EPL668: Computer Vision

Lecture 2: Fundamentals

Dr Paris Kaimakis

University of Cyprus
Department of Computer Science
pariskaimakis@cs.ucy.ac.cy

January 2015

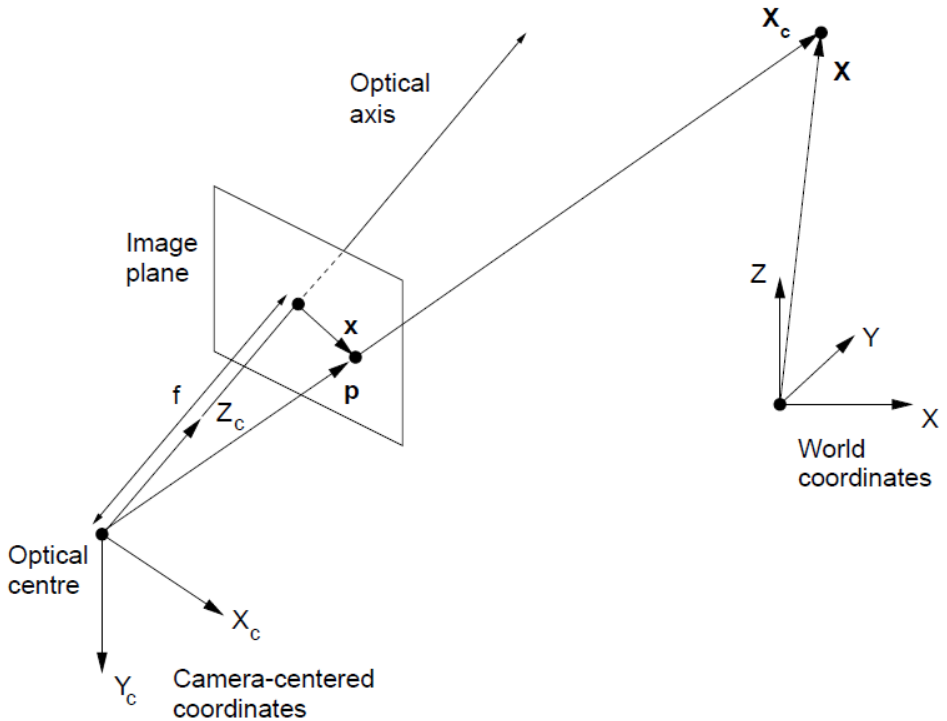
Typically, the first stage in computer vision is to perform **feature extraction**. The aim is to reduce the data content of the images while preserving their useful information.

Corner detection is also useful and achieves higher compression rates. Corner features are particularly useful for motion analysis.

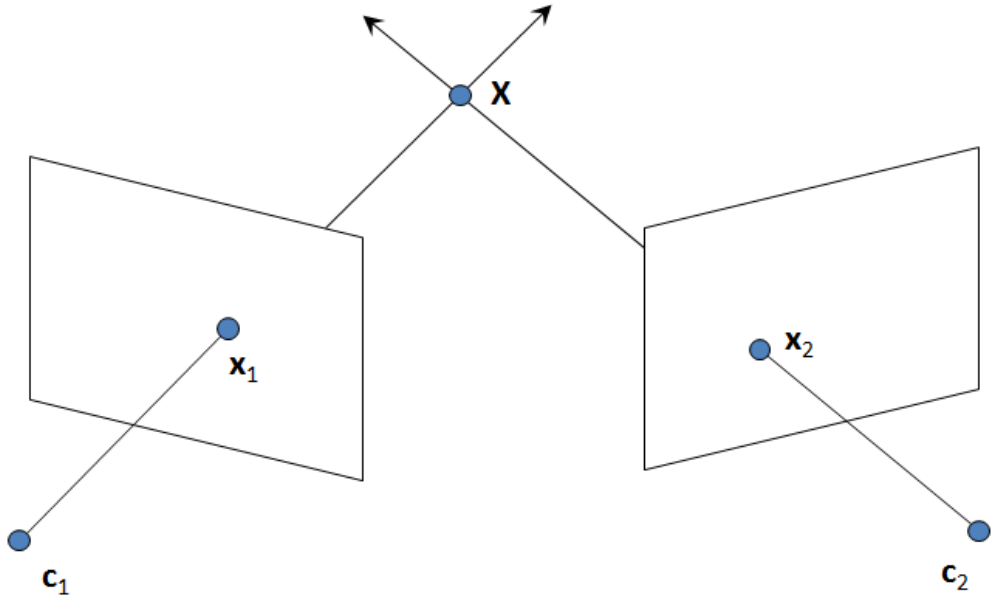
Typically, the first stage in computer vision is to perform **feature extraction**. The aim is to reduce the data content of the images while preserving their useful information.

Edges are perhaps the most intuitive feature to use. Edge-detection algorithms produce a line-drawing version of the image.

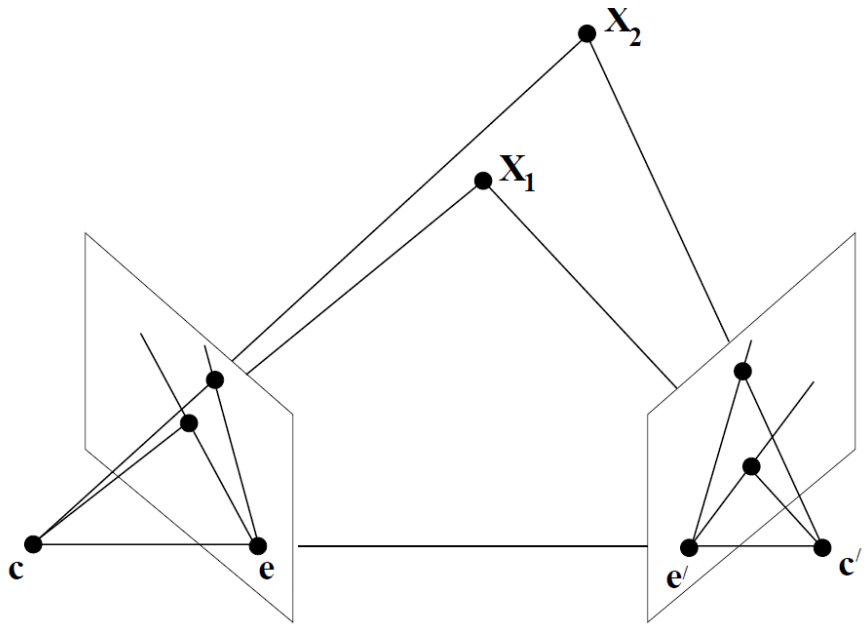
Before we attempt to interpret the features extracted from the image, it is crucial to understand how they were formed. In other words we must understand how a camera works.



Stereo vision is the luxury of having more than one cameras. By capturing images of the same scene from different viewpoints we can obtain depth information based on triangulation of corner features.

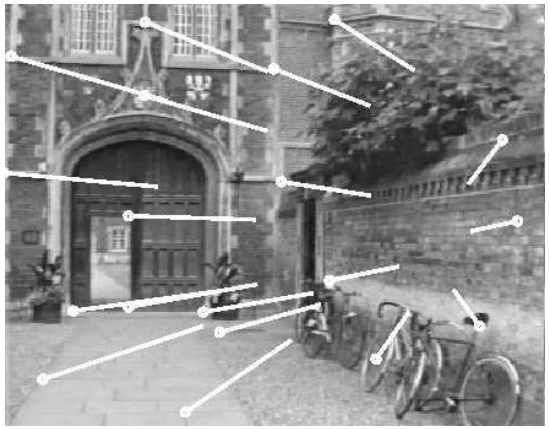


In most cases we have hundreds of features in each image. In **shape from stereo vision** we must ensure that features in the left and right images are correctly associated. This is the well-known and difficult **correspondence problem**.



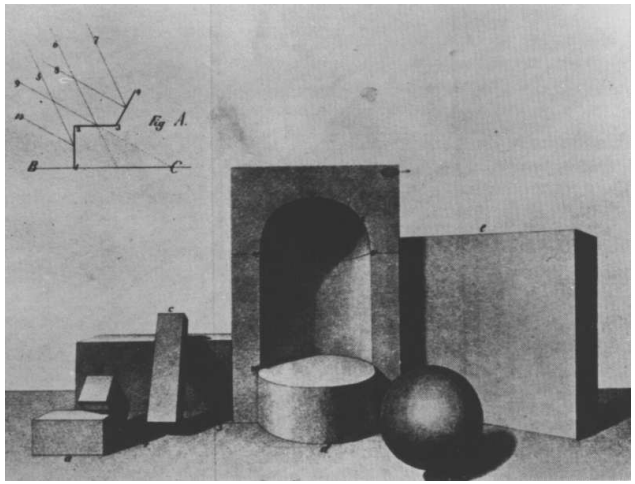
We can also allow a single camera to move and capture several images of the scene. This is known as **shape from motion**.

As the camera moves, the motion of some features (in this case corner features) is *tracked*. The feature trajectories allow us to calculate the shape of the scene and the motion of the camera.



It is possible to infer the shape of the scene from the **shading** observed in the image.

We commonly assume isotropic surface reflectance, and a top-lit scene.



Our own visual system goes beyond the mechanisms that we have described so far, and allows us to successfully interpret images under a wide range of conditions.

More surprising is our ability to interpret a wide range of images with limited cues.

Remember...

The challenge for computer vision scientists is to resolve image ambiguity by using extra information. But where does the extra information come from?

We have two options:

- Use more than one image of the scene.
- Make assumptions about the scene.

In the **geometrical framework**, we:

1. reduce the information content of the images by extracting salient features.
2. use projective transformations to model the perspective projection that has created the features in the image.
3. take as many different images of the scene as it is necessary, in order to force the transformation to become invertible.
4. **calculate** 3D shape and motion by inverting the projective transformation.

In the **statistical framework**, we:

1. reduce the information content of the images by extracting salient features.
2. use projective transformations to model the perspective projection that has created the features in the image.
3. use models of the objects being represented in the images.
4. **infer** shape and motion based on the combination of available visual and nonvisual information.