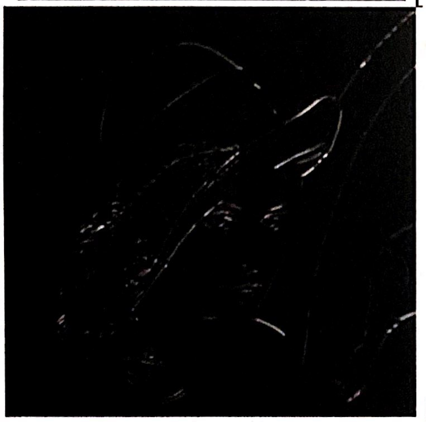
a)  [input image]

b)  [X- Derivative of Gaussian]

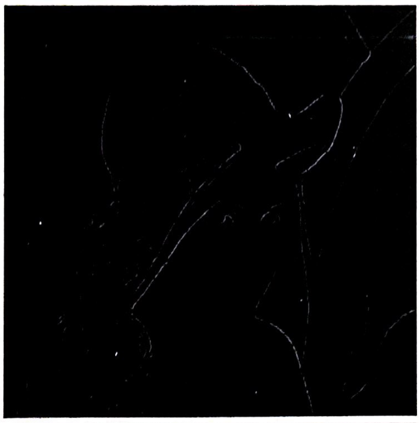
c)  [Y- Derivative of Gaussian]



d) [Gradient Magnitude]

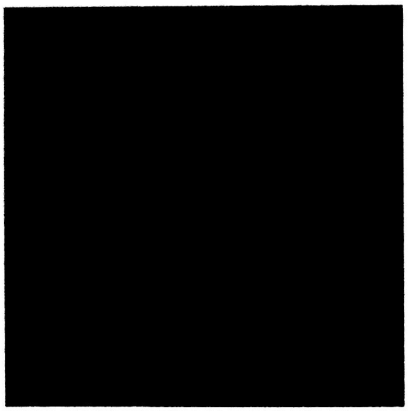


e) [get orientation that $\alpha = \mathrm{atan2}(g_y, g_x)$]



f) [After non-max suppression]

g) [Hysteresis thresholding]

h)  [Final Canny Edges]

Problem 4. (10 points). Consider a linear, position-invariant image degradation system with impulse response

$$h(x-\alpha, y-\beta) = e^{-[(x-\alpha)^2+(y-\beta)^2]}$$

Suppose that the input to the system is an image consisting of a line of infinitesimal width located at $x = a$, and modeled by $f(x,y) = \delta(x - a)$, where δ is an impulse. Assuming no noise, what is the output image $g(x,y)$?

Solve:

Due to the impulse response of the system

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)h(x-\alpha, y-\beta)\,d\alpha\, d\beta \quad \cdots (1)$$

Because of $f(x,y) = \delta(x-a)$

So (1) is equal to $g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\alpha - a) h(x-\alpha, y-\beta)\,d\alpha\, d\beta$

$$= \int_{-\infty}^{\infty} \delta(\alpha - a) e^{-(x-\alpha)^2} d\alpha \int_{-\infty}^{\infty} e^{-(y-\beta)^2} d\beta$$

$$= e^{-(x-a)^2}\int_{-\infty}^{\infty} e^{-(y-\beta)^2} d\beta \quad \cdots (2)$$

I note that $\int_{-\infty}^{\infty} e^{-[(y-\beta)^2]} = \sqrt{2\pi\cdot\frac{1}{2}}\left[\frac{1}{\sqrt{2\pi\cdot 1/2}} e^{-\frac{1}{2}\frac{(\beta-y)^2}{1/2}}\right]$ is the form of a constant times a Gaussian density with variance $\sigma^2 = 1/2$

So $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi 1/2}} e^{-\frac{1}{2}\frac{(\beta-y)^2}{1/2}} d\beta = 1$

The (2) is equal to $g(x,y) = \sqrt{\pi}\, e^{-(x-a)^2}$

is a blurred version of the original image

Problem 5. (10 points). Compute the Fourier transform of the 2-D continuous sine function $f(x,y) = A\sin(\mu_0 x + v_0 y)$ and show it is the pair of conjugate impluses. (Hint: express the sine in terms of exponentials)

Solve: due to $\sin X = \frac{e^{jX} - e^{-jX}}{2j}$

So $f(x,y) = A\sin(u_0X + v_0y) = A\,\frac{e^{j(u_0X+v_0y)} - e^{-j(u_0X+v_0y)}}{2j}$

$$F(u,v) = \frac{A}{j2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[e^{j(u_0X+v_0y)} - e^{-j(u_0X+v_0y)}\right]e^{-j2\pi(uX+vy)}\,dx\,dy$$

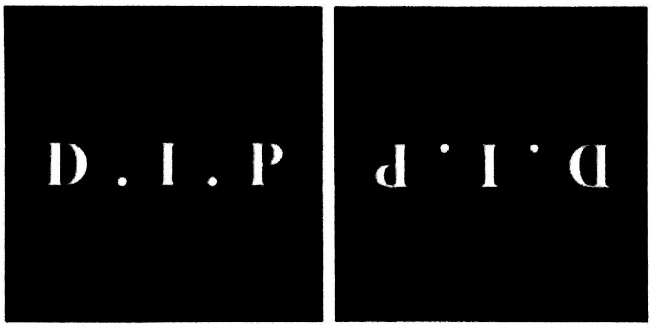
$$= \frac{A}{j2}\int_{-\infty}^{\infty} e^{-(-ju_0+j2\pi u)X}\,dx\int_{-\infty}^{\infty} e^{-(-jv_0+j2\pi v)y}\,dy$$

$$- \frac{A}{j2}\int_{-\infty}^{\infty} e^{-(ju_0+j2\pi u)X}\,dx\int_{-\infty}^{\infty} e^{-(jv_0+j2\pi v)y}\,dy$$

$$= \frac{A}{j2}\cdot 2\pi\,\delta(2\pi u - u_0)\cdot 2\pi\,\delta(2\pi v - v_0) - \frac{A}{j2}\cdot 2\pi\cdot\delta(2\pi u + u_0)\cdot 2\pi\cdot\delta(2\pi v + v_0)$$

$$= -j2\pi^2 A\left[\delta(2\pi u - u_0,\ 2\pi v - v_0) - \delta(2\pi u + u_0,\ 2\pi v + v_0)\right]$$

Problem 6. (10 points, reused from midterm, and the grading will be relatively strict w.r.t. the reference solutions). Consider the images shown. The image on the right was obtained by: (a) multiplying the image on the left by $(-1)^{x+y}$; (b) computing the DFT; (c) taking the complex conjugate of the transform; (d) computing the inverse DFT; and (e) multiplying the real part of the result by $(-1)^{x+y}$. Explain (mathematically) why the image on the right appears as it does. Hint: derive the process step by step from (a) to (e)



Solve: Let the left image be $f(x,y)$

Step a: $f(x,y)\cdot(-1)^{x+y}$

Step b: $DFT[f(x,y)\cdot(-1)^{x+y}] \longrightarrow F(u-M/2, v-N/2)$

due to $f(x,y)(-1)^{x+y} \Leftrightarrow F(u-M/2, v-N/2)$

Step c: Because of $F(u-u_0, v-v_0) \Leftrightarrow f(x,y)(-1)^{x+y} e^{j2\pi(u_0x/M + v_0y/N)}$

$\therefore F(u-M/2, v-N/2) \Leftrightarrow f(x,y)e^{j2\pi(\frac{x}{2}+\frac{y}{2})}$

Step d: Due to $F^*(u,v) = [f(M-x, N-y)]^*$

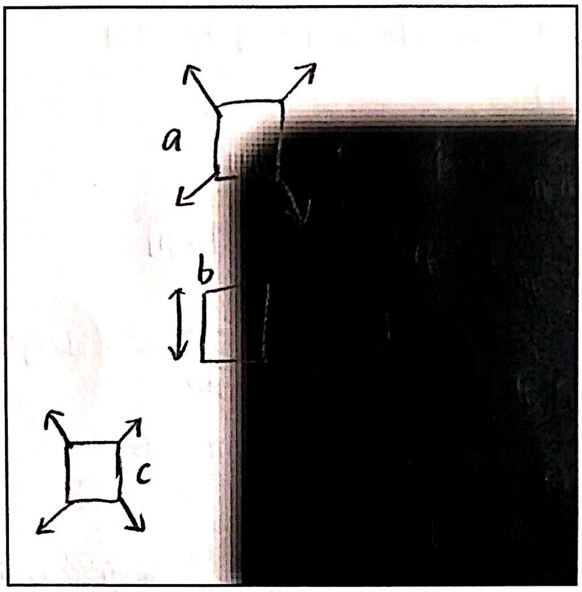
So $[f(M-x, N-y)e^{-j2\pi(\frac{x}{2}+\frac{y}{2})}]^* = f(M-x, N-y)e^{j2\pi(\frac{x}{2}+\frac{y}{2})}$

Step e: $f(M-x, N-y)\cos(\pi x+\pi y)\cdot(-1)^{x+y}$

when x and y both odd, the output image is $f(M-x, N-y)$
when x and y both even, the output image is $f(M-x, N-y)$
when x odd and y even, the output image is $f(M-x, N-y)$
when x even and y odd, the output image is $f(M-x, N-y)$

Problem 7. (15 points).

- Show the basic idea of detecting corners using the given toy image.



Solve:

In order to detect corners, we are supposed to easily recognize the point by looking through a small window. If we shift a window in any directions and then a large change in intensity appears, there'll be corners that we want. For example, c is the "flat" region due to no change in all directions; b is the "edge" due to no change along the edge direction; however, a is the "corner" due to significant change in all directions.

- Show why the second moment matrix is important in detecting corners using detailed derivation. ***Define your nations and explain steps in the derivation.***

Solve: Change in appearance of window w for the shift $[u,v]$

$$E(u,v) = \sum_{(x,y)\in w} [I(x+u, y+v) - I(x,y)]^2 \cdots (1)$$

We want to find out how this function behaves for small shifts.

due to first-order Taylor approximation for small motions $[u,v]$

$$I(x+u, y+v) \approx I(x,y) + I_x \cdot u + I_y \cdot v \cdots (2)$$

Substitute (2) into (1):

$$E(u,v) = \sum_{(x,y)\in w} [I(x+u, y+v) - I(x,y)]^2$$

$$\approx \sum_{(x,y)\in w} [I(x,y) + I_x \cdot u + I_y \cdot v - I(x,y)]^2$$

$$= \sum_{(x,y)\in w} [I_x \cdot u + I_y \cdot v]^2 = \sum_{(x,y)\in w} (I_x^2 u^2 + 2 I_x I_y u \cdot v + I_y^2 \cdot v^2)$$

The quadratic approximation can be written as

$$E(u,v) \approx [u \; v]\, M \begin{bmatrix} u \\ v \end{bmatrix}$$

where M is a second moment matrix computed from image derivatives:

$$M \begin{bmatrix} \sum_{x,y} I_x^2 & \sum_{x,y} I_x \cdot I_y \\ \sum_{x,y} I_x \cdot I_y & \sum_{x,y} I_y^2 \end{bmatrix}$$ The sums are over all the pixels in the window w

Therefore, the M matrix plays an important role in detecting corners

The surface $E(u,v)$ is locally approximately by a quadratic form.

Consider the axis-aligned case (gradients are either horizontal or vertical)

$$M = \begin{bmatrix} \sum_{x,y} I_x^2 & \sum_{x,y} I_x \cdot I_y \\ \sum_{x,y} I_x \cdot I_y & \sum_{x,y} I_y^2 \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

In order to find corners, we need to check whether a and b are both large

Problem 8. (10 points). Write down the steps of Lowe's SIFT algorithm. **Your steps should include the SIFT point detection and the SIFT descriptor generation.** Explain why SIFT keypoints are scale invariant to certain degree, and why SIFT descriptor is robust w.r.t. scale changes, illumination changes, orientation changes and viewpoint changes up to certain degree.

Solve:

Run DoG detector
- Find maxima in location/scale space
- Remove low-contrast and edge points

↓

Find all major orientations
- Bin orientations into 36 bin histogram
  - Weight by gradient magnitude
  - Weight by distance to center (Gaussian-weighted mean)
- Return orientations within 0.8 of peak
  - Use parabola for better orientation fit

↓

For each (x, y, scale, orientation), create descriptor:
- Sample 16×16 gradient mag. and rel. orientation
- Bin 4×4 samples into 4×4 histograms
- Threshold values to max of 0.2, divide by L2 norm
- Final descriptor: 4×4×8 normalized histograms

A SIFT keypoint is a circular image region with an orientation. It's described by a geometric frame of four parameters: the keypoint center coordinates x and y, its scale (the radius of the region), and its orientation (an angle expressed in radians) → makes SIFT keypoints scale invariant to certain degree

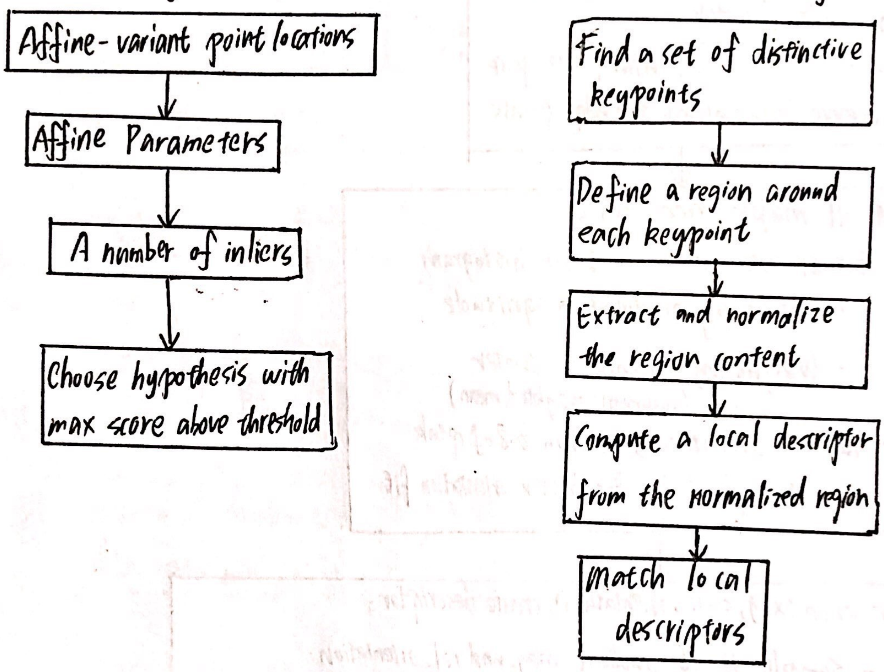
Also, the SIFT descriptor: $\boxed{\text{Image gradients}} \rightarrow \boxed{\text{keypoint descriptor}}$

The histogram of oriented gradients can { capture important texture information; be robust to small translations / affine deformations } → makes SIFT descriptor robust with respect to scale changes, illumination changes, orientation changes and viewpoint changes up to certain degree

Problem 9. (15 points). Draw the workflows and explain the workflows for the following three tasks.

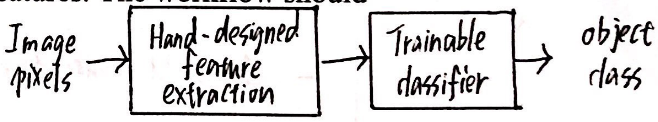
a) Instance-based (object or image patch) matching or recognition.

Solve: Instance-based recognition: Instance-based matching:



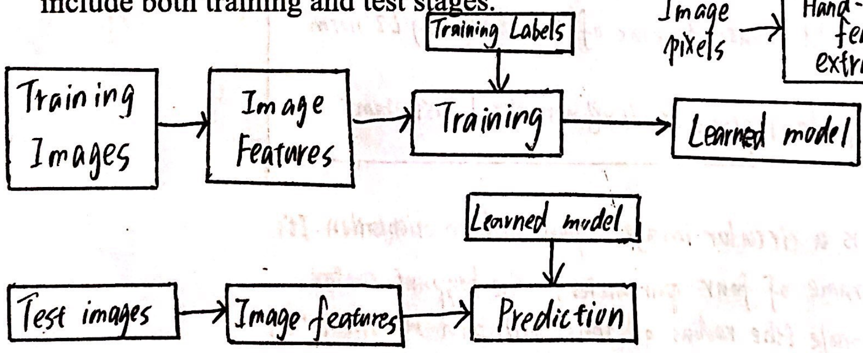
b) Image / object categorization with hand-crafted features. The workflow should include both training and test stages.

Solve:



Training:

Testing:



Apply a prediction function to a feature representation of the image to get the desired output,

$$y = f(x)$$

output — prediction function — Image feature

Training: given a training set of labeled examples $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, estimate the prediction function f by minimizing the prediction error on the traing set.

Testing: apply f to a never before seen test example x and output the predicted value $y = f(x)$

c) RANSAC algorithm for line fitting.

Solve:

Random sample consensus (RANSAC): Very general framework for model fitting in the presence of outliers

Choose a small subset of points uniformly at random

↓

Fit a model to that subset

↓

Compute error function

↓

Find all remaining points that are "close" to the model and reject the rest as outliers

↓

Do this many times and choose the best model