Chapter 2

Omnidirectional Vision

This chapter introduces the concept of omnidirectional vision and mentions why it is useful to have wide fields-of-view. This is followed by a description of the different types of camera systems that achieve omnidirectional vision, together with their advantages and disadvantages. The chapter concludes by mentioning why it is beneficial for omnidirectional camera systems to have a single viewpoint of the surrounding world.

2.1 Omnidirectional Vision

The word omnidirectional is derived from the Latin word *omnis*, meaning *all*. And **omnidirectional vision** refers to the ability to *sense light in all directions*. This means that the field of view covers a full sphere – a field of view of 360° by 360° in both horizontal and vertical directions. A related term is **panoramic vision**. Although this term is normally used in a more generic sense, and includes those views that cover a 360° field of view (or less) in the horizontal direction only (or any other 360° cross-section) and with a limited field of view in the vertical direction.

Achieving full omnidirectional vision is not possible in practice, since the sensor itself (however small) must hide part of the view. Therefore the term omnidirectional vision is taken to include any sensor that can see a single hemisphere of view (or at least provide a panoramic view).

2.2 Motivation

The main motivation for the use of omnidirectional sensors is the very wide field-of-view they offer, when compared to conventional (linear)[1] sensors. This has several advantages.

The importance of large fields-of-view is illustrated by the case of human vision. In human vision, the detailed (*foveal*) part of the retina covers a mere 6° [YAGI99]. This is augmented by *peripheral vision*, which extends human vision to a much wider field-of-view of approximately 200° by 135° (although still far from being panoramic). This leads to a better *sense of awareness* of what is happening in the world and the *feeling of immersion*, of being part of the world.

A look at the different eye designs in the biological world shows that having a wide field-of-view is an important characteristic for certain kinds of animals [TINB80 §2]. And most probably, this characteristic played an important role in the evolution of animal vision – it is marked as one of the major eye design categories in biologist M. Land's 'landscape of eye evolution' (as quoted in [FERM98]). Examples include the spherical compound eyes of insects with their wide field-of-view (but low visual acuity), fish eyes, and the corneal eyes of birds that approximate panoramic vision by having their eyes on opposite sides of their heads (while maintaining visual acuity). The requirements of flight played a major role for the evolution of these types of eyes. [FERM98] shows mathematically that a 'spherical eye' is superior as regards to certain problems like 3D motion estimation (especially for the 6 degrees of freedom of movement in flight).

Similarly, computer vision applications benefit from large fields-of-view, for example, because of their unobstructed views (in the case of automated surveillance systems) and by helping to reduce the *aperture problem* (in the case of applications that derive motion from images) [TRUC97 §8.3.2].

[1] Throughout this document, when the term *conventional sensor* or *conventional camera* is mentioned, it is used to mean *linear cameras*.

2.3 Omnidirectional Camera Systems

There are several different camera systems available for acquiring omnidirectional (panoramic) images. These can be broadly classified into two main categories – camera systems that can acquire an omnidirectional image in a *single shot* (that is, acquire the full field-of-view at a single moment in time), and others that build an image in successive steps (different parts of the field-of-view acquired over an extended period of time). Those cameras that capture an image in a single shot can be viewed as being *true* omnidirectional sensors. Below is a brief mention of the different designs currently available.

2.3.1 Rotating Camera Systems

Rotating cameras were one of the earliest methods used for generating panoramic photographs. The basic idea is to use a conventional camera (that is, with a limited field-of-view[2]) and to rotate it around its focal point to acquire the panorama in successive steps. A full rotation gives rise to a cylindrical 360° panorama. This category can be further sub-divided into two: those cameras systems that generate a single panoramic image directly[3] and those that generate multiple conventional images as output.

Examples of the first type are *swing lens cameras*, where the lens moves from side to side (first patented in 1843), and *panning cameras*, where the whole camera rotates around its focal point (first commercial model appeared in 1904) [IAPP]. Because of how they operate, these cameras are also referred to as *slit cameras*, or *line-scan cameras*, as the panoramic image is built "line by line" while the camera/lens rotates/swings along the horizontal direction.

The multiple image system, also called the *panoramic mosaicing* method or *multi-shot* method, can be easily implemented using a standard camera. This is done by taking a sequence of individual images while the camera is rotated. Some process of *image stitching* is then required in order to combine the sequence of images into a

[2] In the context of this document, a limited field-of-view camera is taken to mean a standard camera. This normally covers a range from 35° to 45°, and rarely exceeds 60° [FREE93 pp. 146-149].

[3] Although these type of cameras produce a single panoramic image, they cannot be termed *single-shot*, as the image is not produced at a single moment in time, but over an extended period of time.

panorama (see Figure 2.1 below). This method has become quite popular in recent years and image-stitching software has become readily available.

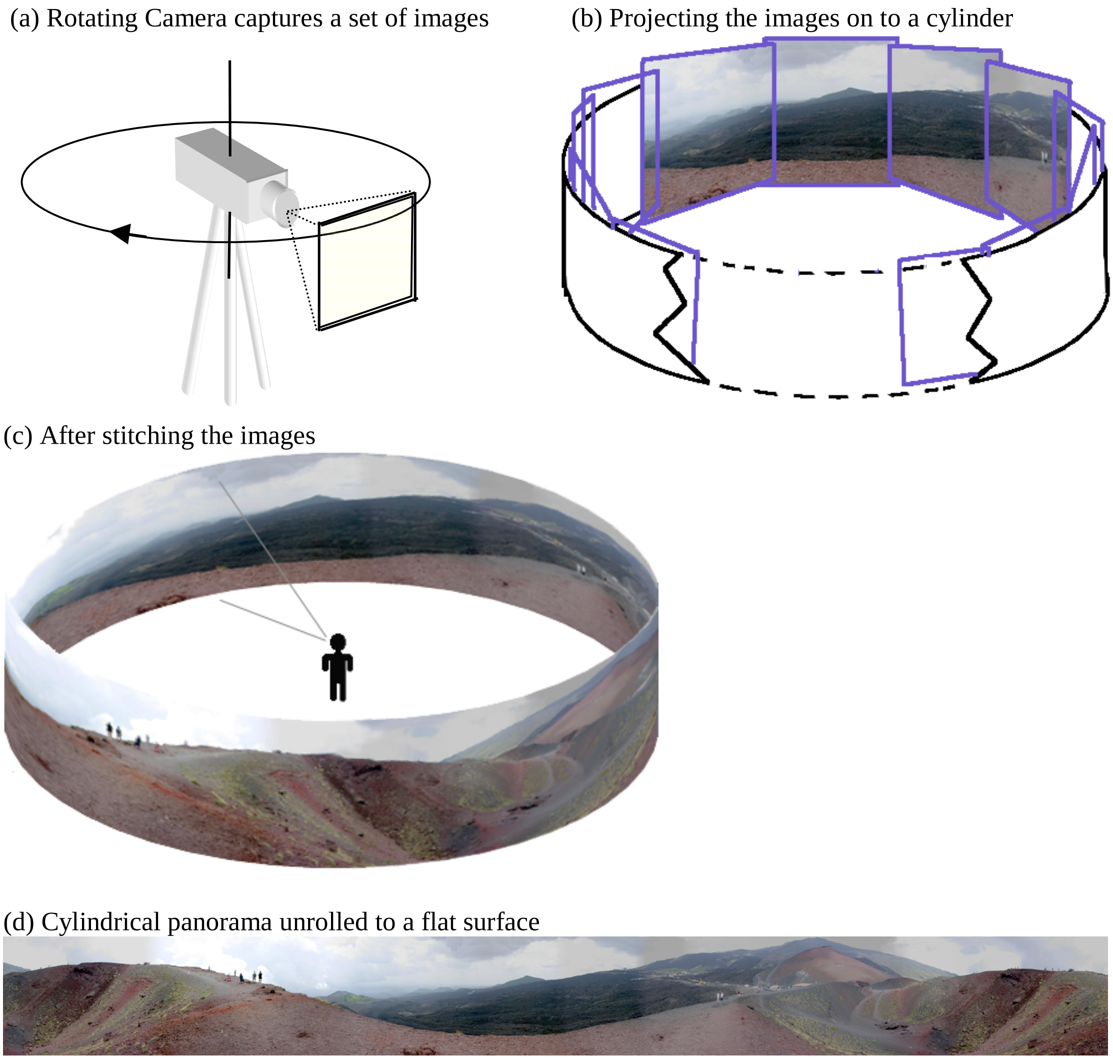

Figure 2.1: A 360° panorama created by rotating a camera around its focal point. (Mt. Etna panorama, Sicily. Taken by the author on Aug 2002. A total of 18 images were used).

2.3.2 Multiple Camera Systems

The basic idea behind multiple camera systems is to use several cameras and to arrange them into a tightly packed camera cluster. The cameras within this cluster all point outwards from the cluster's centre in different directions, so that the overlap between their individual fields-of-view is minimised (see Figure 2.2). The combined field of view of the cluster will then cover a much wider area and this can be used to approximate an omnidirectional sensor, with each camera's image covering a portion of the final omnidirectional image.

This is somewhat similar in idea to the *compound eyes* of insects. A compound eye consists of several units (*ommatidia*), each with its own lens and light-sensitive cells. The camera cluster represents the compound eye, and each camera represents the individual ommatidium.

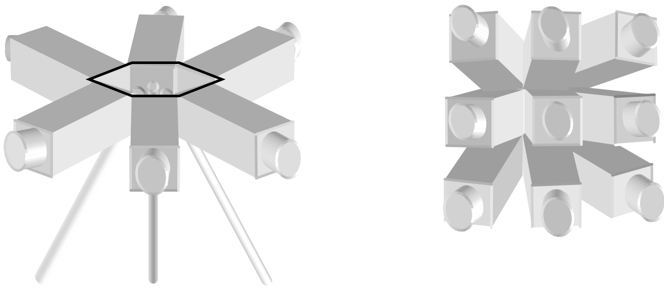

Figure 2.2: Multi-Camera systems

2.3.3 Fish-Eye Lens Systems

Cameras equipped with fish-eye lenses are usually able to acquire a whole hemisphere (360° by 180°) of view or more in a single image. This is achieved through the use of a lens system with very short focal lengths and lens elements with a high refractive power. The word *fish-eye lens* was first introduced by R. Wood in a paper in 1911 and the first commercial patent was issued in 1964 [KUML00].

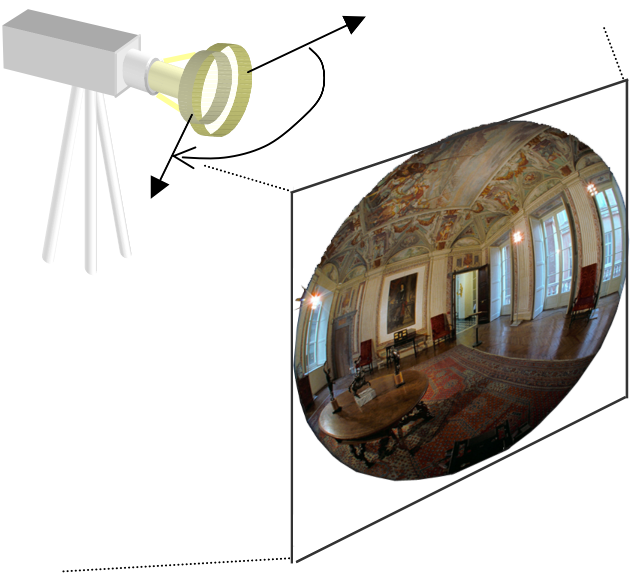

Figure 2.3: Fish-Eye camera systems. (Fish-eye image courtesy of Gianni Maselli).

Fish-eye lenses can be divided into two categories [FREE93 p.150]. The first type includes those lenses that generate a *circular* image, where all of the incoming light is diverted into a disc-based image. When used with a rectangular sensing device (or film), this leaves the corners empty (see example given in Figure 2.3). The second type of fish-eye lenses generates a *full-frame* image, and leaves no unused areas, but loses some parts of the scene (180º across the diagonal of the image).

2.3.4 Catadioptric Systems

Catadioptric camera systems acquire omnidirectional images through the use of curved mirrors. A conventional camera is used and it is oriented in such a way so that it looks towards the mirror and captures what the mirror is reflecting of the surrounding scene (see Figure 2.4). Because of its curved shape, the mirror will in general capture a wide field-of-view, typically a hemisphere or more. The first patent for a catadioptric system was issued in 1970 [REES70].

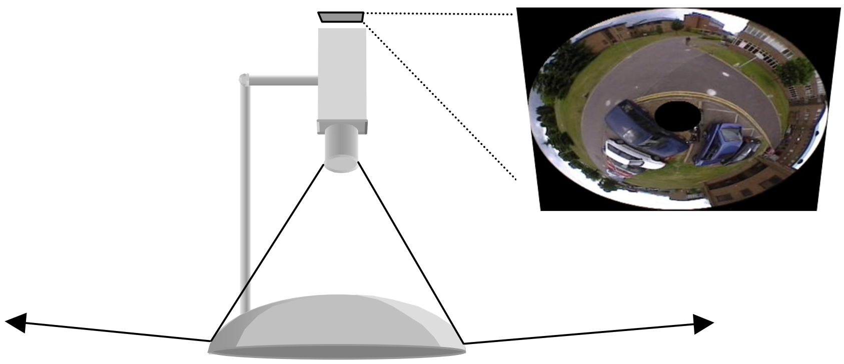

Figure 2.4: Catadioptric camera systems. (Omnidirectional image is a single frame from one of the PETS2001 dataset).

2.3.5 Others

There are some other omnidirectional camera systems that do not fit in neither of the above categories, are specialised systems or hybrids of one or more of the above. Amongst these, one can find the *Panoramic Annular Lens*, which combines reflective surfaces with refractive surfaces (glass block) in a compact design [ZHU99].

A variation of the multiple camera cluster systems that use mirrors and inward-looking cameras (looking towards the mirrors), such as the system implemented by

[MAJU99] and another by [TAN02]. Another system is the *slice camera* as proposed by [NAYA00]. This combines the rotating camera method with a convex mirror to get high-resolution panoramas.

2.4 Comparison of the Omnidirectional Camera Systems

Each of the omnidirectional camera designs introduced in the preceding sections has its own advantages and disadvantages. These are explored in this section, together with a brief mention of some applications using these systems.

2.4.1 Rotating Cameras

The main advantages of rotating cameras are:

- A1. These systems provide a very simple and cheap way of acquiring very wide field-of-view images (especially the panoramic mosaicing method).
- A2. The horizontal resolution of the final panoramic image is not limited by the physical resolution of the camera. It only depends on the angular speed of rotation.
- A3. Rotating cameras do not suffer from *blind spots* – that is, areas blocked by the camera itself.

The main disadvantages are:

- D1. They do not acquire a panoramic view in a single shot. So these systems are limited to acquiring panoramic images of static scenes and are not usually used for real-time applications. This greatly limits their use in computer vision. In some cases, an offline image stitching process is also needed.
- D2. These cameras have moving parts, so requiring more power, introduce more delays to image acquisition and response, and are prone to mechanical failure.
- D3. There are also alignment and calibration issues: the camera must rotate around an axis that is perpendicular to the optical axis of the lens, rotation must be done exactly around the focal point (to avoid *parallax effects*), the angular speed of rotation must be constant (to achieve uniform horizontal resolution) and some image overlap (data redundancy) is required for image stitching to be done successfully.

D4. Rotating cameras are usually limited to acquiring horizontal panoramas, with a limited vertical field-of-view.

Applications:

Because of their non real-time nature, rotating cameras are usually limited to capturing static environments. Some applications are listed below:

- Perhaps the best-known application is the *QuickTime VR* image-based rendering application described in [CHEN95]. This uses a sequence of images acquired by a rotating camera to build a panoramic image.
- Another system is the one by [SHUM97], where the image sequences are captured by a hand-held video camera. To avoid image registration problems, a local alignment technique is used to overcome small motion effects (partially solving disadvantage D3, above).

2.4.2 Multiple Cameras

Advantages:

A1. This camera design allows omnidirectional images to be acquired in real-time.

A2. Multiple camera clusters can be designed to acquire very high-resolution omni-directional images.

A3. Arbitrary field-of-views can be covered by these systems and the cluster's shape can be adjusted as required.

Disadvantages:

D1. These systems are quite expensive. The design is usually not very compact, especially for systems with a large number of cameras, and it is quite hard to achieve a tight cluster.

D2. It is hard to maintain the alignment and calibration of a multiple camera cluster – if the position of one camera changes, then the whole cluster needs to be re-calibrated.

D3. The process of stitching the images from each camera into a final omnidirectional image can be computationally expensive and camera clusters tend to require large data bandwidths.

D4. It is usually not possible to make the centre of projection of the cluster's cameras to coincide, as the centre of projection of each individual camera resides inside it. This problem can be minimised by making the cluster more compact.

Applications:

- A multiple camera cluster system is used by [SWAM99], called the 'panoramic polycamera'. This cluster consists of 4 cameras arranged in a compact cube-like structure, capturing a horizontal 360º panorama. Wide-angle lenses were selected for the individual cameras, in order to minimise the number of cameras used, and hence achieve a tighter cluster (to reduce D4, above). A 6-camera version is also suggested to achieve a full sphere of view.
- Another camera cluster is that of [MAJU99], which is used for immersive teleconferencing. This system uses a cluster of 12 standard cameras to capture a 360º by 90º panorama. To achieve a (nearly) single centre of projection (partially solving D4), the mirror-based variant of the camera cluster method is used (see §2.3.5). Local alignment methods are also used to avoid having to re-calibrate the cluster if a single camera is modified (solving D2). This allows each camera to be treated individually when stitching the panoramic image.
- Finally, a multi-camera system is implemented by [NEUM00] for the purpose of offline and online remote viewing over a network. It is made up of 5 video cameras, which capture a 360º by 76º panorama with a final resolution of 3520 x 480 pixels – showing the advantage of multi-camera systems for producing high resolution panoramas (mentioned in A2; the individual cameras had a horizontal resolution of 704 pixels).

2.4.3 Fish-Eye Cameras

Advantages:

A1. Fish-eye cameras are 'true' omnidirectional sensors – they acquire a very wide-of-view through the use of a single camera and can operate in real-time.

A2. These systems are quite easy to implement and existing conventional cameras can be made omnidirectional by attaching a fish-eye lens.

Disadvantages:

D1. Fish-eye lenses produce non-uniform resolution images, by their very nature of having to 'compress' a wide field-of-view into a flat disc-shaped image.

Normally, the central part of a fish-eye image has high resolution, and poor resolution at the periphery (*edge compression*) – called the 'magnifying glass' effect in photography.

D2. Another disadvantage (related to D1), is that since only one camera is being used, the overall image resolution of the omnidirectional tends to be quite low (especially when compared to the other previous methods of the rotating cameras and multi-cameras).

D3. Good quality fish-eye lenses can be quite expensive and tend to be complex in structure and hence, large. For example, the fish-eye lenses reviewed by [KUML00], consist of between 8 and 12 individual refractive lens elements.

D4. Fish-eye lenses have widely varying (radial) distortions across their hemispherical field-of-views. These distortions are due partly to manufacturing error, but more seriously due to the very nature of fish-eye image formation [PERŠ02]. Hence, cameras equipped with fish-eye lenses, diverge from the pinhole camera model and it is usually not possible to achieve a single centre of projection. Correcting for these distortions is usually computationally expensive.

D5. Fish-eye lenses suffer from a number of optical defects, amongst these: lateral colour shifts, field curvature, and astigmatism [KUML00]. These defects are made worse by the short focal length nature of fish-eye lenses. They also suffer from uneven illumination – light fall-off at the periphery of the image causes the border areas to appear darker (due to *vignetting* and other effects). This problem increases as the field-of-view of the lens increases [OKAT01].

D6. As mentioned in §2.3.3, the image generated by a fish-eye lens can be either circular or full-frame. The circular image means that the corner parts of the rectangular sensing device are not used (no light falls on them), and so the effective resolution is lower than the device's true resolution. For the full-frame case, some of the light gathered by the fish-eye lens is never registered by the sensing device. Also, the vertical field-of-view varies along the image.

Applications:

- A fish-eye lens with a field-of-view of 183º is used in [BAKS02] to acquire omnidirectional images with a hemispherical view.

2.4.4 Catadioptric Cameras

Advantages:

A1. Like in the case of fish-eye lenses, catadioptric systems acquire omnidirectional images with a single camera and operate in real-time.

A2. One of the great advantages of catadioptric systems is the ability to design a system that has a single centre of projection (unlike fish-eye lenses).

A3. Because of the use of a reflective surface, rather than a refractive one, catadioptric systems do not suffer from optical effects such as light fall-off (vignetting), astigmatism, colour shifts, and other effects usually associated with (wide-angle) lenses. It is also easier to make curved mirrors fee of optical distortion than lenses – light is reflected from the surface and does not have to pass through the material.

Disadvantages:

D1. Catadioptric systems suffer from non-uniform resolution variation across the omnidirectional image. This is caused by the curvature of the mirror, which on the other hand, is a necessary requirement for achieving wide field-of-views.

D2. As only one camera is used, catadioptric systems tend to have limited resolution. This unlike the case of a multiple camera cluster, where the effective (total) resolution of the cluster is more than that of an individual camera, even after allowing for image overlap.

D3. Because of the use of curved mirrors, catadioptric cameras suffer from the optical effect of *field curvature* (or *spherical aberration*) [NAYA99]. The curvature of the mirror causes points at infinity to be focused onto a curved surface[4] (instead of a plane, as happens for pinhole cameras). Two points in the scene, which are at the same depth, will require different focal settings depending on where their image appears on the mirror. The effect of this is that it is harder to get a focused image for catadioptric systems, resulting in *focal blur* (especially for the peripheral parts of the omnidirectional image). This focusing problem is made worse, the smaller and more compact the mirror is.

D4. The introduction of the mirror, in addition to the camera sensor, makes alignment and calibration of catadioptric systems harder.

[4] This curved surface is called a *Petzval surface* in optics.

D5. Catadioptric systems suffer from a *blind spot* in the centre of the image, where the camera sensor sees its reflection in the mirror. This reduces the effective field-of-view of the sensor.

D6. Catadioptric systems generate circular omnidirectional images (see Figure 2.4). As a result (like fish-eye lenses), when used in conjunction with a rectangular sensing device (film), some parts of the image are not used. The effective resolution of the mirror-camera system is lower than the resolution of the camera on its own.

Applications:

Catadioptric cameras are often the method of choice in computer vision for acquiring omnidirectional images. A selected list of applications using catadioptric camera systems will be given at the end of chapter 3, after a more detailed discussion about the different sub-types of catadioptric systems.

2.5 The Single Viewpoint Constraint

In computer vision, in addition to being able to acquire images in real-time, it is often highly desirable for a camera system to have a **single centre (point) of projection**. This means that all the rays of light entering the camera intersect at a single point, the centre of projection (see Figure 2.5(a)).

Cameras with a single centre of projection are also called *central projection cameras* or *central cameras* for short. And those violating this property, as *non-central cameras* [HART00 §5].

The benefits provided by this are that it makes working with the camera mathematically convenient and simplifies the geometric modelling of the system (for example, by using the *pinhole camera model*). Most of the work in computer vision has been done for camera systems having a single projection centre and their geometric properties are well understood (the field of *projective geometry*). Some work, for example that involving the relationships between multiple cameras (multi-view geometry [HART00]), in fact requires such single centre of projection.

Cameras having a single centre of projection are also said to have a **single viewpoint**, since the camera 'sees' the world from a single point, the centre of projection. This is an especially important property for an omnidirectional camera. Because of its ability to see all around it, an omnidirectional camera can be said to come close to being able to sample the *plenoptic function* at a single point in 3D space. The plenoptic function is an ideal function proposed by [ADEL91] to represent the complete visual information (what a sensor can potentially see) in 3D space, and is defined as: $P = f(\theta, \phi, \lambda, t, V_x, V_y, V_z)$. An omnidirectional camera is able to see (most) of the horizontal and vertical viewing directions (ϑ, ϕ), from its single viewpoint in space (V_x, V_y, V_z) at any particular instance in time *t* (see Figure 2.5(b)).

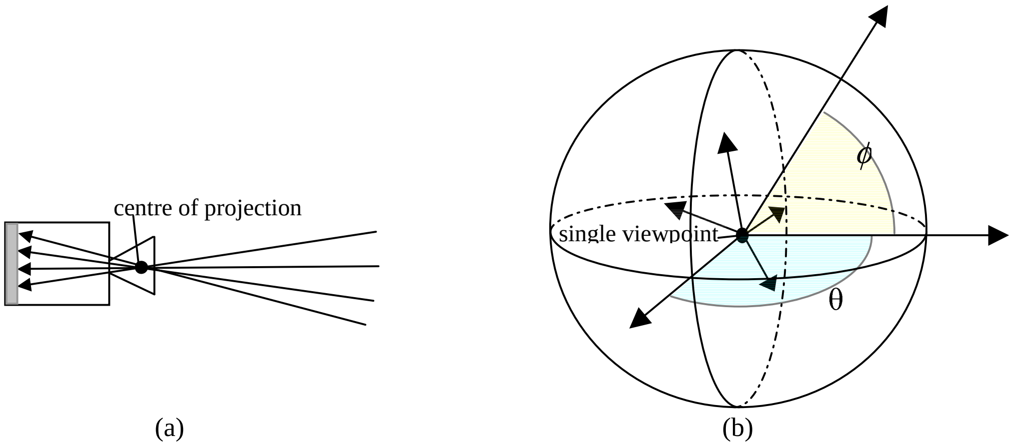

Figure 2.5: (a) Single Centre of Projection, and (b) Single Viewpoint.

From the omnidirectional image it is then possible to construct *virtual views* by mapping the brightness values sensed by the omnidirectional camera to the 'image' that would be generated by some other camera, for example, the *perspective images* of conventional cameras. It is as if a virtual perspective camera is placed so that its viewpoint coincides with that of the omnidirectional camera. The single viewpoint property ensures that any virtual views generated from the omnidirectional images are geometrically correct (that is, in terms of projective geometry).

For omnidirectional cameras that do not produce perspective views (what the human eye sees), such as fish-eye and catadioptric ones, the process of generating virtual perspective views can be used to 'dewarp' the omnidirectional image into a human-viewable form.

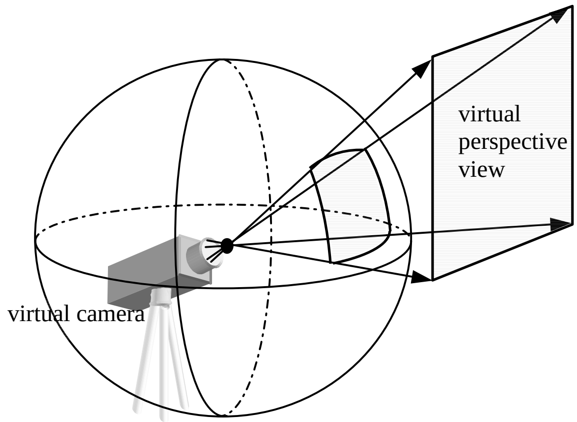

Figure 2.6: Generating virtual perspective views.

2.5.1 Non-Central Omnidirectional Cameras

Even though it is highly desirable for omnidirectional cameras to obey the single viewpoint constraint, some applications use non-central omnidirectional cameras because of the greater freedom it allows in the camera design [BAKS01]. In these cases, light entering the camera system does not pass through a unique point, but rather passes through a set of points. This locus of viewpoints forms a surface in 3D space called a *caustic*[5]. In many cases, the omnidirectional camera can be designed in such a way as to make the caustic surface small and compact, to approximate a single viewpoint [SWAM01].

2.6 Conclusion

This chapter has introduced the concept of omnidirectional vision and listed some of the advantages it provides. The different ways in which omnidirectional vision can be realised were described, mainly, rotating cameras, multi-camera systems, the use of fish-eye lenses, and catadioptric cameras. This chapter also described the importance of the single viewpoint constraint for omnidirectional vision sensors and how this can be used later on to create virtual views.

[5] Caustic is a mathematical process of deriving a new curve (the 3D locus of viewpoints in this case) based on an existing curve (the curved optical element – lens, mirror – used by the omnidirectional camera system) and a point (the sensor's pinhole). The word caustic is usually used interchangeably for both the derivation process and the derived curve.

As a conclusion, one can say that from among the different camera types mentioned in this chapter, catadioptric systems are often the method of choice in computer vision, because they operate in real-time, acquire a hemisphere (or more) of view and may be designed to obey the single viewpoint constraint. The rest of this document will concentrate on catadioptric cameras, which is the camera system used for the application developed for this thesis.