Computer Vision CITS4240

School of Computer Science & Software Engineering
The University of Western Australia

Single View Reconstruction

Introduction

We, as humans, can look at a photograph or a well executed painting and deduce considerable 3D information. Computer vision techniques that emulate aspects of this process are starting to emerge.

This work has resulted from the fusion of two areas of research:

- The use of projective invariants for object recognition. This involves finding properties of objects that are invariant to perspective projection. This allows you to recognize an object or feature no matter what view you have of it.

  This is typified by the work of Zisserman, Forsyth and Mundy at Oxford in the early 90s.

- Camera Autocalibration: This involves automatically determining camera calibration parameters from matched image points in stereo pairs of images, or motion sequences.

  Major contributors to this area of research were the group at INRIA (e.g. see Figure 1), France, headed by Olivier Faugeras, and the group at Oxford.

A Brief History of Perspective

Early art depicted scenes in a very symbolic form (see Figures 2, 3, and 4)

- Cave paintings showed animals in profile.
- Egyptian art depicted people with their heads in profile, torso in front view, and waist and legs in profile.
- Medieval art depicted people and objects very much like cardboard cut-outs stuck on a screen.

Figure 1: Merton College, Oxford. From a single view such as this we can deduce considerable 3D information

It was during the Italian Renaissance around 1430 that concepts of perspective were first developed and artists were first able to depict a sense of three dimensionality. Uccello was one of the first artists to use perspective.

Figure 2: The Battle of San Romano, by Uccello 1430. Note the alignment of the spears on the ground, the foreshortened image of the fallen soldier on the left and the road in the background. These features were sensational for the time.

The first book that described perspective was produced by Leon Battista Alberti in 1435. It is fascinating reading.

Vanishing Points and Vanishing Lines

Establishing vanishing points and vanishing lines are the fundamental operations in reconstructing 3D information from a perspective image.

Figure 3: The Hunt, Uccello 1460. Note the pattern of the trees vanishing towards an apparent horizon in the distance.

In perspective a set of parallel lines meet at a point. Two sets of parallel lines in different directions will give two vanishing points. These two vanishing points form a vanishing line for the collection of parallel planes defined by these two sets of parallel lines.

Figure 5 shows an example of a number of parallel lines and their vanishing point, V, on the image plane. Note that these parallel lines in 3D do not need to lie on the same plane. Geometrically, the vanishing point is the intersection point of the image plane with one of these parallel lines that passes through the optical centre of the camera. For instance, in Figure 5, the line OV, which has the same direction as lines AB, CD, and EF and which contains the optical centre O, intersects the image plane at the vanishing point V.

The above idea extends to *vanishing lines*. If we have a set of parallel planes in the scene, then their images should all overlap to form one identical line, namely the vanishing line. This idea is depicted in Figure 6. Alternatively, we can think of a set of parallel lines lying on these parallel planes in the scene (as in the Figure 5) giving a vanishing point in the image. Another set of parallel lines on these planes in the scene will give another vanishing point, and so on. All these vanishing points will be collinear in the image, forming the vanishing line. Geometrically, the vanishing line is the intersection of the image plane with one of the parallel planes passing through the camera's optical centre (see Figure 6).

We see that vanishing points and vanishing lines are both camera dependent.

The Horizon

The horizon is the vanishing line for the ground plane (see Figures 7 and 8). Anything below this line will be projected to a point below the horizon, and anything above is projected above the horizon. Note that the horizon is a property of the camera: two viewers at different heights will perceive different horizons.

Where do lines vanish to?

The camera calibration matrix gives you the location of the vanishing point. It maps 3D coordinates in the world to 2D coordinates in the image (see Figure 9). For example, the

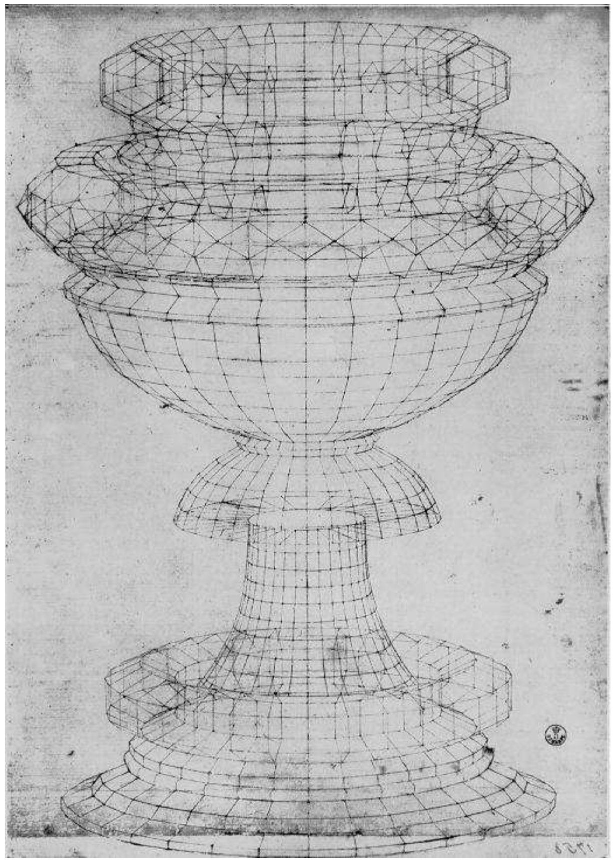


Figure 4: Drawing of a Chalice (by Uccello date unknown). This is a remarkable piece of work. It emulates what we take for granted with computer generated wire-frame graphics—except that it was done 600 years ago with pencil and paper!

vanishing point of the X axis is the location in the image where a point at infinity on the X axis will appear in the image.

What are the coordinates of a point at infinity on the X axis? Homogeneous coordinates allow us to represent, and do calculations with, points at infinity. A point at infinity on the X axis is represented by

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

Where will this appear in the image?

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (1)$$

$$= \begin{bmatrix} q_{11} \\ q_{21} \\ q_{31} \end{bmatrix} \quad (2)$$

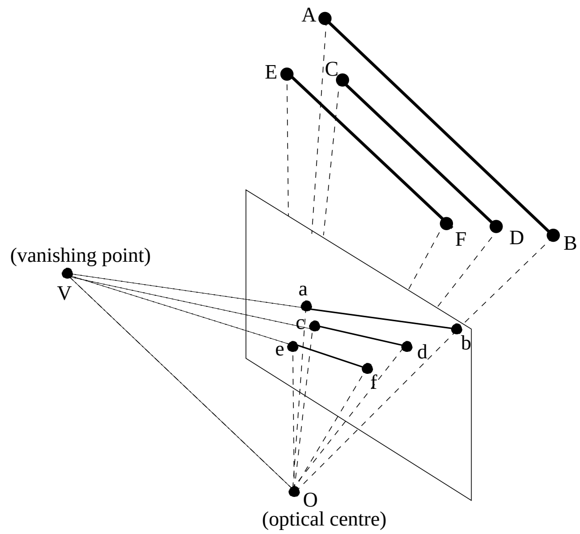


Figure 5: Parallel lines in the scene meet at a vanishing point in the image plane. Note that the vanishing point can be (and is often) outside the image boundary.

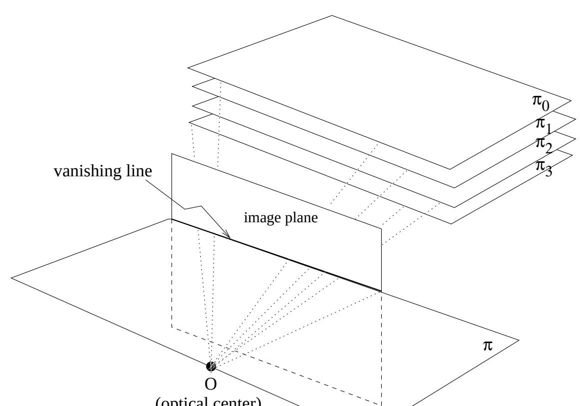


Figure 6: The vanishing line of a collection of parallel planes, $\{\pi_0, \pi_1, \cdots\}$, in the scene. The plane π is parallel to these planes and contains the camera's optical centre. Its intersection with the image plane gives the vanishing line.

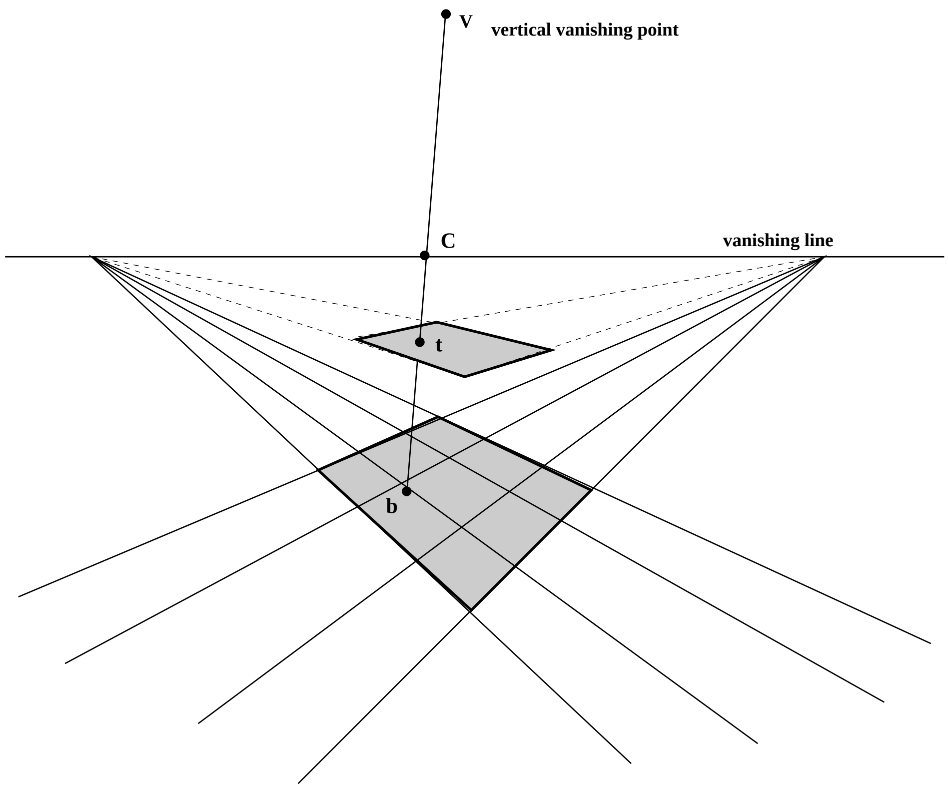


Figure 7: Perspective representation of the world. Parallel lines which are not orthogonal to the optical axis all meet at vanishing points, all of which lie on the horizon. The point **b** is in the base plane and **t** is a point directly above **b** in the world. The line through **b** and **t** defines the vertical, and meets the horizon at **C**, representing the camera height. Any two such vertical lines will meet at the vertical vanishing point **v**.

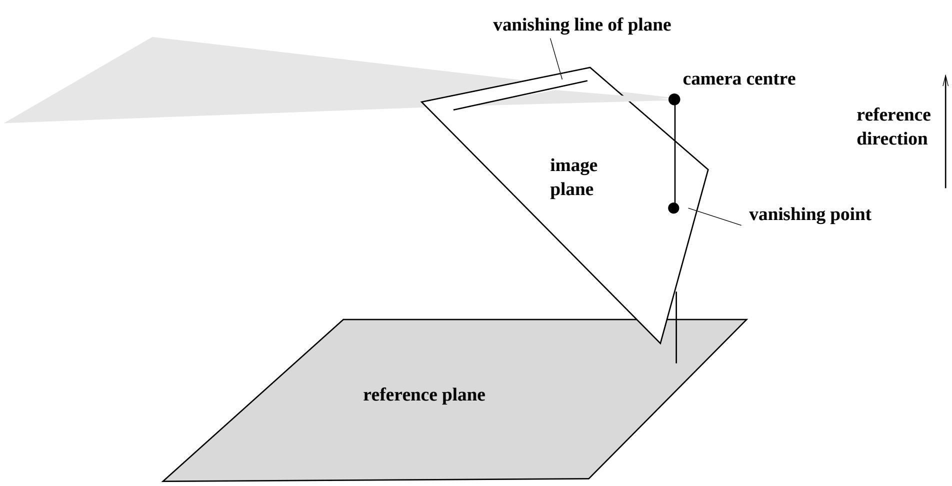


Figure 8: The horizon is defined either as the line joining two vanishing points defined from the reference plane, or as the intersection of a plane through the projection centre and parallel to the reference plane. A vertical vanishing point is defined as the intersection of the image plane with a line through the projection centre and perpendicular to the reference plane.

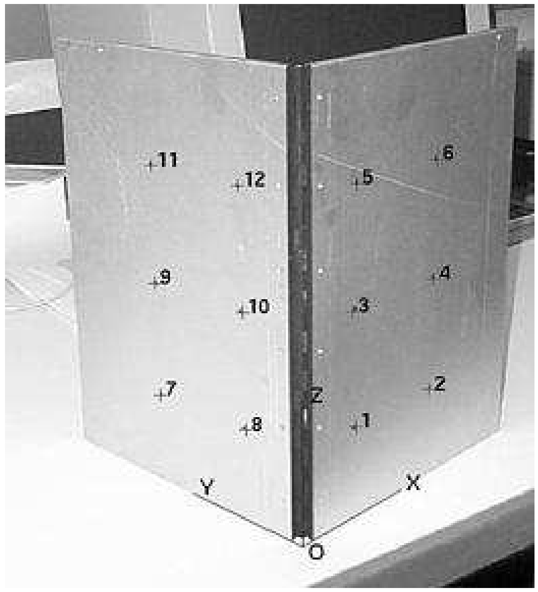


Figure 9: A calibration frame.

Thus, we see that the first, second and third columns provide the homogeneous image coordinates of the X, Y and Z vanishing points respectively. The fourth column tells you where the origin is.

Establishing Relative Sizes of Objects in a Perspective Image

We can use these ideas to take measurements in an image, whenever we have a clear way of establishing planar surfaces and measuring reference objects in the image. The image shown in Figure **??** is video footage of a suspect leaving a bank foyer. In this case, the police were interested in measuring the height of the suspect.

Figure 10: If the podium is 1.12m tall, how tall is the suspect?

Let the point h_r denote the given reference height (see Figure 11).

If we draw a line through the base points b_r and b_u to the vanishing line we get the

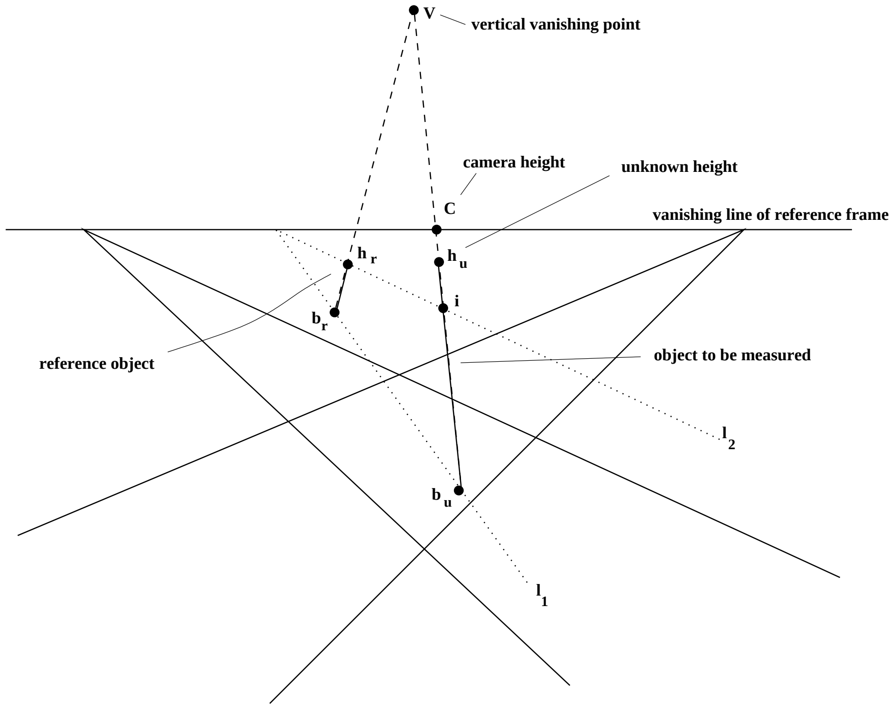


Figure 11: Geometry of relative height measurements in perspective. This figure is actually distorted, because the vertical vanishing point should really be underneath the reference plane when the view is taken from above. This is clear from the figure above.

vanishing point of that line.

A line from h_r to this vanishing point represents a line that is parallel (in the 3D world) to the line through the base points.

The point i, which is the intersection of this line with the line joining b_u and h_u, is the same height above b_u as h_r is above b_r.

However, we *cannot* use the ratio of the lengths $(h_u - b_u)/(i - b_u)$ times $h_r - b_r$ to estimate the unknown height, as ratios of lengths are not preserved under perspective. To solve this problem we have to find a property of the scene that is invariant (independent of) the perspective projection.

Geometric Invariants

An invariant of a geometric configuration is a function of the configuration whose value is unchanged by a particular transformation. For example, the distance between two points, or the angle between two line segments, are unchanged under a Euclidean transformation (translation or rotation).

There are a number of geometric invariants for projective transformations. Here we will illustrate one of the most important, the cross-ratio of four points on a line.

Suppose we are given a configuration of four collinear points, as shown below (Figure 12).

If we denote the distance between two points X_i and X_j and Δ_{ij} then one definition of the *cross ratio* is

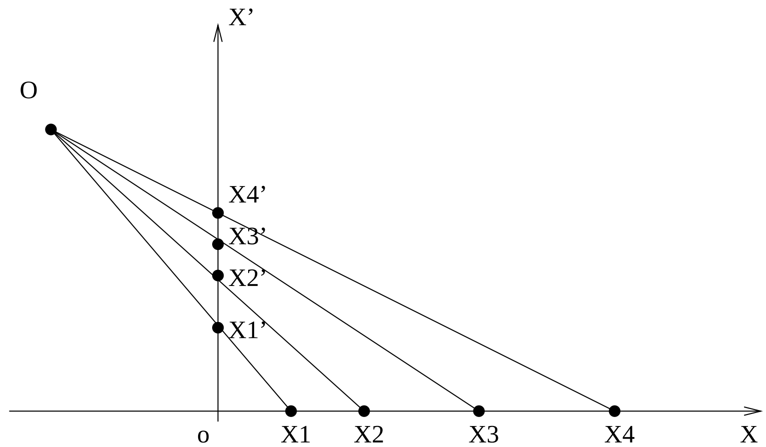


Figure 12: A one-dimensional construction of perspective viewing. The optical centre of the camera is **O**. Under perspective projection, the length, and ratios of lengths, on a line are not invariant, but ratios of ratios of lengths are.

$$I = \frac{\Delta_{13}/\Delta_{14}}{\Delta_{23}/\Delta_{24}} \tag{3}$$

$$= \frac{\Delta_{13}\Delta_{24}}{\Delta_{14}\Delta_{23}}. \tag{4}$$

There are many ways of forming the cross ratio between four points. One simply selects one point, say X_1, as a reference point. We compute the ratio of distances from that point to two others, say X_3 and X_4. Then compute the ratio of distances from the remaining point, X_2, to the same two points. The ratio of these ratios is invariant to perspective transformation.

The perspective transformation between the lines X and X' is given by

$$\begin{pmatrix} X'_i \\ 1 \end{pmatrix} = k_i \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} X_i \\ 1 \end{pmatrix}.$$

Now to see why the cross-ratio of four points on a line is preserved under such a transformation we note that the distance $(X'_i - X'_j)$ can be written as a determinant:

$$X'_i - X'_j = |S(X'_i, X'_j)| = \begin{vmatrix} X'_i & X'_j \\ 1 & 1 \end{vmatrix}.$$

Under the projective transformation above, the matrix $S(X'_i, X'_j)$ transforms as follows:

$$\begin{pmatrix} X'_i & X'_j \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} k_i X_i & k_j X_j \\ k_i & k_j \end{pmatrix},$$

and taking the determinant of both sides gives

$$|S(X'_i, X'_j)| = k_i k_j |M|.|S(X_i, X_j)|.$$

Substituting this relation into the expression for the cross-ratio gives

$$\begin{aligned}\frac{(X'_3 - X'_1)(X'_4 - X'_2)}{(X'_3 - X'_2)(X'_4 - X'_1)} &= \frac{|S(X'_3, X'_1)||S(X'_4, X'_2)|}{|S(X'_3, X'_2)||S(X'_4, X'_1)|} \\ &= \frac{k_3k_1|M||S(X_3, X_1)|k_4k_2|M||S(X_4, X_2)|}{k_3k_2|M||S(X_3, X_2)|k_4k_1|M||S(X_4, X_1)|} \\ &= \frac{(X_3 - X_1)(X_4 - X_2)}{(X_3 - X_2)(X_4 - X_1)}. \qquad (5)\end{aligned}$$

In summary, the cross-ratio is an invariant of any sets of four collinear points in projective correspondence. It is unaffected by the relative position of the line or the position of the optical centre, as shown in Figure 13.

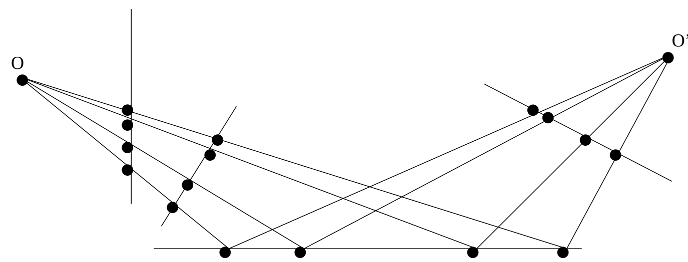


Figure 13: The cross-ratio of every set of four collinear points shown in this figure has the same value.

Returning to our height measurement problem, we take the points C, h_u, i and b_u as our four collinear points. We can form the cross ratio from distances measured in the image (in pixels)

$$I = \frac{(h_u - b_u)}{(h_u - i)} \frac{(C - i)}{(C - b_u)} .$$

Then we can solve for the world distances (metres) because we have

$$\begin{aligned}(h_u - b_u)_{world} &= I \times \frac{(C - b_u)_{world}}{(C - i)_{world}} \times (h_u - i)_{world} \qquad (6) \\ &= h_u - i + i - b_u \qquad (7)\end{aligned}$$

and thus we can solve for $h_u - i$ from the last equation, and then $h_u - b_u$ from the first. In these equations we have

- $(h_u - b_u)_{world}$ is the height of the reference object in the world;
- $(C - b_u)_{world}$ is the height of the camera from the reference plane, and
- $(C - i)_{world}$ is the difference between the reference height and the camera height.

Note that if we use the vanishing point V as one of the four points we get a degeneracy in calculating the cross ratio in world coordinates because we will have infinite distances. However, a solution can be found by other means (which will not be covered here).

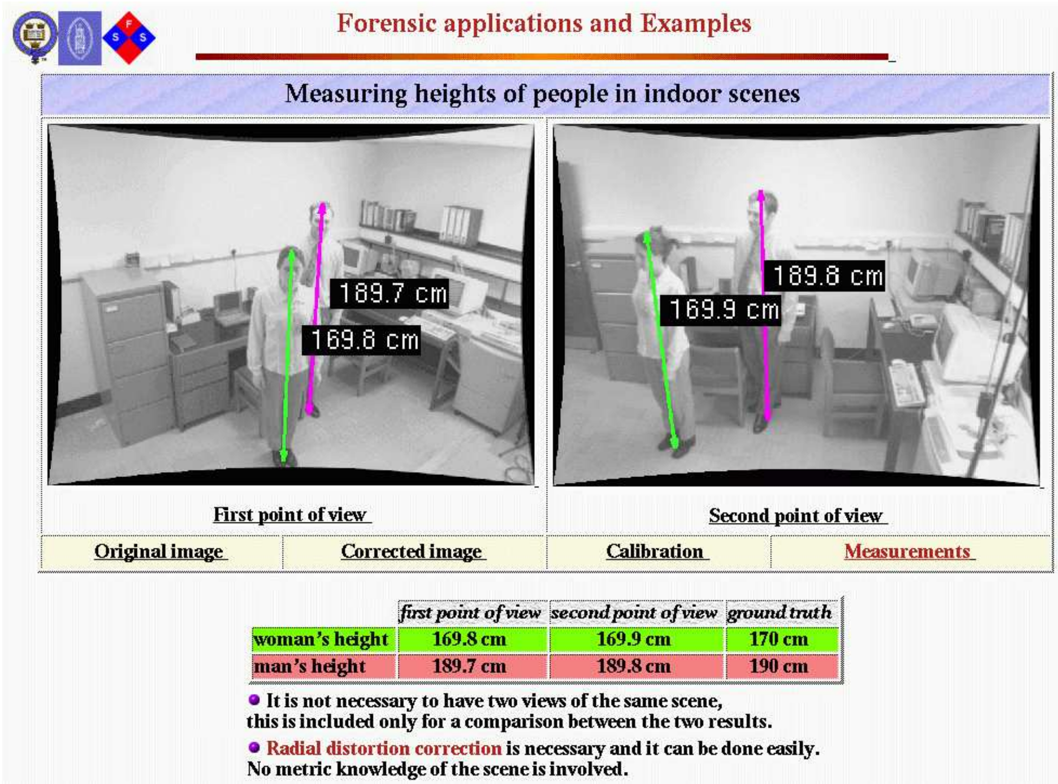


Figure 14: An example of height measurement taken from Andrew Zisserman's web pages.

See also Figure 14 for another example.

Clearly, if the four collinear points are permuted then the cross ratio computed using the formula given in (3) (or (4)) will be different. As an exercise, show that from the 24 possible permutation of the 4 points in general position (i.e., all the points are distinct) there are only 6 distinct cross ratio values.

Image Rectification

It is possible to undo the perspective distortion of a plane in the scene if we can find the vanishing line of the plane, and if we have two reference measurements of known lengths, or angles, in the scene. Before we do this we need to review some operations in 2D homogeneous coordinates.

2D Homogeneous Coordinates

Recall that a point in Euclidean 2-space is represented in homogeneous coordinates be a 3-vector $[x \;\; y \;\; 1]^T$.

The homogeneous coordinates $[x \;\; y \;\; 1]^T$ can represent both a point and a line.

The line passing through points p_1 and p_2 is obtained via the cross product

$$l = p_1 \wedge p_2.$$

For example, the line passing through the two points $[0\ \ 2\ \ 1]^T$ and $[3\ \ 0\ \ 1]^T$ as shown in Figure 15 is given by

$$\det \begin{bmatrix} i & j & k \\ 0 & 2 & 1 \\ 3 & 0 & 1 \end{bmatrix} = 2i + 3j - 6k \tag{8}$$
$$= \begin{bmatrix} 2 \\ 3 \\ -6 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

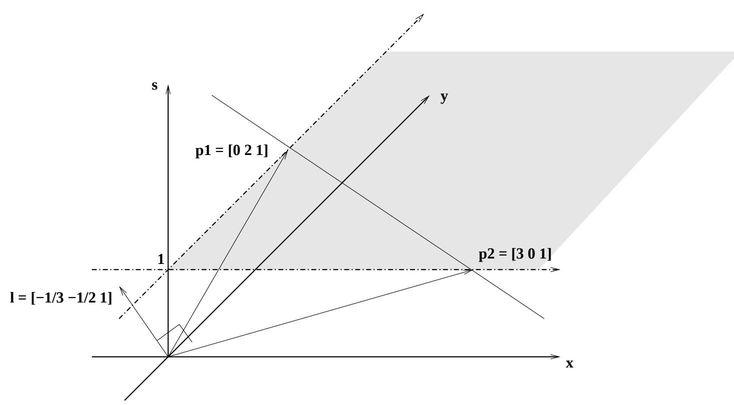


Figure 15: An example showing how to compute the coordinates of a line passing through two given points.

This cross product generates a vector that is perpendicular to both vectors, $[0\ \ 2\ \ 1]^T$ and $[3\ \ 0\ \ 1]^T$. This is the normal of the plane that passes through the origin and through the points $[0\ \ 2\ \ 1]^T$ and $[3\ \ 0\ \ 1]^T$. The intersection of this plane and the $z = 1$ plane forms the line.

This representation may seem a bit abstract, but is very concise, very powerful and is capable of representing points and lines at infinity.

There is a duality between lines and points in homogeneous coordinates. The intersection of two lines l_1 and l_2 is obtained via the cross product

$$p = l_1 \wedge l_2$$

For example, to find the intersection point of the two lines given by $[-\frac{1}{3}\ \ 0\ \ 1]^T$ and $[-\frac{1}{3}\ \ -\frac{1}{2}\ \ 1]^T$ (see Figure 16), we compute

$$\det \begin{bmatrix} i & j & k \\ -1/3 & 0 & 1 \\ -1/3 & -1/2 & 1 \end{bmatrix} = 1/2i + 0j + 1/6k \tag{9}$$
$$= [3\ \ 0\ \ 1]^T$$

This cross product generates a vector that is perpendicular to both plane normals, $[-\frac{1}{3}\ \ 0\ \ 1]^T$ and $[-\frac{1}{3}\ \ -\frac{1}{2}\ \ 1]^T$. If a vector is perpendicular to both plane normals then it must lie in both planes, and hence be the intersection of the two lines.

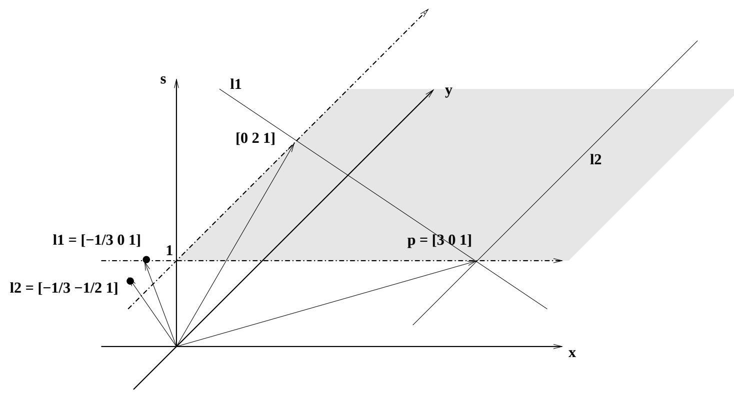
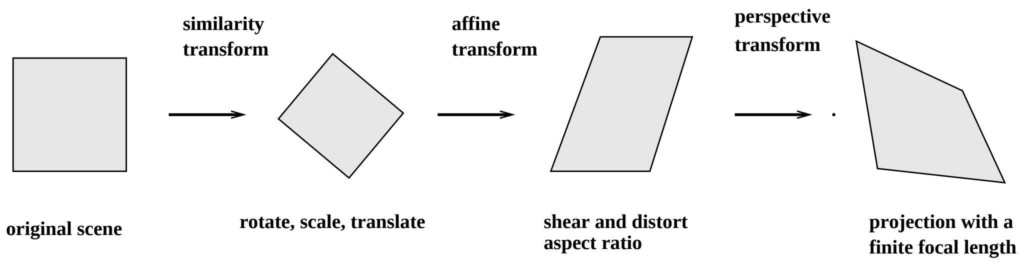


Figure 16: An example showing how to compute the intersection point of two given image lines.

Image Transformation of a Plane

We can think of the projection of a plane into an image as being the composition of three transforms.

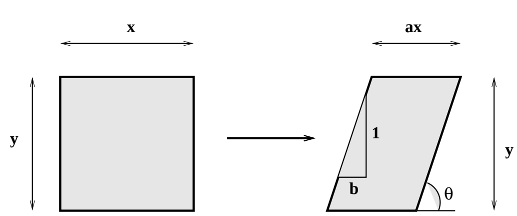


The similarity transform corresponds to you looking directly down on a rotated and translated scene with orthographic projection (infinite focal length). There may be a scale change involved, depending on how far away you are. The affine transform corresponds to you looking at the scene obliquely with orthographic projection. Here you get foreshortening and shearing (but parallel lines remain parallel). Finally the perspective transformation corresponds to keeping the same oblique viewing angle but now we introduce a finite focal length to introduce the perspective.

The affine transform has two degrees of freedom

$$\begin{bmatrix} a & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by \\ y \\ 1 \end{bmatrix},$$

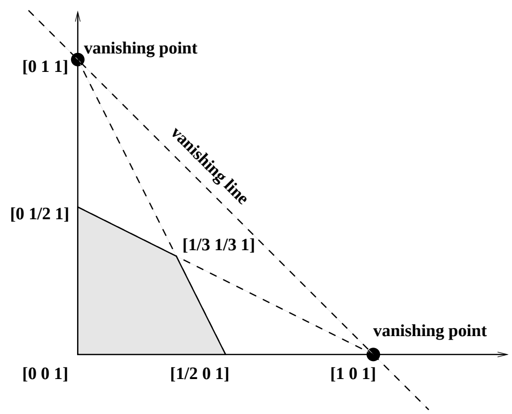
where a represents the scaling in x relative to y, and b is the cotangent of the shearing angle θ.



The perspective transform is represented in homogeneous coordinates as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c & d & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ cx+dy+1 \end{bmatrix} = \begin{bmatrix} \frac{x}{cx+dy+1} \\ \frac{y}{cx+dy+1} \\ 1 \end{bmatrix}$$

where $[-c \quad -d \quad 1]^T$ represents the homogeneous coordinates of the vanishing line of the plane.



Example: The scene above has vanishing points at $[0 \quad 1 \quad 1]^T$ and $[1 \quad 0 \quad 1]^T$. The homogeneous vector representing the vanishing line of the plane is $[-1 \quad -1 \quad 1]^T$, obtained from the cross product of these two points. The transformation describing this projection is thus

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

To invert this transformation and recover the view of the plane (up to an affine transformation) we simply apply the inverse of the transformation to the image points.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} . \begin{bmatrix} 0 & 0 & 1/3 & 1/2 \\ 0 & 1/2 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1/3 & 1/2 \\ 0 & 1/2 & 1/3 & 0 \\ 1 & 1/2 & 1/3 & 1/2 \end{bmatrix}.$$

Thus the four points $[0 \ \ 0 \ \ 1]$, $[0 \ \ 1/2 \ \ 1]$, $[1/3 \ \ 1/3 \ \ 1]$ and $[1/2 \ \ 0 \ \ 1]$ are transformed to the unit square having coordinates $[0 \ \ 0 \ \ 1]$, $[0 \ \ 1 \ \ 1]$, $[1 \ \ 1 \ \ 1]$ and $[1 \ \ 0 \ \ 1]$, because each of the resulting points is a homogeneous vector, and rescaling to ensure that the scale factor is 1 gives the desired outcome.

Rectification

To rectify (obtain a plan view of) a surface in an image all we need to do is find its vanishing line. The equation of this line allows us to invert the perspective transformation. The result will be an image that will still have an affine transform. If one can find two constraint conditions, such as knowledge of the ratio of two lengths in the image, and/or

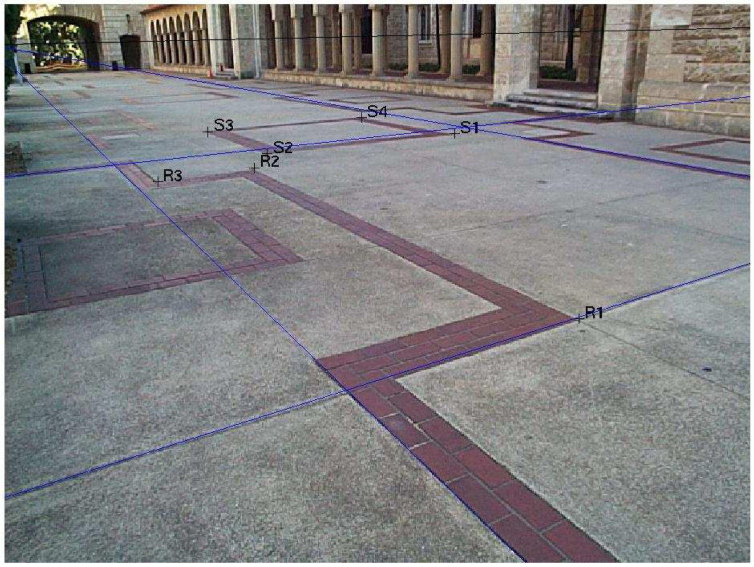


Figure 17: Image of the Vice Chancellery.

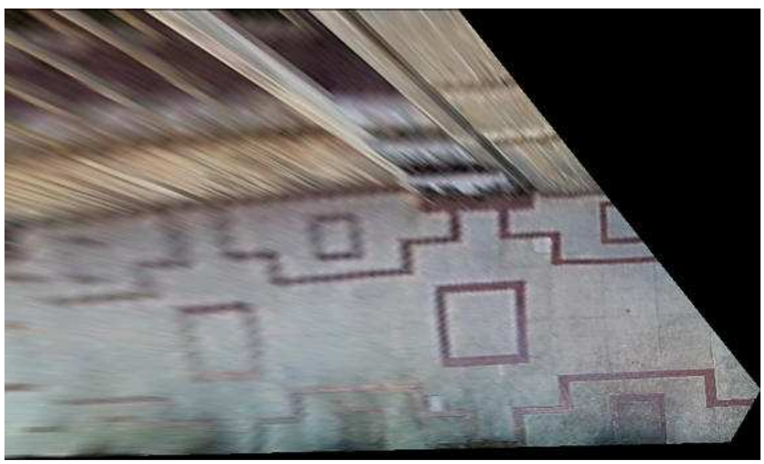


Figure 18: Rectified image of the base plane about the square tiles in front of the entrance.

knowledge of an angle between line segments in the image, one can then solve for the two degrees of freedom in the affine transformation.

Example (see Figure 17): Points are chosen manually from the image. Any two points on a straight line define the line by taking their cross product. The cross product of two such parallel lines defines a vanishing point. The cross product of two such vanishing points defines the vanishing line. Points R1–R3 and S1–S4 were also picked out manually. These are used to provide constraints on the affine transformation. These lines lead to the two vanishing points. The line across the top of the image is the vanishing line of the plane. It indicates the height of the camera relative to the objects in the scene.

There is still some distortion in the rectified image (see Figure 18) due to uncorrected lens distortion—lines are not quite straight in the image. There is also some poor digitising of reference points in the image on my behalf. However, this is a generally pleasing result, especially for the regions that were close to the camera. Look at the portion of the square in the paving that is cut off at the very left of the rectified image—now try to find it in the original image.

Note that the rectified image cannot include points that are too close to the vanishing line as these points are at infinity! This is why the rectified image is cut off on the left hand side where it is.

References

[1] Vishvjit S. Nalwa. *A Guided Tour of Computer Vision.* Chapter 9, Addison-Wesley Publishing Company, 1993.

[2] Leon Battista Alberti 1404-1472 *On painting and On sculpture : The Latin texts of De pictura and De statua by Leon Battista Alberti.* Edited with translations, introduction and notes by Cecil Grayson. Phaidon, London, 1972.

[3] J.G. Semple and G.T. Kneebone, *Algebraic Projective Geometry* Clarendon Press, Oxford, 1952.

[4] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision.* Cambridge University Press, 2000.

[5] David Liebowitz and Andrew Zisserman. “Metric Rectification for Perspective Images of Planes”. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 482-488, June, 1998.

[6] A. Criminisi, I. Reid and A. Zisserman. “Single View Metrology” *International Journal of Computer Vision*, vol. 40, no. 2, pages 123-148, November, 2000.