

Introduction to Computer Vision Assignment 4

Lam Nguyen - 500838417
Toronto Metropolitan University
lam.nguyen@ryerson.ca

I. INTRODUCTION

This assignment introduces the basic concepts of imaging process and single view geometry. In addition, implementations of the SIFT keypoint detector will be investigated.

II. PART 1

A. Problem 1

1. The canonical form of the plane at infinity is given as following:

$$\pi_\infty = (0, 0, 0, 1)^T$$

2. Recall the affine matrix

$$H_A = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$$

and a 3D plane is transformed as

$$\pi' = H_A^{-T} \pi$$

Apply affine transformation on a plane at infinity $\pi_\infty = (0, 0, 0, 1)^T$

$$\pi_\infty' = H_A^{-T} \pi_\infty = \begin{bmatrix} A^{-T} & 0 \\ -t^T A^{-T} & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

As the result, the plane at infinity is fixed under 3D affine transformation.

3. Given the 3D point transformation $X' = HX$, then the inverted point transformation is

$$X = H^{-1} X'$$

The 3D plane transformation obtained by applying inverted point transformation is

$$\pi^T X = \pi^T H^{-1} X' = (H^{-T} \pi)^T X'$$

Therefore, the result of the 3D plane transformation is

$$\pi' = H^{-T} \pi$$

B. Problem 2

1. The general form of a finite projective camera's intrinsic parameter matrix K is defined as:

$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where

- α_x and α_y : the focal lengths of the camera
- s : the skew parameter, which is non-zero if the image axes are not perpendicular
- x_0 and y_0 : the coordinates of the principal point (optical center)

2. Given the projection matrix $P = [M \mid p_4]$, we can rearrange the matrix as follow

$$P = M[I \mid M^{-1} p_4]$$

Substitute the center $C = -M^{-1} p_4$ and $KR = M$

$$P = KR[I \mid -\tilde{C}]$$

$$P = K[R \mid -R\tilde{C}]$$

$$P = K[R \mid t]$$

where

- K: intrinsic matrix
- [R | t]: extrinsic matrix

Therefore, the internal and external parameters are recovered from the projection matrix P.

3. Points which can be written as $X_\infty = (d^T, 0)^T$ are mapped to the image plane by a projection matrix $P = [M \mid p_4]$

$$x = PX_\infty = [M \mid p_4]X_\infty = Md$$

For the point at infinity, the mapping is only affected by sub-matrix M.

C. Problem 3

1. Given the projection matrix P , the optical center P of the camera is the right null space of the matrix P

$$PC = 0$$

Direct computation for optical center

$$X = \det([p_2, p_3, p_4])$$

$$Y = -\det([p_1, p_3, p_4])$$

$$Z = \det([p_1, p_2, p_4])$$

$$T = -\det([p_1, p_2, p_3])$$

2. Recall the projection matrix P

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

consist of 4 columns

$$p_1 = \begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \end{bmatrix}, p_2 = \begin{bmatrix} p_{12} \\ p_{22} \\ p_{32} \end{bmatrix}, p_3 = \begin{bmatrix} p_{13} \\ p_{23} \\ p_{33} \end{bmatrix}, p_4 = \begin{bmatrix} p_{14} \\ p_{24} \\ p_{34} \end{bmatrix}$$

and 3 rows

$$p^{1T} = [p_{11} \ p_{12} \ p_{13} \ p_{14}],$$

$$p^{2T} = [p_{21} \ p_{22} \ p_{23} \ p_{24}],$$

$$p^{3T} = [p_{31} \ p_{32} \ p_{33} \ p_{34}]$$

Suppose having three homogeneous points $(1, 0, 0, 0)^T$, $(0, 1, 0, 0)^T$, and $(0, 0, 1, 0)^T$ represented X , Y , and Z axes of the world system, respectively, then the vanishing points in x , y , and z directions are

$$P(1, 0, 0, 0)^T = p_1,$$

$$P(0, 1, 0, 0)^T = p_2,$$

$$P(0, 0, 1, 0)^T = p_3$$

Therefore, the first three columns of the projection matrix correspond to the vanishing points of the X , Y , and Z axes of the world system.

3. The principle plane is plane through the camera center parallel to the image plane. It consists of the set of points X which are imaged on the line at infinity of the image

$$\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = PX$$

$$P^{3T}X = 0$$

As the result, the last row of the projection matrix corresponds to the principle plane of the camera.

D. Problem 4

1. Let a coplanar 3D point $X = (X, Y, 0, 1)^T$, then the mapping on the image plane is

$$x = PX$$

$$x = K \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$$

$$x = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$x = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Therefore, coplanar 3D points and their images are related by a 2D homography given by $H = K[r_1, r_2, t]$.

2. The back-projection of a line l is the set of points X in the scene such that $l^T(PX) = 0$. Let $\pi = P^Tl$

$$\pi^T X = (P^T l)^T X = l^T P X = 0$$

The result signifies that the back-projection of an image line is a 3D plane given by $\pi = P^T l$.

3. Let x and x' be the images of a point X before and after zooming, respectively

$$x = K[I \mid 0]X$$

$$x' = K'[I \mid 0]X$$

$$x' = K'K^{-1}(K[I \mid 0]X)$$

$$x' = K'K^{-1}x = Hx$$

Therefore, the image captured by a zooming camera are related by a 2D homography given by $H = K'K^{-1}$.

E. Problem 5

1. Recall that the absolute conic is on the plane at infinity π_∞ and a point on π_∞ is mapped to an image point as

$$\mathbf{x} = \mathbf{P}\mathbf{X}_\infty = \mathbf{K}\mathbf{R}[\mathbf{I} | -\tilde{\mathbf{C}}] \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix} = \mathbf{K}\mathbf{R}\mathbf{d}$$

The mapping between π_∞ and an image is given by a planar homography $\mathbf{x} = \mathbf{H}\mathbf{d}$ with

$$\mathbf{H} = \mathbf{K}\mathbf{R}$$

The image of the absolute conic (the IAC) is the conic

$$\mathbf{C} = \Omega_\infty = \mathbf{I}$$

$$\mathbf{C} \mapsto \mathbf{H}^{-T}\mathbf{C}\mathbf{H}^{-1}$$

$$\omega = (\mathbf{K}\mathbf{R})^{-T}\mathbf{I}(\mathbf{K}\mathbf{R})^{-1} = \mathbf{K}^{-T}\mathbf{R}\mathbf{R}^T\mathbf{K}^{-1} = (\mathbf{K}\mathbf{K}^T)^{-1}$$

Therefore, the image of the absolute conic is given by $\omega = (\mathbf{K}\mathbf{K}^T)^{-1}$.

2. The angle between two rays is given by

$$\cos\theta = \frac{\mathbf{x}_1^T \omega \mathbf{x}_2}{\sqrt{\mathbf{x}_1^T \omega \mathbf{x}_1} \sqrt{\mathbf{x}_2^T \omega \mathbf{x}_2}}$$

Two image points correspond to orthogonal directions if $\cos\theta = 0$ or $\mathbf{x}_1^T \omega \mathbf{x}_2 = 0$.

3. In the case of planar projective geometry we can define two characteristic points, so called circular points (or, absolute points). The circular points \mathbf{I} and \mathbf{J} are a pair of complex conjugate ideal points

$$\mathbf{I} = \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix}, \mathbf{J} = \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix}$$

In 3D projective space every plane π intersects the plane at infinity π_∞ at the line at infinity l_∞ of that particular plane π . Also, this line l_∞ intersects the absolute conic in two circular points of plane π . Furthermore the images of those circular points lie on the IAC. Thus, if we can find the homography $\mathbf{H} = [h_1, h_2, h_3]$ between some plane π and the camera image plane we can easily apply the homography \mathbf{H} on the circular points to find images of circular points $h_1 \pm ih_2$. Knowing the images of circular points we have the linear constraint on the elements of the IAC

$$(h_1 \pm ih_2)^T \omega (h_1 \pm ih_2) = 0$$

The imaginary and real parts give respectively

$$h_1^T \omega h_2 = 0$$

and

$$h_1^T \omega h_1 = h_2^T \omega h_2$$

4. assume that the camera has zero skew and the pixels are square. This removes some of the constraints on the image of the absolute conic ω , in that zero skew has $\omega_{12} = \omega_{21} = 0$,

and square pixels has $\omega_{11} = \omega_{22}$. As such, the image of the absolute conic has the form

$$\omega = \begin{bmatrix} a & 0 & b \\ 0 & a & c \\ b & c & 1 \end{bmatrix}$$

The final value of ω set to 1 due to scale invariability. In addition, if we know that two lines l_1 and l_2 are orthogonal, then their respective vanishing points also satisfies $v_1^T \omega v_2 = 0$. When this is expanded, we get

$$\begin{aligned} v_1^T \omega v_2 &= [x_1 \ x_2 \ x_3] \begin{bmatrix} a & 0 & b \\ 0 & a & c \\ b & c & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \\ &= ax_1x_2 + bx_1z_2 + ay_1y_2 + cy_1z_2 + bz_1x_2 + cz_1y_2 + z_1z_2 \\ &= [x_1x_2 + y_1y_2, x_1z_2 + z_1x_2, y_1z_2 + z_1y_2, z_1z_2] \begin{bmatrix} a \\ b \\ c \\ 1 \end{bmatrix} \\ &= \mathbf{A}\mathbf{w} = 0 \end{aligned}$$

Thus, for a pair of vanishing points from orthogonal lines, we are able to obtain one constraint on ω , and thus for the three vanishing points all of which are from orthogonal parallel line, we are able to obtain 3 total constraints on ω . Because zero skew and square pixels resolves another two constraints, this yields the 5 needed constraints to solve ω . Thus, we are then able to solve the above $\mathbf{A}\mathbf{w} = 0$ through SVD and reshaping to yield the 3×3 matrix of ω and dividing by the last entry in ω for the scale invariance (setting it to 1).

The image of the absolute conic ω can be equivalently related to the intrinsic matrix by the following equation $\omega = (\mathbf{K}\mathbf{K}^T)^{-1}$. Note that the intrinsic matrix is an upper triangular matrix, the inverse of an upper triangular matrix is an upper triangular matrix, and the transpose of an upper triangular matrix is a lower triangular matrix. The Cholesky decomposition is a method of decomposing an $n \times n$ matrix into $L L^T$, where L is a lower triangular matrix. Thus, by applying the Cholesky decomposition onto ω , we are able to obtain the lower triangular matrix \mathbf{K}^{-T} which, by transposing and then inverting the matrix, yields the intrinsic matrix \mathbf{K} .

III. PART 2

SIFT feature matching and image stitching

A fully automated method of stitching panoramic images using invariant features has a number of benefits over earlier ones. First of all, despite rotation, zooming, and illumination changes in the input photos, our use of invariant features permits accurate matching of panoramic image sequences. By considering image stitching as a multi-image matching issue, we can automatically identify panoramas in unordered datasets and find the relationships that the photos match. Thirdly, we render seamless output panoramas utilising multi-band blending to produce high-quality outputs.

The extraction and matching of SIFT characteristics from all of the photos is the initial stage in the panorama recognition method. The scale-space maxima and minima of a difference in a Gaussian function are where SIFT characteristics are found. A distinctive scale and orientation is established for each feature location. This provides a measurement frame that is similarity-invariant. The similarity invariant descriptor is actually calculated by adding up local gradients in orientation histograms, despite the fact that simply sampling intensity values in this frame would be similarity invariant. This gives some robustness to affine change by allowing edges to move a little bit without changing the descriptor vector.

The group of transformations that the images may go through is a particular group of homographies, assuming that the camera rotates about its optical centre. Each camera is parameterized with a rotation vector $\theta = [\theta_1, \theta_2, \theta_3]$ and focal length f . This gives pairwise homographies $\tilde{u}_i = H_{ij}\tilde{u}_j$ where

$$H_{ij} = K_i R_i R_j^T K_j^{-1}$$

The 4 parameter camera model is defined by

$$K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Use of picture features that are invariant to this set of transformations is ideal. Nevertheless, for minor adjustments to image position

$$u_i = u_{i0} + \left. \frac{\partial u_i}{\partial u_j} \right|_{u_{i0}} \Delta u_j$$

or equivalently $\tilde{u}_i = A_{ij}\tilde{u}_j$ where

$$A_{ij} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

is an affine transformation that was created by lining up the homography around u_{i0} . The use of SIFT features, which are largely invariant under affine change, is justified by the implication that each little image patch experiences an affine transformation.

Features must be matched after being extracted in linear time from all n photos. Each feature is matched to its k

nearest neighbours in feature space because many photos could overlap a single ray.

Later, connected collections of image matches would develop into panoramas. This problem looks to be quadratic in the number of photos at first glance due to the possibility that every image could possibly match every other image. To acquire a satisfactory solution for the picture geometry, it is only essential to match each image to a few pairs of adjacent images.

We have found photographs that have a lot of matches between them thanks to the feature matching stage. We look for probable picture matches among a fixed set of m photos that most closely match the present image in terms of features. First, we choose a set of inliers consistent with a homography between the pictures using RANSAC. After that, a probabilistic model is used to confirm the match.



Fig. 1: The result image showed that the algorithm recognised multiple panoramas in unordered image sets, and stitch them fully automatically without user input. The system is robust to orientation of the input images and ensures smooth transitions between images despite illumination differences, while preserving high frequency details.