# 1 Projective Geometry

1. Why is the Euclidean geometry not sufficient to model image formation ?
2. How many vanishing points the perspective projection of the edges of a cube define ?
3. Assume that a point $P_i$ is linearly interpolated between two points $P_1$ and $P_2$ in 3D. Is the perspective projection of $P_i$ the same linear interpolation between the projection of $P_1$ and $P_2$ ? Same question with an orthographic projection ?
4. What are the homogeneous coordinates of the line of $\mathcal{P}^2$ going through the points with homogeneous coordinates $(1, 0, 0)$ and $(0, 1, 0)$ respectively ?
5. Demonstrate that parallelism is preserved by affine transformations of $\mathcal{P}^2$.

# 2 Perspective Projection

Consider a perspective projection with focal length $f$:

1. In such a projection why do objects further away appear smaller in the image ?
2. Given an object (perspectively) projected in an image how should I modify the focal length of the projection so that the size of the object in the image is divided by 2 ?
3. Assume that two discs $S_1$ and $S_2$ of radius $R$ and $2R$ are perpendicular to the optical axis with their centers on the optical axis at distances $D_1$ and $D_2 \geq D_1$ from the projection center respectively.
   (a) Show that we observe two nested discs.
   (b) Assume $D_1$ fixed, above which distance $D_2$ will $S_2$ be fully occluded by $S_1$ ?
   (c) Assume the distance between the two discs to be fixed, i.e. $D_2 - D_1$ is constant, at which distance $D_1$ will $S_2$ and $S_1$ project onto the same disc ?

# 3 3D Modeling

1. Considering a point on a shape silhouette in an image, can we tell whether the corresponding point in 3D is a convex, concave or saddle point on the observed shape ?
2. An algorithm estimates the visual hull associated to $n$ silhouettes using a voxel grid of size $d^3$. What is the theoretical maximum number of *inside silhouette* tests required ? Is there a theoretical minimum number of such tests ?

3. Depending on the number of viewpoints available we can perceive a scene in 2D or 3D using adapted displays. How can we perceive 3D with a mobile phone, a stereo screen or a head mounted display ? explain how they differ.

4. In a multi-view stereo reconstruction, we seek for points that are photoconsistent, what does it mean to be photoconsistent and what are the assumptions made in such reconstruction ?

5. A 4D model (hologram) is composed of shape, appearance and sometimes motion information, explain what are these information ?

# 4 Image Mosaics

Assume that a camera acquires images while rotating about its optical center and consider the case where the camera takes **two** images. Between the two images, it carries out a rotation about the $Y$ axis:

$$\mathsf{R} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}$$

As for the intrinsic parameters of the camera, we suppose that they correspond to a simplified calibration matrix whose only unknown is the focal length $\alpha$:

$$\mathsf{K} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

It is known that there exists a projective transformation (homography) that links the two images. The goal is to estimate the transformation from a single point correspondence.

1. Write down the homography $\mathsf{H}$ in terms of the two unknowns, the rotation angle $\beta$ and the focal length $\alpha$.

2. Assume that two point correspondences are required to estimate the transformation $H$. Assume further that we have access to $10$ point correspondences among which $80\%$ are good correspondences. What is the probability to get a good estimation after 2 RANSAC iterations ?

3. Can we estimate $H$ using a single point correspondence ?

# 5 Plane Projection

The perspective projection of the point $P$ with homogeneous coordinates $(x, y, z, 1)$ onto the image point with coordinates $(u, v)$ can be modeled with:

$$(wu, wv, w)^t \sim K[R\,T](x, y, z, 1)^t, \tag{1}$$

where $K$ is the intrinsic parameter matrix:

$$K = \begin{bmatrix} k_u f & 0 & u_0 \\ 0 & k_v f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$R$ a rotation matrix in $\mathbb{R}^3$ and $T$ the $3 \times 1$ position vector of the world origin in the camera coordinate frame.

1. We consider points in the plane with equation $z = 0$, what kind of transformation becomes the above projection (1) with such points ?
2. Assume that the image plane is parallel to the plane with equation $z = 0$.
   (a) What kind of transformation is the above projection (1) with the points in the plane $z = 0$ ?
   (b) What is the minimum number of point correspondences required to estimate the transformation and how can we compute it given this number of correspondences ?
   (c) Consider two lines in the plane $z = 0$ that are parallel to the $x$ axis. Where does the projections of those two lines intersect in the image plane ?

# 6 Correction

## 6.1 Projective Geometry

1. The transformations that occur during image formation can modify lengths and angles (affine transformations) or even parallelism (projective transformations) in a perspective projection. The Euclidean geometry can only model rigid transformations which is not sufficient.
2. The edges of a cube in 3D are along 3 different directions and each direction in 3D defines one vanishing point in a perspective projection, thus the cube defines 3 vanishing points.
3. $P = (x, y, z)^t$. (t -> transpose)
   Linear interpolation: $(x_i, y_i, z_i)^t = \lambda(x_1, y_1, z_1)^t + (1 - \lambda)(x_2, y_2, z_2)^t, \lambda \in [0..1]$.
   Hence: $(x_i, y_i, z_i, 1)^t = \lambda(x_1, y_1, z_1, 1)^t + (1 - \lambda)(x_2, y_2, z_2, 1)^t$.
   Assume $M$ is a $3 \times 4$ projection matrix and $(u, v)$ the point image coordinates:
   $(w_i u_i, w_u v_i, w_i)^t = M \cdot (x_i, y_i, z_i, 1)^t = \lambda M \cdot (x_1, y_1, z_1, 1)^t + (1 - \lambda)M \cdot (x_2, y_2, z_2, 1)^t$.
   $(w_i u_i, w_u v_i, w_i)^t = \lambda(w_1 u_1, w_1 v_1, w_1)^t + (1 - \lambda)(w_2 u_2, w_2 v_2, w_2)^t$.
   $(u_i, v_i)$ is the linear interpolation of $(u_1, v_1)$ and $(u_2, v_2)$ with coefficient $\lambda$ only if $w_i = w_1 = w_2$ which is generically true only with an orthographic projection for which $w_i = w_1 = w_2 = 1$.
4. $L \sim (1, 0, 0)^t \times (0, 1, 0)^t \sim (0, 0, 1)^t$.
   Equivalently, assuming $L \sim (0, 0, 1)^t$ is the line at infinity, it is easy to see that the 2 points lie on this line since their $3rd$ coordinate is zero.
5. Assume two parallel lines of $\mathcal{P}^2$, they intersect at a point at infinity with therefore homogeneous coordinates $(a, b, 0)$. Given that the third row of an affine transformation of $\mathcal{P}^2$ is $(0\ 0\ 1)$ this point is transformed by any affine transformation into a point $(a', b', 0)$ which is itself on the line at infinity. Since linear transformations of $\mathcal{P}^2$ preserve incidence, this point is at the intersection of the transformed lines that are therefore parallel.

## 6.2 Perspective Projection

1. By definition in a perspective projection: $x(y) = X(Y)f/Z$ where $(x, y)$ are image coordinates and $(X, Y, Z)$ are 3D world coordinates. Looking at a length $L$ along the $X$ axis, at a distance $Z$, in 3D it is easy to see that its projections $l$ writes $l = Lf/Z$. Hence when $Z$ increases the length of the projection $l$ decreases.

2. Using $l = Lf/Z$ from the previous question we see that the focal length must be divided by 2.

3. (a) Consider the 3D point $(X, Y, D_1)$ on the external circle of $S_1$, i.e. $X^2 + Y^2 = R$. This point projects onto the point $(Xf/D_1, Yf/D_1)$ in the image and we can check that: $(Xf/D_1)^2 + (Yf/D_1)^2 = Rf/D_1$. Hence the disc projection in the image is delimited by a circle of radius $Rf/D_1$ hence it is a disc. The same reasoning applies to $S_2$ and we thus observe two nested discs centered on the optical axis with radius $Rf/D_1$ and $2Rf/D_2$ respectively.

   (b) $D_2 = 2D_1$.

   (c) Assume $D_2 - D_1 = l$ then the radius of disc $S_2$ writes $2Rf/D_2 = 2Rf/(l + D_1)$ and it equals the radius of $S_1$ when $Rf/D_1 = 2Rf/(l + D_1)$, hence when $D_1 = l$.

## 6.3 3D Modeling

1. In the image the curvature along the occluding contour (the projection of the 3D curve that delimits the visible region on the observed shape) is either convex or concave in which cases the shape is locally convex or hyperbolic (saddle point) respectively. Concavities are not observed by silhouettes due to occlusion.

2. In practice $n$ inside tests can be required, e.g., for a voxel inside the visual hull, and 1 test can be sufficient, e.g., for a voxel that projects outside all silhouettes.

3. To observe a scene in 3D we need to generate different viewpoints of that scene. On a mobile phone, we can generate 2D views that depend on the position and orientation of the mobile phone hence creating a 3D feeling when navigating around the scene with the device. On a stereo screen 2 views are generated from a fixed viewpoint. These views are processed by the human brain to generate 3D information, depth information in practice (fixed viewpoint). With a head mounted display, 2 views are generated that depend on the device position and orientation. These solutions differ by the number of views that are generated at a given time, one or two, and by whether these views depend on the position and orientation of the device over time. Only the head mounted display is providing a full 3D experience.

4. A 3D point is photo consistent in several image projections when the colors, at the corresponding image locations, are consistent, i.e. , similar. This is true for a point on the surface of the observed shape when the surface is Lambertian (diffuse) and not specular, since specularities appear differently depending on the viewpoint (brighter when the viewpoint is in front of the specular region).

5. A 4D model is composed of geometric information, typically a triangular mesh in 3D, appearance information, typically in the form of a 2D texture image that is associated to the geometry information, typically with texture coordinates for mesh vertices. Additional motion information can be provided with for instance vertex displacements over time in which case the texture information can also evolved over time, as an image sequence for instance.

## 6.4 Image Mosaics

1. By construction the transformation between the two image projections of the same 3D points writes (see the lecture Geometry 2): $H \sim K \cdot R \cdot K^{-1}$.
   Thus:
   $$H \sim \begin{pmatrix} \cos\beta & 0 & \alpha\sin\beta \\ 0 & 1 & 0 \\ -\alpha^{-1}\sin\beta & 0 & \cos\beta \end{pmatrix}$$

2. The probability to get 2 good correspondences is $0.8^2 = 0.64$. The probability to not get 2 good correspondences after 2 iterations is $(1 - 0.64)^2 = 0.1296$ and thus the probability to have a good estimation after 2 RANSAC iterations is $0.8704$.

3. In principle yes. Given a point correspondence we get 3 equations with 3 unknowns: $\beta, \alpha$ and the scale factor in the correspondence equation.

## 6.5 Plane Projection

1. For points with $z = 0$:
   $(wu, wv, w)^t \sim K[R\,T](x, y, 0, 1)^t$.
   Assume $R_1, R_2$ are the two first columns of $R$ then:
   $(wu, wv, w)^t \sim K[R_1\,R_2\,T](x, y, 1)^t$.
   The above equation above corresponds to a projective transformation (or homography) of the plane $z = 0$.

2. (a) The rotation $R$ is around the $z$ axis only and the two column vectors $R_1$ and $R_2$ present both a third coordinate which is zero (the rotation applies to $x$ and $y$ only). The transformation is then an affine transformation of the plane $z = 0$.

   (b) Let $A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix}$ be the associated transformation. A point correspondence brings 2 equations hence 3 correspondences are therefore required to estimate $A$.
   Given 3 correspondences $(x'_i, y'_i, 1)^t = A \cdot (x_i, y_i, 1),\ i \in [1..3]$., we can write:
   $A \cdot \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ 1 & 1 & 1 \end{bmatrix}$, and:
   $$A = \begin{bmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{bmatrix}^{-1},$$
   that can be solved assuming non colinear points (i.e., full rank matrices). With more than 3 correspondences the above system can be solved in the least square sense.

   (c) Affine transformations preserve parallelism hence the two lines intersect at infinity the image plane.