Solutions to Chapter 1

Prob. 1.1 — Define image processing and image analysis.

Answer (Prob. 1.1) — Image processing is defined in this book as the field of study in which algorithms process input images to produce output images, whereas image analysis is the field of study in which algorithms process input images to produce other types of (typically higher-order) data.

Prob. 1.2 — Even though machine vision and computer vision are nearly synonymous, there are some subtle distinctions between them. List at least two of these differences.

Answer (Prob. 1.2) — Machine vision typically deals with images of a 2D scene captured in an industrial environment in which the lighting, camera placement, and objects being viewed are carefully controlled. Computer vision relaxes these restrictions and deals with images of generic, everyday, unconstrained scenes.

Prob. 1.3 — Image analysis, as defined in this book, is very closely related to computer vision. What is the key difference?

Answer (Prob. 1.3) — Computer vision is usually focused upon images obtained by an optical camera, whereas image analysis is a broader term (as used in this book) that encompasses the processing of any type of image. However, the two terms can, for the most part, be used interchangeably.

Prob. 1.4 — Image processing, as defined in this book, produces an output image from an input image. What are the two primary purposes for such output images?

Answer (Prob. 1.4) — The images produced by an image processing algorithm are either displayed to a human viewer or serve as input to a computer vision (or image analysis) algorithm.

Prob. 1.5 — Another way to categorize the information in this book would be in terms of low-, mid-, and high-level vision. Explain how you would map image processing, image analysis, machine vision, and computer vision into these alternative categories.

Answer (Prob. 1.5) — Image processing generally deals with low-level processing, machine vision deals with mid-level processing, and the other two areas deal with both mid-level and high-level processing.

Prob. 1.6 — List three basic image processing problems, and three basic problems in image analysis.

Answer (Prob. 1.6) — Three basic problems in image processing are enhancement, restoration, and compression. Three basic problems in image analysis are segmentation, classification, and shape from X.

Prob. 1.7 — Skim the table of contents to identify at least one topic for each of the six basic problems mentioned in the previous question (list the chapter and/or section number for each, along with the title). Can you identify a topic that overlaps more than one basic problem? Can you identify a topic that does not fit into any of the basic categories?

Answer (Prob. 1.7) — Answers may vary, but some examples follow (chapter and section numbers can be found in the table of contents). *Thresholding* can be viewed not only as segmentation but also as classification, since the goal is to classify each pixel as being either above the threshold (usually indicating the foreground) or below the threshold (background). Similarly, *stereo correspondence* and *optical flow* can be viewed either as a form of classification in which a set of pixels in one image is compared with a set of pixels in another image to determine whether they belong to the same category, or as segmentation. If the segmentation were known, these problems would reduce to simple model fitting, using relatively straightforward optimization techniques that have been known for centuries, in order to estimate the disparity or displacement function. Similarly, determining the boundary of an object† can be viewed as either a classification problem — because the boundary defines which pixels are inside the object and which are outside, or as a segmentation problem — because the boundary separates the object from the rest of the scene. *Image mosaicking* can be seen as either restoration, with captured images being cropped version of the desired original, or as enhancement, with the final image being a more complete view of the scene than was available in the individual images. Two topics that do not fit any of the basic categories are calibration and albedo estimation.

Prob. 1.8 — Explain the statement, "Computer vision is the inverse of computer graphics."

Answer (Prob. 1.8) — Computer graphics is the problem of generating an image viewable by a human from a 3D model of the world stored in the computer, whereas computer vision is the problem of inferring some model of the world from an image that is generally viewable by a human.

Prob. 1.9 — The three main problems in machine learning are unsupervised learning, supervised learning, and reinforcement learning. Relate any two of these to the main problems in image analysis.

Answer (Prob. 1.9) — Unsupervised learning is closely related to image segmentation, while supervised learning is closely related to classification. Reinforcement learning is sometimes used in interactive or tracking systems.

Prob. 1.10 — Provide the make and model of an automobile that processes images from one or more cameras permanently mounted on the vehicle, and explain the purpose of the processing. Search the Web if needed.

Answer (Prob. 1.10) — Automobile manufacturers are increasingly installing cameras in vehicles. Examples include Infiniti's advanced lane departure warning system, Infiniti and Nissan's around view monitor, Chrysler's smart headlight detection, Lexus' intelligent parking assist, Toyota's automatic parking feature, Volvo's blind spot information system, Mercedes' pedestrian detector, and Tesla's AutoPilot feature.

Prob. 1.11 — Give an example of an application that you have used personally in the past month that involves image processing and/or analysis.

Answer (Prob. 1.11) — Answers may vary.

Prob. 1.12 — List three psychologists whose work has been influential in understanding the human visual system.

Answer (Prob. 1.12) — Hermann von Helmholtz conducted some of the first psychophysical experiments, Max Wertheimer proposed gestalt psychology, and J. J. Gibson contributed to our understanding of visual perception. Other answers are possible.

Prob. 1.13 — Give a real-world example of technology using each of the following fields: (a) photogrammetry, (b) signal processing, (c) computer graphics, and (d) machine learning.

Answer (Prob. 1.13) — Answers may vary, but (a) online mapping software uses techniques from photogrammetry, (b) digital music players employ decompression techniques from signal processing, (c) all modern movies with special effects use algorithms from computer graphics, and (d) automated speech recognition that is prevalent in smartphones is based upon decades of research in machine learning.

† § **??**

Prob. 1.14 — Search the Web for job openings in computer vision. List three jobs that you found, along with the qualifications needed to apply.

Answer (Prob. 1.14) — Answers may vary.

Prob. 1.15 — Suppose we have the following image: $I = \begin{bmatrix} 4 & 5 & 2 \\ 1 & 3 & 8 \end{bmatrix}$. Using the conventions of this book, what are the values of $I(0,1)$, $I(1,1)$, and $I(2,1)$?

Answer (Prob. 1.15) — $I(0,1) = 1$, $I(1,1) = 3$, and $I(2,1) = 8$.

Prob. 1.16 — Suppose an image has 640 columns and 480 rows and is stored in row-major order. Convert the coordinates $(x, y) = (38, 52)$, $(592, 241)$, and $(33, 0)$ to 1D indices. Conversely, convert the following 1D indices to (x, y) coordinates: $i = 8092$, 24061, and 38190.

Answer (Prob. 1.16) — Conversion to 1D indices is achieved by applying Equation (1.3):

$$\begin{aligned} (38, 52) &\rightarrow i = 52 \cdot 640 + 38 = 33318 \\ (592, 241) &\rightarrow i = 241 \cdot 640 + 592 = 154832 \\ (33, 0) &\rightarrow i = 0 \cdot 640 + 33 = 33 \end{aligned}$$

Conversion to 2D coordinates is achieved via Equations (1.4) and (1.5):

$$\begin{aligned} i = 8092 &\rightarrow (\mathrm{mod}(8092, 640), \lfloor 8092/640 \rfloor) = (412, 12) \\ i = 24061 &\rightarrow (\mathrm{mod}(24061, 640), \lfloor 24061/640 \rfloor) = (259, 38) \\ i = 38190 &\rightarrow (\mathrm{mod}(38190, 640), \lfloor 38190/640 \rfloor) = (430, 59). \end{aligned}$$

For example, $\lfloor 8092/640 \rfloor = 12$, so $\mathrm{mod}(8092, 640) = 8092 - 12 \cdot 640 = 412$.

Prob. 1.17 — Equations (1.3)–(1.5) apply to an image stored in row-major order. Write the equivalent expressions to convert between 2D coordinates and 1D indices for an image stored in column-major order.

Answer (Prob. 1.17) — The equivalent expressions are obtained by interchanging x and y and replacing *width* with *height*:

$$\begin{aligned} i &= x \cdot \mathit{height} + y \\ y &= \mathrm{mod}(i, \mathit{height}) = i - x \cdot \mathit{height} \\ x &= \lfloor i \,/\, \mathit{height} \rfloor. \end{aligned}$$

Prob. 1.18 — Suppose the following 1D array of bytes in memory stores a 2×2 color image (in blue-green-red order): 52, 68, 31, 133, 192, 88, 255, 208, 32, 233, 161, 25.

(a) Assuming that the image is stored in interleaved format, convert to planar format. What are the RGB values of the pixel at location $(1, 1)$?

(b) Assuming that the image is stored in planar format, convert to interleaved format. What are the RGB values of the pixel at location $(0, 1)$?

Answer (Prob. 1.18) — The answer is as follows:

(a) In interleaved format, the image is

$$\begin{bmatrix} (52, 68, 31) & (133, 192, 88) \\ (255, 208, 32) & (233, 161, 25) \end{bmatrix}.$$

Therefore, the RGB values at $(x, y) = (1, 1)$ are $(233, 161, 25)$, that is, blue = 233, green = 161, and red = 25.

(b) In planar format, the image is

$$\text{blue: } \begin{bmatrix} 52 & 133 \\ 255 & 233 \end{bmatrix} \qquad \text{green: } \begin{bmatrix} 68 & 192 \\ 208 & 161 \end{bmatrix} \qquad \text{red: } \begin{bmatrix} 31 & 88 \\ 32 & 25 \end{bmatrix}.$$

Therefore, the RGB values at $(x, y) = (0, 1)$ are $(255, 208, 32)$, that is, blue = 255, green = 208, and red = 32.

Prob. 1.19 — Suppose the following 1D array of bytes in memory stores 8 consecutive pixels of a binary image: $0, 0, 0, 255, 255, 0, 255, 0$. Show how to store these pixels in a single packed byte.

Answer (Prob. 1.19) — Each pixel is either ON or OFF, so we can convert the sequence of values into 0s and 1s: 00011010, or `1A` in hexadecimal.

Prob. 1.20 — For increased fidelity, medical images are often stored using more than 8 bits. Suppose you needed to store a 12-bit-per-pixel grayscale image. Would you try to pack 3 pixels into 2 bytes to avoid wasted bits? Why or why not?

Answer (Prob. 1.20) — Packing 3 pixels into 2 bytes would remove the wasted bits, thus saving memory — at the expense of making it more difficult to access a single pixel. Depending on the application, it might make sense to pack the pixels when storing to disk, but unpack them when reading into memory.

Prob. 1.21 — Convert the following grayscale image to set notation: $\begin{bmatrix} 112 & 195 & 48 \\ 97 & 203 & 125 \end{bmatrix}$.

Answer (Prob. 1.21) — Set notation captures the x and y coordinates, along with the values:

$$\{(0, 0, 112), (1, 0, 195), (2, 0, 48), (0, 1, 97), (1, 1, 203), (2, 1, 125)\}.$$

Prob. 1.22 — Convert the following 3×2 binary image back to array notation: $\{(1, 0), (0, 1), (2, 1)\}$.

Answer (Prob. 1.22) — In array notation, the image is represented as

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Prob. 1.23 — Consider the following 2D array, which has 3 columns and 2 rows: $\begin{bmatrix} 167 & 30 & 245 \\ 41 & 127 & 87 \end{bmatrix}$.

(a) If the array is an image I, what is the value of $I(1, 1)$?

(b) If it is a matrix $\mathbf{A}$, what is the value of a_{11}?

(c) In which case would you write the dimensions as 2×3? As 3×2?

Answer (Prob. 1.23) — The answers are as follows:

(a) Using the conventions of this book, images are accessed by column, then row, both zero-based, so that $I(1, 1) = 127$.

(b) Matrices are accessed by row, then column, both one-based, so that $a_{11} = 167$.

(c) As a matrix, the dimensions of the array are 2×3; as an image, its dimensions are 3×2.

Prob. 1.24 — List three mathematical prerequisites for studying the material in this book.

Answer (Prob. 1.24) — The material in this book draws from topics in probability, linear algebra, calculus, and signal processing.

Prob. 1.25 — Explain the difference between a convergent problem and a divergent problem.

Answer (Prob. 1.25) — A convergent problem has a single solution, and the more you study the problem the more you know about it. In contrast, a divergent problem has many solutions, making it appear that the more you study the problem the less you know about it.

Prob. 1.26 — Briefly explain why computer vision is so difficult.

Answer (Prob. 1.26) — Robust computer vision algorithms are difficult due to the many variables that affect the capturing of an image, such as the pose of the object, shadows, and lighting conditions.

Prob. 1.27 — Download a software library (e.g., OpenCV), and write a program to load an image from a file and display it in a window.

Answer (Prob. 1.27) — Answers may vary.