IIST, Thiruvananthapuram AVD624 - Computer Vision

Lecture Notes: Week - 1

18/01/2022 - 28/01/2022

Lecturer: Dr. Deepak Mishra *Summarised by: Vatsalya Gupta*

This is a weekly summary of the lectures given by *Dr. Deepak Mishra*, Prof. and Head of Department of Avionics, for the course on *Computer Vision* conducted during the even semester 2021-22.

1 Introduction

The fundamental problem of computer vision is to convert light into meaning, i.e. use the observed image/video data (semantics, geometry etc.) to infer something about the world. Mapping the 3D world onto a 2D image plane leads to the loss of sense of perception, depth etc. Hence, a major challenge in computer vision is to recover unknowns from insufficient information.

Computer Vision Pipeline - A traditional computer vision system involves the following flow: input, sensing device (e.g. camera), interpreting device and the output. Interpretation is a major task in computer vision. In classical systems, it involved image preprocessing, feature extraction and classification, i.e. labelling of symbols into an output. Features are the unique characteristics which help us differentiate between different entities.

In a modern computer vision system, all of this is replaced with deep learning models using end-to-end enabled neural networks. Humans perceive the world using shape, colour, size etc. but for a computer an image is just a matrix of numbers which needs to be interpreted in terms of several complex variables, such as viewpoint differences, varying scales and instances, and occlusion.

2 Image Formation

An object and a film form the basic requirements for image formation. But a major caveat is that multiple light rays would reflect from a single point on the object towards the film, not leading to a reasonable image. The solution is to put a barrier with a small aperture (opening) between the object and the film. This helps in blocking most of the rays and hence reducing the blurring.

This concept forms the basis of a basic pinhole camera. The distance between the centre of the camera (barrier) and the image plane (film) is known as the focal length. Hence, this image formation can be better understood using projective geometry.

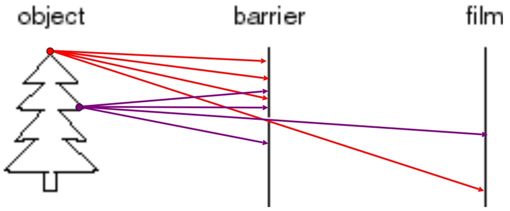

Figure 1: A basic Pinhole Camera setup

3 Projective Geometry

Projective geometry models the mapping of the 3D world onto a 2D image and also provides a mathematical representation appropriate for computations. But this leads to dimensionality reduction and change/loss of important information such as length, area, size, depth perception etc. Position of the image plane is an important factor in determining the length, angle etc. of the objects mapped onto the image.

An important feature is that straight lines are preserved when mapped onto the image plane, but their angles may differ from the real world. Parallel lines in the world intersect in the image at a *vanishing point* due to the perspective projection.

Vanishing Point - A point on the image plane of a perspective drawing where the 2D perspective projections of mutually parallel lines in 3D space appear to converge. Generally, an image may have three vanishing points.

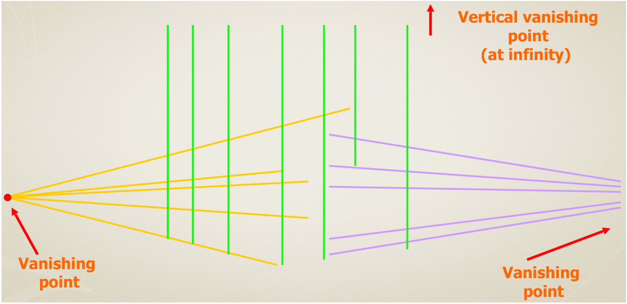

Figure 2: Depiction of Vanishing points

4 CCD and CMOS Camera

Pros and Cons of CCD (Charge Coupled Device) Camera

+ Larger fill factor(the ratio of a pixels light sensitive area to its total area)
− Harder to read out image
− More expensive
− Shutter which stops exposing while data read off

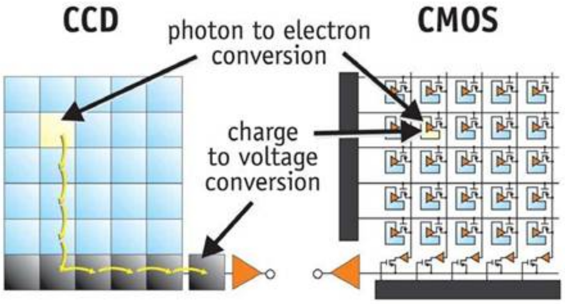

Figure 3: Functioning of CCD and CMOS

Pros and Cons of CMOS (Complementary Metal Oxide Semiconductor) Camera
+ Random access to pixels
+ Cheaper to manufacture
− Lower fill factor
+ Shutter can expose all the time (rolling shutter)
− Causes some new artifacts

5 Homogeneous Coordinate System

We introduce an augmented vector based upon its Euclidean representation as

- Converting to homogeneous coordinates (image and scene)

$$(x, y) \implies \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad\qquad (x, y, z) \implies \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\mathbb{R}^2 \to \mathbb{R}^3 \qquad\qquad \mathbb{R}^3 \to \mathbb{R}^4$$

- Converting from homogeneous coordinates (image and scene)

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \implies (x/w, y/w) \qquad\qquad \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \implies (x/w, y/w, z/w)$$

$$\mathbb{R}^3 \to \mathbb{R}^2 \qquad\qquad \mathbb{R}^4 \to \mathbb{R}^3$$

So we have mapped a point from 2D image plane to homogeneous coordinate system. Hence, we can define our projective space $\mathbb{P}^2$ as the set of equivalence classes of vectors in $\mathbb{R}^3 - (0, 0, 0)^T$.

Couple of properties are immediately evident that the homogeneous coordinate system is *invariant to scaling* and a *point* in Cartesian space is a *ray* in Homogeneous space.

6 Camera Projection Matrix

Now that we have defined the homogeneous coordinates, we can extend the idea to get the projection matrix for a camera setup.

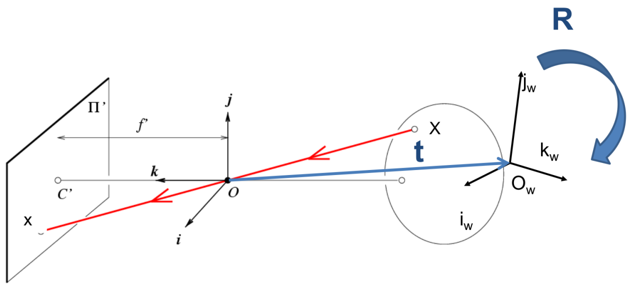

Figure 4: Alignment of Camera setup and the Image plane

$$x = K \begin{bmatrix} R & t \end{bmatrix} X$$

where, x: Image Coordinates (u, v, 1)
K: Intrinsic Matrix (3 x 3)
R: Rotation (3 x 3)
t: Translation (3 x 1)
X: World Coordinates (X, Y, Z, 1)

As a preliminary case, we first consider some assumptions

- Intrinsic assumptions: Unit aspect ratio, Optical centre at (0,0), No skew
- Extrinsic assumptions: No rotation, Camera at (0,0,0)

Hence, our expression becomes

$$x = K \begin{bmatrix} I & 0 \end{bmatrix} X \implies w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

where, f is the focal length.
We can generalise the expression by removing our assumptions one by one.

6.1 Remove assumption: Known optical centre

The optical centre is no longer at origin. Let it be translated to (u_0, v_0), we get

$$x = K\begin{bmatrix} I & 0 \end{bmatrix} X \implies w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

6.2 Remove assumption: Square pixels

Since, the pixels are no longer in square shape, hence focal length in x-direction (α) will be different from focal length in y-direction (β). So we get

$$x = K\begin{bmatrix} I & 0 \end{bmatrix} X \implies w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

6.3 Remove assumption: Non-skewed pixels

The skew parameter (s) defines the skewness in the intrinsic system. It defines the property of x and y not being linearly related, i.e. x may be varying at a different rate as compared to y. We get

$$x = K\begin{bmatrix} I & 0 \end{bmatrix} X \implies w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & s & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

6.4 Allow camera translation

Now we consider that the 3D world coordinates are not perfectly aligned to image plane and there can be translation along x, y, and z axis.

$$x = K\begin{bmatrix} I & t \end{bmatrix} X \implies w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

6.5 Allow camera rotation

We assume that we know the rotation along all 3 axes, defined by 3 matrices as

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\alpha & -sin\alpha \\ 0 & sin\alpha & cos\alpha \end{bmatrix}, \; R_y(\beta) = \begin{bmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{bmatrix}, \; R_z(\gamma) = \begin{bmatrix} cos\gamma & -sin\gamma & 0 \\ sin\gamma & cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Multiplying these rotations matrices in a certain order (since matrix multiplication is not commutative), we get simplified expression considering the camera rotation.

$$x = K\begin{bmatrix} I & t \end{bmatrix} X \implies w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

Hence, we have relaxed all the assumptions and obtained a generalised expression for mapping a 3D world point onto a 2D image plane using a projection matrix P, defined as the matrix multiplication of intrinsic (5 degrees of freedom) and extrinsic (6 dof) matrix, simplified as

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P_{3\times 4}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

P matrix has 12 elements which can be thought of as 11 independent ratios and hence 11 degrees of freedom. Upon relaxing the conditions such as square pixels and skewness, the degrees of freedom come down to 8.

7 Camera and Lens Properties

Field of View - It is the maximum area of a sample that a camera can image. The higher the focal length, the higher the image will be zoomed in and vice versa.

Radial Distortion - It refers to bending of straight lines into circular arcs, violating the main invariance preserved in the pinhole camera model, in which straight lines in the world map to straight lines in the image plane.

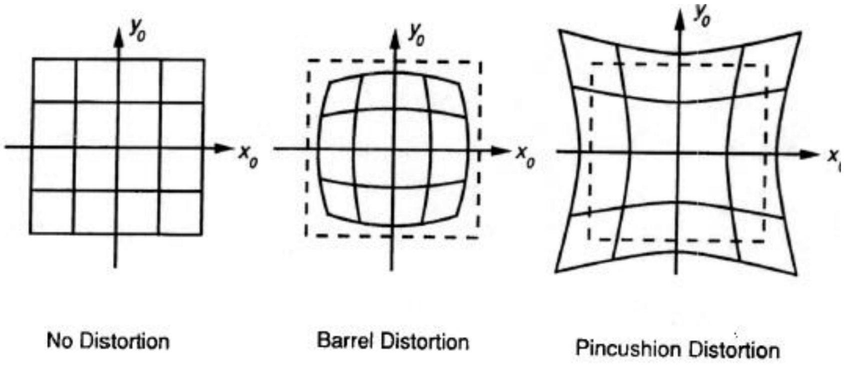

Figure 5: Lens distortion in Camera

8 Two-dimensional Points, Lines and Conics

8.1 Points

An *inhomogeneous* vector x is converted to a *homogeneous* vector $\tilde{x}$ as follows

$$\tilde{X} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ 1 \end{pmatrix} = \bar{X} \; (augmented\, vector)$$

$$\bar{X} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ 1 \end{pmatrix} = \frac{1}{\tilde{w}}\tilde{X} = \frac{1}{\tilde{w}} \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{pmatrix} = \begin{pmatrix} \tilde{x}/\tilde{w} \\ \tilde{y}/\tilde{w} \\ 1 \end{pmatrix}$$

Ideal Points or Points at Infinity - The homogeneous points which have last element as $\tilde{w} = 0$. These cannot be represented using inhomogeneous coordinates.

8.2 Lines

2D lines can also be expressed using homogeneous coordinates

$$\{\bar{X} \mid \tilde{l}^T \bar{X}\} \iff \{x, y \mid ax + by + c = 0\}$$

where, $\tilde{l} = [a, b, c]^T$ or $(n_x, n_y, d)^T = (n, d)$ in terms of normal vector and distance from origin.

Line at Infinity - $l_\infty = (0, 0, 1)$ passes through all the ideal points.

Hence, we get two important results, **intersection of two lines** ($\tilde{X} = \tilde{l}_1 \times \tilde{l}_2$) and **line joining two points** ($\tilde{l} = \tilde{X}_1 \times \tilde{X}_2$). Here, × denotes the cross product.

8.3 Conics

The general expression for a conic in affine space is given by

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

This can be homogenised by substituting $x = X_1/X_3$ and $y = X_2/X_3$ as

$$aX_1^2 + bX_1X_2 + cX_2^2 + dX_1X_3 + eX_2X_3 + fX_3^2 = 0 \implies X^TCX = 0$$

where, C is a symmetric homogeneous representation of the conic.

Complex algebraic objects can be represented using **polynomial homogeneous** $\bar{X} \mid \bar{X}^TQ\bar{X} = 0$.

Some important results which can be observed are - the line l tangent to a conic C at any point x is given by $l = Cx$; and the degenerate conic of rank 2 can be defined by two lines l and m as $C = lm^t + ml^t$, with the condition that $l^t x = 0$, satisfied by $(x^t l)(m^t x) + (x^t m)(l^t x) = 0$. Also, the dual conic $xy^t + yx^t$ represents lines passing through x and y.

Note - A similar idea can be extended to three-dimensional points, planes and quadrics using two-dimensional analogy and an increase in the number of parameters.

* A trickier representation of 3D lines (involving 6 parameters and 4 dof) can be done as a **linear combination** of two points *p* and *q* on the line

$$\{x \mid x = (1 - \lambda)p + \lambda q \wedge \lambda \in \mathbb{R}\}$$

9 Two-dimensional Transformations

We define *projectivity* as an invertible mapping *h* from $\mathbb{P}^2$ to itself such that three points x_1, x_2 and x_3 lie on the same line if and only if $h(x_1)$, $h(x_2)$ and $h(x_3)$ do. Hence, we can say that a mapping $h : \mathbb{P}^2 \rightarrow \mathbb{P}^2$ is a projectivity if and only if there exists a non-singular 3 x 3 matrix *H* such that for any point in $\mathbb{P}^2$ represented by a vector *x* it is true that *h(x) = Hx*.

Some general 2D transfrmations are as shown in the figure.

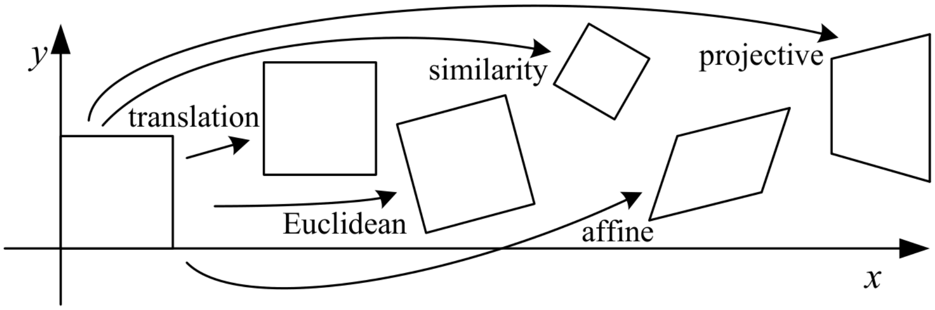

Figure 6: 2D planar transformations

Generalised expression for all these transformations can be written as $\bar{x}' = [M]_{3\times3}\bar{x}$, where [*M*] is a matrix defining the transform as listed here.

Transformation	**Matrix**	**# DoF**	**Preserves**
translation	$\begin{bmatrix}\mathbf{I} & \mathbf{t}\end{bmatrix}_{2\times 3}$	2	orientation
rigid (Euclidean)	$\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}_{2\times 3}$	3	lengths
similarity	$\begin{bmatrix}s\mathbf{R} & \mathbf{t}\end{bmatrix}_{2\times 3}$	4	angles
affine	$\begin{bmatrix}\mathbf{A}\end{bmatrix}_{2\times 3}$	6	parallelism
projective	$\begin{bmatrix}\tilde{\mathbf{H}}\end{bmatrix}_{3\times 3}$	8	straight lines

Each transformation also preserves the properties listed in the rows below it, i.e. similarity preserves not only angles but also parallelism and straight lines. The 2 x 3 matrices are extended with a third $[0^T\ 1]$ row to form a full 3 x 3 matrix for homogeneous coordinate transformations.

Here, **I** is the identity matrix, **t** is the translation, **R** is an orthonormal matrix $\begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$, s is an arbitrary scaling factor, **A** and $\tilde{\boldsymbol{H}}$ are arbitrary 2 x 3 and 3 x 3 matrices respectively.

Note - The action of a projective transformation on a co-vector such as a 2D line or 3D normal can be represented by the transposed inverse of the matrix, i.e. $\tilde{l}' = \tilde{H}^{-T}\tilde{l}$.

* 3D transformations are defined analogously to 2D transformations. 3 x 4 matrices are extended with a fourth $[0^T\ 1]$ row for homogeneous transforms as shown.

Transformation	**Matrix**	**# DoF**	**Preserves**	**Icon**
translation	$\begin{bmatrix}\mathbf{I} & \mathbf{t}\end{bmatrix}_{3\times 4}$	3	orientation	□
rigid (Euclidean)	$\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}_{3\times 4}$	6	lengths	◇
similarity	$\begin{bmatrix}s\mathbf{R} & \mathbf{t}\end{bmatrix}_{3\times 4}$	7	angles	◇
affine	$\begin{bmatrix}\mathbf{A}\end{bmatrix}_{3\times 4}$	12	parallelism	▱
projective	$\begin{bmatrix}\tilde{\mathbf{H}}\end{bmatrix}_{4\times 4}$	15	straight lines	⏢

10 Estimation of Homography

Images may not always be perfectly aligned. If we know the homography, then we can apply the necessary rectification. We follow the following procedure:

1. Select 4 points in a plane with known coordinates.
2. Form the system of equations.
3. Set $h_{33} = 1$ and solve the equations.

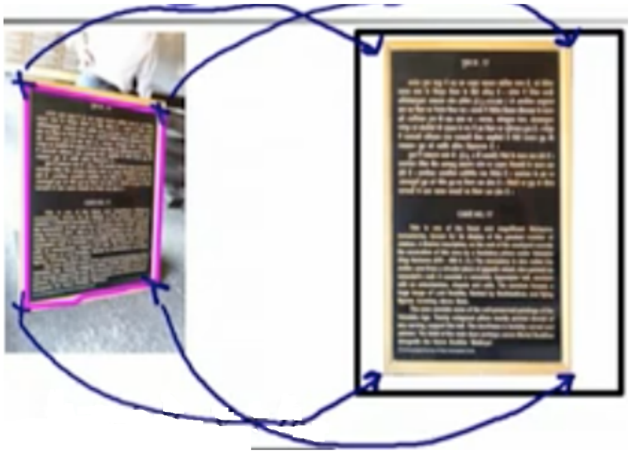

Figure 7: Sample image from different perspectives

$$x'(h_{31}x + h_{32}y + h_{33}) = h_{11}x + h_{12}y + h_{13}$$
$$y'(h_{31}x + h_{32}y + h_{33}) = h_{21}x + h_{22}y + h_{23}$$

We assume that the 4 correspondence points are known: $(x_i, y_i) \rightarrow (x'_i, y'_i)$. Setting $h_{33} = 1$, we have 8 variables, and 8 equations (4 points, 1 pair each). Hence, we can solve the system of equations and obtain the homography matrix.

10.1 Direct Linear Transformation (DLT)

The fundamental principle of DLT is to use number of correspondence points. We set x'_i $(= Hx_i)$ such that it is in the same direction as x_i. Thus, it satisfies $x'_i \times Hx_i = 0$. **H** is a 3 x 3 matrix, so we can also define it as

$$H = \begin{bmatrix} h^{1^T} \\ h^{2^T} \\ h^{3^T} \end{bmatrix} \implies Hx_i = \begin{pmatrix} h^{1^T} x_i \\ h^{2^T} x_i \\ h^{3^T} x_i \end{pmatrix}$$

- Now satisfy the cross product condition

$$\begin{bmatrix} 0^T & -w'_i x_i^T & y'_i x_i^T \\ w'_i x_i^T & 0^T & -x'_i x_i^T \\ -y'_i x_i^T & x'_i x_i^T & 0^T \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0$$

  $A_i h = 0$ and A_i is a 2 x 9 matrix.

- Setting $h_{33} = 1$, we get a system of equations of the form $\tilde{A}_i \tilde{h} = b_i \implies A\tilde{h} = b$. Here, we need to minimise $||\ A\tilde{h} - b\ ||$ and we will get a solution of the form $\tilde{h} = (A^T A)^{-1} A^T b$.

  where, $h = [\tilde{h}\ 1]^T$, dimension of *A*: 2n x 8, rank: 8, dimension of *h*: 8 x 1, dimension of *b*: 2n x 1.

- Another approach is to solve the homogeneous equation $Ah = 0$. Then, need to minimise $||\ Ah\ ||$ such that $||\ h\ || = 1$. We will get the solution as unit eigen vector of smallest eigen value of $A^T A$.

  where, dimension of *A*: 2n x 9, rank: 8, dimension of *h*: 9 x 1, dimension of *Ah*: 2n x 1, dimension of $A^T A$: 9 x 9.

■ ■ ■