Computer Vision and Deep Learning

Exam 1

2021.12.23

1. **(20%, Camera Model)** Figure 1 shows the pinhole camera model.
   1) Please write the 4x4 transformation matrix (homography matrix) H for the relationship between a point $W(X, Y, Z)$ in the 3D world coordinate and the corresponding point $C(x, y, z)$ in the 3D camera coordinate. This transformation matrix will include the extrinsic parameters R (rotation) and T (translation). Please use $C(4x1)=H(4x4)W(4x1)$ format. **(5%)**
   2) Please write the 3x3 transformation matrix A for the relationship between a point $C(x, y, z)$ in the 3D camera coordinate and the corresponding point $I(u, v)$ in the 2D image coordinate. This matrix consists of 5 intrinsic parameters, $(\alpha, \beta), \gamma, (u_0, v_0)$. Please use $\lambda I(3x1)=A(3x3)C(3x1)$ format. **(5%)**
   3) Based on 2) and Homography matrix A, please simplify to the equation form to $u =$ ———, $v =$ ———. **(5%, if only one is correct, then 3%)**
   4) Matrix H is also called _____? Matrix A is called _____? **(5%)**
   (Hint: Bilinear Transform, 8-Parameters Plane Transform, Affine Transform, Projection Transform, Parallel Transform)

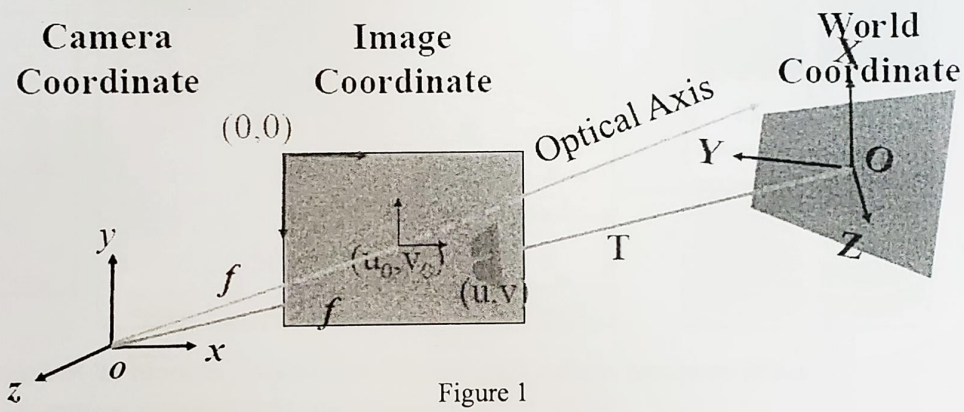


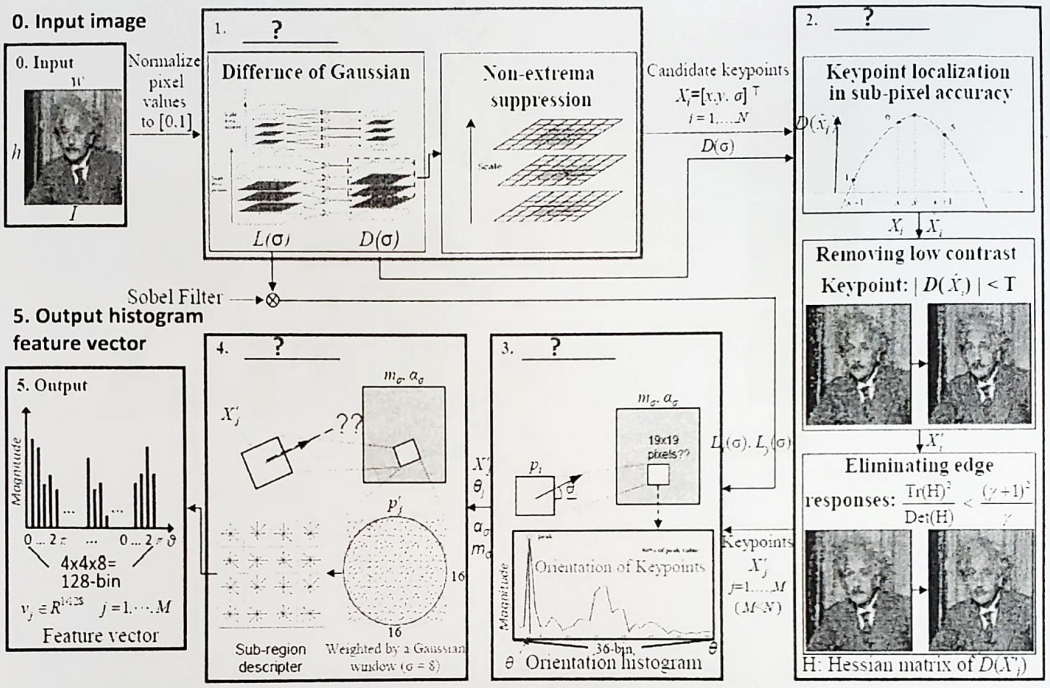
Figure 1

2. **(16%, Camera Calibration)**
   1) For multi-dimensional Gaussian model in Zhang's camera calibration paper, the exponential term is $\sum_i (m_i - \tilde{m}_i)^T \Lambda_{m_i}^{-1} (m_i - \tilde{m}_i)$, where m is the locations of corner detection points and $s\tilde{m} = A[R\ t]\tilde{M}$,
   This term is called _____ **(5%)**. It is for similarity measure.
   (Hint: Euclidean distance, Manhattan distance, Mahalanobis distance, or Kullback–Leibler distance)
   2) Assume $\Lambda_{m_j} = \sigma^2 I$ for all i. Then max MAP problem becomes
   $min_H \sum_i \| m_i - \hat{m}_i \|^2$.
   We can solve this nonlinear minimization by using _____? **(5%)**
   (Hint: Closed-form solution, Pseudo inverse, Singular Value Decomposition, Levenberg-Marquardt algorithm, The first order Taylor series expansion)
   3) Based on the camera calibration and AR programming at homework, please order

following function calls in digits. (**Hint:** Only 3 functions are needed.) **(6%)**

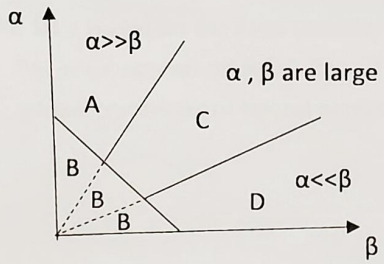
(1) cv2.calibrateCamera()
(2) cv2.GetPerspectiveTransform()
(3) cv2.projectPoints()
(4) cv2.warpPerspective()
(5) cv2.findChessboardCorners()

3. **(16%, SIFT)** For the SIFT (Scale Invariant Feature Transform) procedure as following framework:



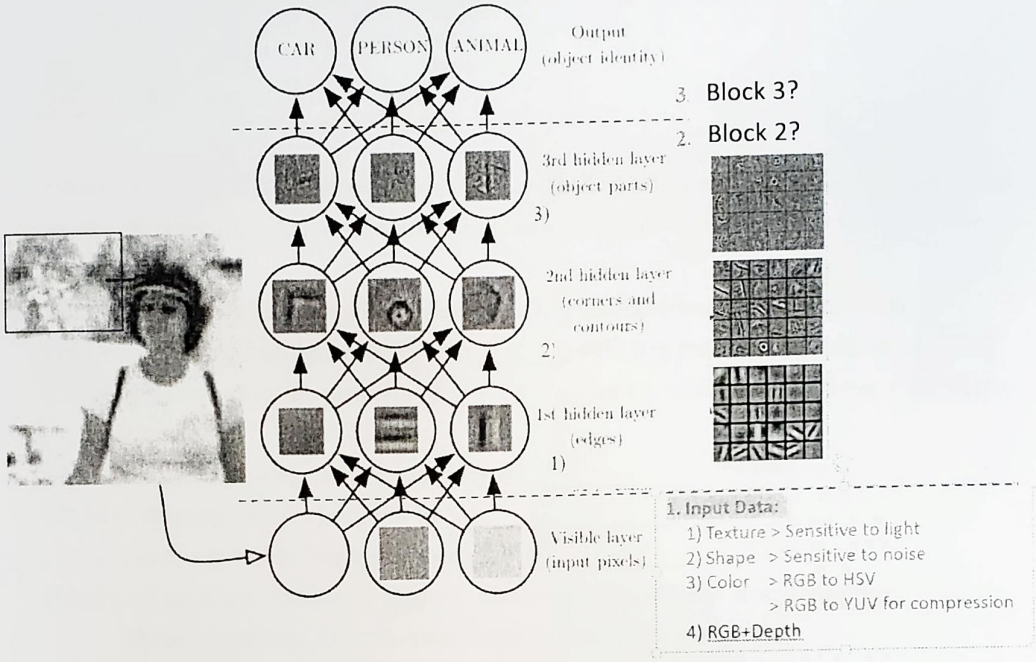
1) Please sort the block in the digit order (from 1. to 4.) of this procedure **(8%):**
   (1) Orientation assignment for each keypoint,
   (2) Histogram equalization
   (3) Scale-space extrema detection,
   (4) Keypoint descriptor,
   (5) Unit vector normalization
   (6) Keypoint localization,

2) To evaluate a feature patch in SIFT, if there is a 2x2 Hessian matrix, and α and β are its corresponding eigenvalues. Based on following α-β figure:



Regions A, B, C, and D correspond to which feature, individually? (1) Flat Feature, (2) Edge Feature, (3) Corner Feature? (Please just write the digit 1, 2, or 3 for each region) A)_____, B)_____, C)_____, D)_____. **(8%)**

4. **(20%, DL)** For the basic CNN (convolutional neural network) structure as following,



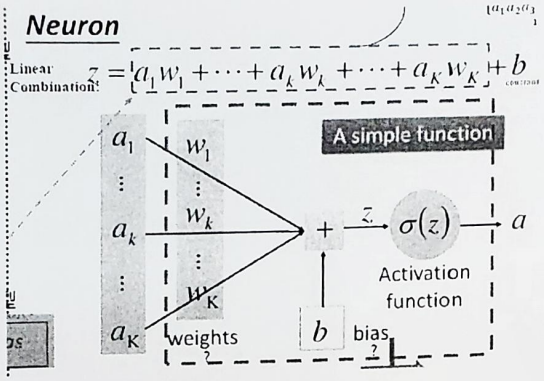
1) For the layer blocks, block 2 and block 3, belong to which property, respectively?
(1) Block 2_____? (2) Block 3 _____?
(Hint: AdaBoost, Classification, Supported Vector Machine, Feature extraction, Vector quantization) **(6%)**

2) For layers 1), 2) and 3), each layer has the property of? (1) Layer 1) _____? (2) Layer 2) _____? (3) Layer 3) _____?
(Hint: Local feature, Medium feature, Global feature, Spectrogram feature, Invariant feature) **(9%)**

3) For AlexNet, VGG16 and ResNet,
Please sort them based on the total number of layers from shallow to deeper? **(5%)**

5. **(20%, DL)** Based on the deep learning lecture:

1) For each neuron as following, if a_i is the given (known) input image pixel, w_i is the weight parameter of neural network, b is the bias parameter, and z is it output result.



Please write its ***WX=Z*** format? Here ***W*** is 1x(K+1)-dimensional unknown parameter vector, ***X*** is the (K+1)x1-dimensional known input pixel vector and ***Z*** is the (1x1) output result. **(5%)**

2) Each of following answers for physical meaning has only one answer selection:

(1) Deep learning has the property of _____, which is the same as AdaBoost, **(3%)**

(2) Deep learning has the property of _____, which is the same as Supported Vector Machine **(3%)**

(3) Convolution process has the property of _____? **(3%)**

(4) Max Pooling has the property of _____? **(3%)**

(5) Activation function has the property of _____? **(3%)**

(Hint: of physical meaning: Feature extraction, Non-linear discrimination, Cascade, Down-sampling, Output normalization, Batch Normalization)

6. **(8%)** For this class so far, please write your suggestions to professor Lien, Jenn-Jier James 連震杰 to improve his lecture? (At least 30 words, written in either English or Chinese. 用中文寫就好)

1) Positive site (Pros.): **(4%)**

2) Negative site (Cons.): **(4%)**