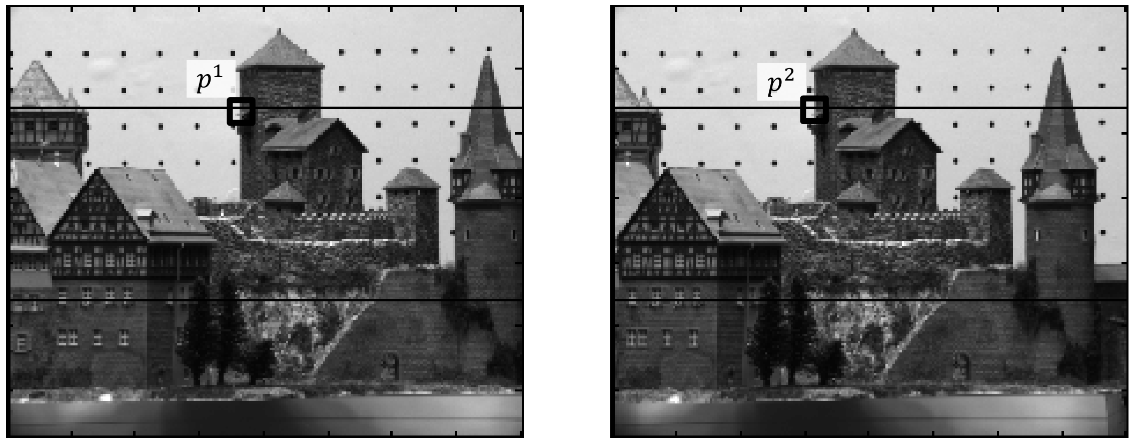


(a) left image (b) right image

1. [20] The above images show the left and the right images of a (rectified) stereo camera. Using epipolar lines (the black horizontal lines), $\boldsymbol{p^1}$ in the left is found to correspond to $\boldsymbol{p^2}$ in the right. Both cameras have the same image sensor with the width 1600 pixels, the height 1200 pixels, and the focal length of 1 cm. The pixel resolutions for both horizontal and vertical directions are 2400 pixels/cm. The baseline is 20 cm. Assume that $\boldsymbol{p^1}$ and $\boldsymbol{p^2}$ are mapped to each camera with their pixel coordinates $(\boldsymbol{u^\ell}, \boldsymbol{v^\ell}) = (\mathbf{786}, \mathbf{304})$ and $(\boldsymbol{u^r}, \boldsymbol{v^r}) = (\mathbf{780}, \mathbf{304})$, respectively. Compute the coordinates of $\boldsymbol{p^2}$ (or $\boldsymbol{p^1}$) $= (\boldsymbol{X^w}, \boldsymbol{Y^w}, \boldsymbol{Z^w})$ in the world coordinate frame whose origin is located at the left camera frame.

$$\lambda_1 \bar{p}^1 = K[I \mid 0]\bar{p}^w = Kp^w$$

focal point $= 1\,\text{cm} \cdot 2400$ pixel/cm Flength

$f = 2400$ pixel

$$K = \begin{bmatrix} 2400 & 0 & 800 \\ 0 & 2400 & 600 \\ 0 & 0 & 1 \end{bmatrix} \qquad \bar{p}^1 = \begin{bmatrix} 786 \\ 304 \\ 1 \end{bmatrix}$$ left image

$B = 20$ cm

$C_x = 1600/2 = 800$

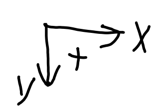
$C_y = 1200/2 = 600$

$$K^{-1} = \begin{bmatrix} 1/2400 & 0 & -\frac{1}{3} \\ 0 & 1/2400 & -1/4 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\text{Depth}: Z^w = \frac{f \cdot B}{u_l - u_r} = \frac{2400 \cdot 20}{786 - 780} = 8000\,\text{cm} = 80\,\text{m}$$

$$p^w = \lambda_1 K^{-1}\bar{p}^1 = 80\text{m}\begin{bmatrix} 1/2400 & 0 & -\frac{1}{3} \\ 0 & 1/2400 & -1/4 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 786 \\ 304 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.467\text{m} \\ -9.867\text{m} \\ 80\text{ m} \end{bmatrix}$$

$$p^1 = p^w = (\underbrace{-0.467}_{X^w}, \underbrace{-9.867}_{Y^w}, \underbrace{80}_{Z^w})\ [m]$$



into page is +z

2. [20] Compute the following four different metrics for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ with respect to $\boldsymbol{w}$, (i.e. compute the metrics for two pairs; $(\boldsymbol{a}, \boldsymbol{w})$ and $(\boldsymbol{b}, \boldsymbol{w})$).

$$a = \begin{bmatrix} 1 & 5 & 3 \\ 2 & 3 & 8 \\ 4 & 6 & 2 \end{bmatrix}, b = \begin{bmatrix} 1 & 9 & 4 \\ 2 & 4 & 3 \\ 5 & 0 & 6 \end{bmatrix}, w = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \\ 1 & 4 & 3 \end{bmatrix}$$

(a) Cross correlation
(b) Convolution
(c) Sum of absolute differences (SAD)
(d) Sum of squared differences (SSD)

no padding

Full 5x5 for cross correlate and convolute calculated using matlab

$$a \text{ xcorr } w = \begin{bmatrix} 3 & 19 & 30 & 17 & 3 \\ 8 & 28 & 52 & 53 & 17 \\ 17 & 50 & 79 & 50 & 32 \\ 10 & 25 & 43 & 50 & 22 \\ 4 & 18 & 28 & 18 & 4 \end{bmatrix}$$

$$b \text{ xcorr } w = \begin{bmatrix} 3 & 31 & 49 & 25 & 4 \\ 8 & 39 & 47 & 47 & 15 \\ 20 & 42 & 72 & 69 & 23 \\ 12 & 15 & 46 & 23 & 24 \\ 5 & 15 & 16 & 18 & 12 \end{bmatrix}$$

$$a \text{ conv } w = \begin{bmatrix} 2 & 13 & 22 & 14 & 3 \\ 7 & 28 & 43 & 40 & 14 \\ 15 & 44 & 83 & 53 & 27 \\ 14 & 33 & 46 & 55 & 28 \\ 4 & 22 & 38 & 26 & 6 \end{bmatrix}$$

$$b \text{ conv } w = \begin{bmatrix} 2 & 21 & 36 & 21 & 4 \\ 7 & 42 & 43 & 35 & 11 \\ 17 & 42 & 77 & 72 & 24 \\ 17 & 17 & 53 & 30 & 21 \\ 5 & 20 & 21 & 24 & 18 \end{bmatrix}$$

Cross - conv

$a \text{ xcorr } w = I'(x,y) = \sum_{s=-1}^{1} \sum_{t=-1}^{1} a(s,t) \cdot w(x+s, y+t)$ rows $= 2(3)-1$, coloums $= 2(3)-1$ $\rightarrow 5\times5$

$I'(2,2) = 1(2) + 5(3) + 3(1) + 2(3) + 3(1) + 8(2) + 4(1) + 6(4) + 2(3)$
$= 79$

$b \text{ xcorr } w = I'(x,y) = \sum_{s=-1}^{1} \sum_{t=-1}^{1} a(s,t) \cdot w(x+s, y+t)$ rows $= 2(3)-1$, coloums $= 2(3)-1$ $\rightarrow 5\times5$

$I'(2,2) = 1(2) + 9(3) + 4(1) + 2(3) + 1(4) + 2(3) + 5(1) + 0(4) + 3(6)$
$= 72$

conv

$A \text{ conv } W \rightarrow 5\times5$ matrix $I'(x,y) = \sum_{s=-1}^{1} \sum_{t=-1}^{1} a(s,t) \cdot w(x-s, y-t)$

$I'(2,2) = 1(3) + 5(4) + 3(1) + 2(2) + 3(1) + 8(3) + 4(1) + 6(3) + 2(2)$
$= 83$

$B \text{ conv } W \rightarrow 5\times5$ $I'(x,y) = \sum_{s=-1}^{1} \sum_{t=-1}^{1} a(s,t) \cdot w(x-s, y-t)$

$I'(2,2) = 1(3) + 9(4) + 4(1) + 2(2) + 4(1) + 3(3) + 5(1) + 0(3) + 6(2)$

$= 77$

SAD

$a \text{ sad } w = \sum_{k=0}^{3} \sum_{l=0}^{3} \left| a(u+k, v+l) - w(u'+k, v'+l) \right|$

$= |1-2| + |5-3| + |3-1| + |2-3| + |3-1| + |8-2| + |4-1| + |6-4| + |2-3|$
$= 26$

$b \text{ sad } w = \sum_{k=0}^{3} \sum_{l=0}^{3} \left| b(u+k, v+l) - w(u'+k, v'+l) \right|$

$= |1-2| + |9-3| + |4-1| + |2-3| + |4-1| + |3-2| + |5-1| + |0-4| + |6-3|$
$= 26$

SSD

$a \text{ sad } w = \sum_{k=0}^{3} \sum_{l=0}^{3} \left(a(u+k, v+l) - w(u'+k, v'+l) \right)^2$

$= (1-2)^2 + (5-3)^2 + (3-1)^2 + (2-3)^2 + (3-1)^2 + (8-2)^2 + (4-1)^2 + (6-4)^2 + (2-3)^2$
$= 64$

$b \text{ sad } w = \sum_{k=0}^{3} \sum_{l=0}^{3} \left(b(u+k, v+l) - w(u'+k, v'+l) \right)^2$

$= (1-2)^2 + (9-3)^2 + (4-1)^2 + (2-3)^2 + (4-1)^2 + (3-2)^2 + (5-1)^2 + (0-4)^2 + (6-3)^2$
$= 98$

3. [20] Apply Sobel filter to the following image patch D and obtain $I_x(u,v) \in \mathbb{R}^{2\times 2}$ and $I_y(u,v) \in \mathbb{R}^{2\times 2}$ matrices for the 2×2 center pixels, i.e. $u \in \{2,3\}$ and $v \in \{2,3\}$.

$$D = \begin{bmatrix} 7 & 6 & 5 & 5 \\ 6 & 5 & 4 & 4 \\ 5 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Using obtained $I_x(u,v)$ and $I_y(u,v)$, compute the 2nd moment matrix for the Harris corner detector;

$$M = \sum_{u=2}^{3}\sum_{v=2}^{3}\begin{bmatrix} I_x^2(u,v) & I_x(u,v)I_y(u,v) \\ I_x(u,v)I_y(u,v) & I_y^2(u,v) \end{bmatrix}$$

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$I_x = G_x * D \rightarrow 6\times 6$$

$$= \begin{bmatrix} X & X & X & X & X & X \\ X & X & X & X & X & X \\ X & X & 7(1)-5(1)+6(2)-4(2)+5(1) & 6-5+5(2)-4(2)+1 & X & X \\ X & X & 6(1)-4(1)+5(2)+1 & 5-4+2 & X & X \\ X & X & X & X & X & X \\ X & X & X & X & X & X \end{bmatrix}$$

$$\rightarrow = \begin{bmatrix} 11 & 4 \\ 13 & 3 \end{bmatrix}$$

$$I_y = G_y * D$$

$$= \begin{bmatrix} X & X & X & X & X & X \\ X & X & X & X & X & X \\ X & X & 7+6(2)+5-5-2 & 6+5(2)+5-1 & X & X \\ X & X & 6+5(2)+4-1 & 5+4(2)+4 & X & X \\ X & X & X & X & X & X \\ X & X & X & X & X & X \end{bmatrix}$$

$$\rightarrow = \begin{bmatrix} 17 & 20 \\ 19 & 17 \end{bmatrix}$$

$$M = \begin{bmatrix} 11^2 + 4^2 + 13^2 + 3^2 & 11(17) + 4(20) + 13(19) + 3(17) \\ 11(17) + 4(20) + 13(19) + 3(17) & 17^2 + 20^2 + 19^2 + 17^2 \end{bmatrix}$$

$$= \begin{bmatrix} 315 & 565 \\ 565 & 1339 \end{bmatrix}$$

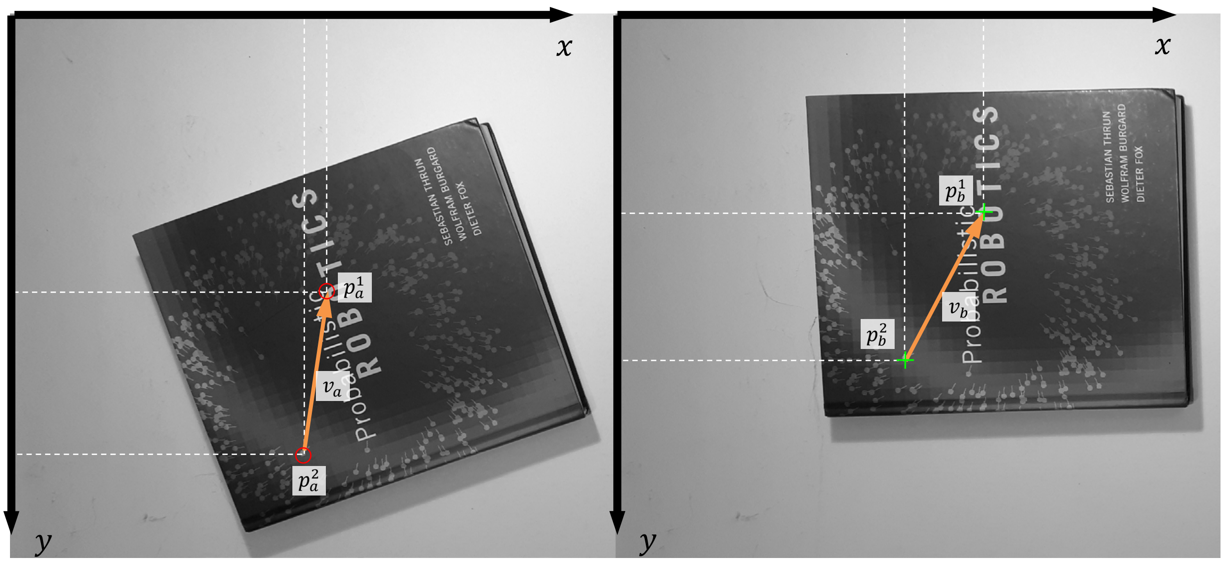


Figure 1 Two images of a book

4. [40] This problem is about SfM (Structure from Motion) with two camera views. In this problem, we like to implement RANSAC to match features so that we can find the SE(2) between two images. Figure 1 shows example photos I took with my old cell phone (Galaxy A5-2017). The two image files ("Book_a.jpg" and "Book_b.jpg") as well as the skeleton Matlab code are provided. There are 6 sections in the Matlab code. Sections 1,2,3,4 and 6 have already been completed. Sections 1,2 and 3 provide the initial feature matching with 328 matched feature points (denoted by "pt_a" and "pt_b" in Matlab code). However, they still have many incorrect matches as shown in Figure 2. Each of the original feature sets "pt_a" and "pt_b" is a 2×328 matrices; each column of these matrices represents (x, y) coordinates (in pixels) of the same feature. You are required to complete Section 5 of the Matlab code to find all the inliers among 328 features by estimating the most credible SE(2) between two images through RANSAC. You only need to complete small part within Section 5 (by replacing "…" by your own code), which include;

   i) The SE(2) can be found by two pairs of matched feature points. As shown in Figure 1, two vectors v_a and v_b can be constructed by two randomly selected feature points. The features p_a^1 and p_b^1 are randomly selected (same) columns from "pt_a" and "pt_b", respectively. They are supposed to represent the same point (if they are matched correctly). Likewise, the features p_a^2 and p_b^2 are another randomly selected (same) columns from "pt_a" and "pt_b", respectively. Using v_a and v_b, we can find the rotation matrix $R \in \mathbb{R}^{2\times 2}$ (through the angle θ) and the translation vector $t \in \mathbb{R}^2$. (**<u>Note</u>**: The image coordinates use the left-handed coordinate frame as shown in Figure 1, so the direction of rotation is opposite to the case with the usual right-handed coordinate frame.)

   ii) The computed R and t can be applied to transform all the feature points (i.e. all the columns) of "pt_b". If the sampled feature pairs (p_a^1, p_b^1) and (p_a^2, p_b^2) were the correct matches, many of other feature points will also be matched. These matched features are "inliers". The more the number of inliers, the more likely that R and t are correct. The inliers can be found by comparing the threshold error ("thresh_e" in the Matlab code) to the norm of the error vector $e_i = p_a^i - p_b^{i\,\prime}$ where $p_b^{i\,\prime}$ is the transform of p_b^i (i.e. the ith feature point of the "Book_b.jpg" or the ith column of "pt_b").

iii) The RANSAC repeats the above process for the specified number of iterations (defined as "N_iter" in the Matlab code) while updating the number of inliers. After the iteration completes, the $[R|t]$ that gives the largest number of inliers will be the correct SE(2).

If you have completed the code correctly, running the plot command in Section 6 of the code will give you all the correct inliers as shown in Figure 3. In your answer, provide the computed $[R|t]$ and the number of inliers (i.e. the value of "Max_count").

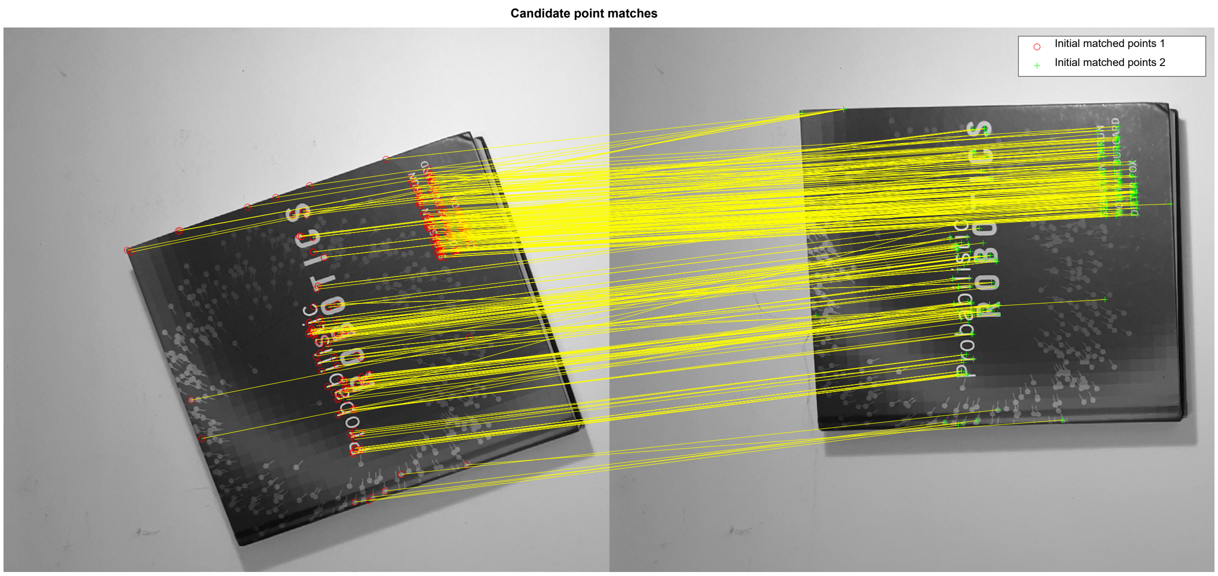


Figure 2 Initial feature matching with many outliers

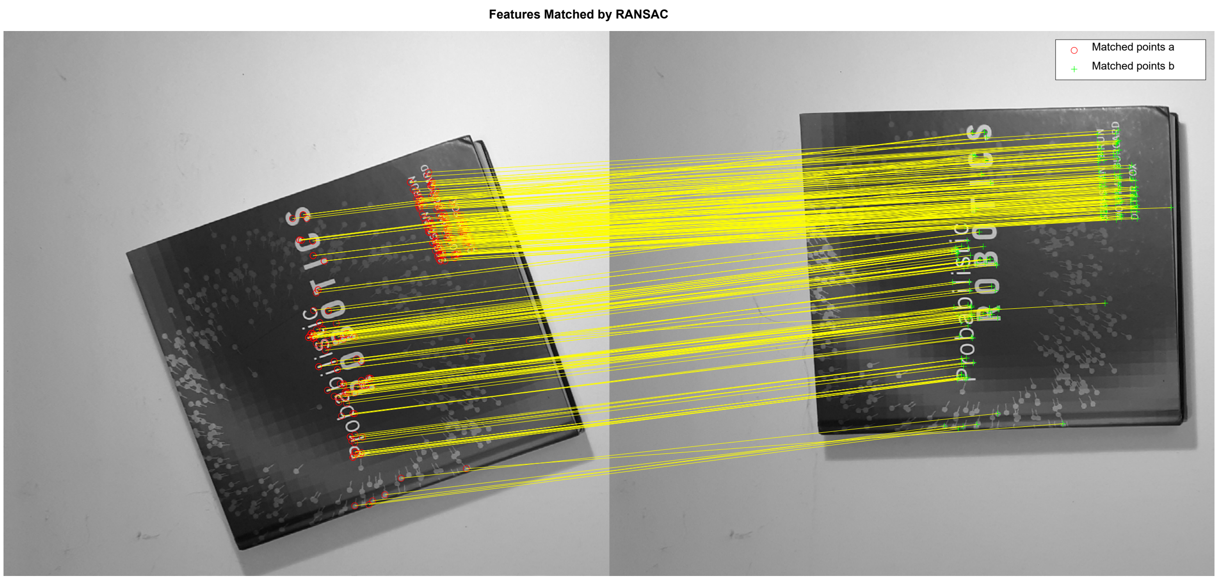
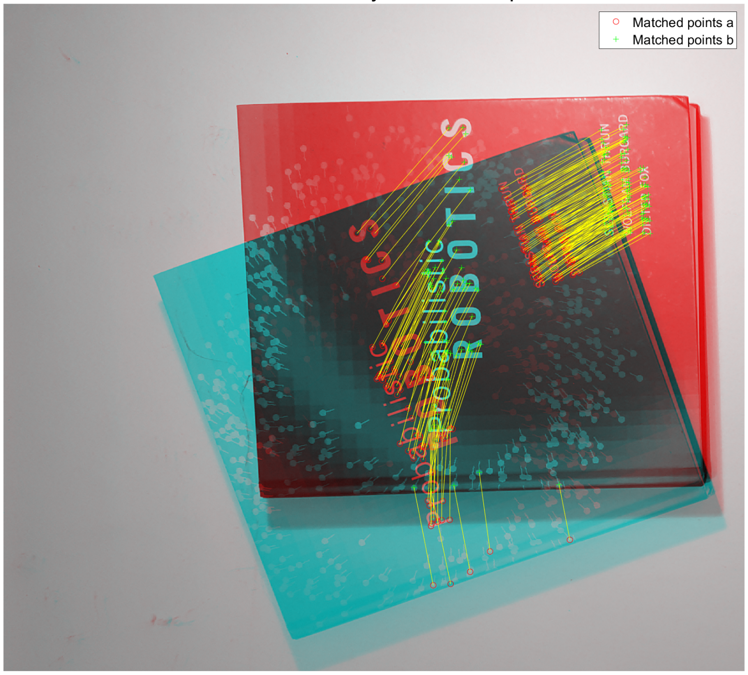


Figure 3 All the inliers found by RANSAC

Features Matched by RANSAC - Overlap



Max count = 281

$$[R|t] = \left[\begin{array}{cc|c} 0.9519 & 0.3063 & -924.9349 \\ -0.3063 & 0.9519 & 2.0894 \end{array}\right]$$

Features Matched by RANSAC - Montage