Chapter 10

Omnidirectional Vision Systems

This chapter describes omnidirectional vision systems and discusses performance aspects.

True omnidirectional cameras should be able to obtain views from all directions. Some authors makes a distinction between omnidirectional and panoramic sensors (for example [1]), but it is widely accepted that the term omnidirectional is applied to any system that has a 360 ° field of view on the horizontal direction, producing a single image from its surroundings. The limits of the field of view on the vertical direction is system dependent.

Omnidirectional systems are not new[1], but they have a renewed interest in computer vision, robotics, surveillance and in other applications, such as in applications where multiple users share one camera. Recently google maps used omnidirectional cameras to store images for their Street View technology [2].

There is a number of architectures developed over the years, including multiple cameras, rotating cameras, wide angle lenses (dioptric systems) and the use of mirrors (catadioptric systems). The word "catadioptric" comes from a combination of mirrors (catoptrics) and lenses (dioptrics). Table 10 shows some of the types of omnidirectional systems encountered in the literature and it is an attempt to classify them. There are systems that do not fit perfectly in a single class, being perhaps more closely defined as a combination of two types.

The special case of the catadioptric system, formed by a single camera associated with a single mirror, is the focus of this chapter. In catadioptric systems the mirror is normally convex, with the camera situated below or above the mirror in order to produce the appropriate images (see an example in figure 10.1). Several curves for the main mirror were evaluated in the past. A complete survey of the different mirror geometries is described in [3] (see also [4] and [5]).

10.1 Catadioptric Systems

The most common mirrors for omnidirectional catadioptric systems are: spherical, conical, paraboloidal, and hyperboloidal. The latter is becoming more popular, despite the technical difficulties of manufacturing it. There are two reasons for its increasing popularity. Firstly, hyperboloids have a single projection centre, allowing to project panoramic images with minimum distortion when compared to other mirror geometries. Secondly, hyperboloidal mirrors have a good balance in terms of the resolution at different points of the mirror.

Spherical mirrors are widely available for other purposes, as this is relatively easy to manufacture. However, this shape does not have a single projection centre. [7] have proposed an approximation and successfully built a catadioptric system using a single sphere with the camera embedded on it, but the study indicates that the resolution of the camera cannot be larger

[1] The first know omnidirectional camera was developed in the early 1900s

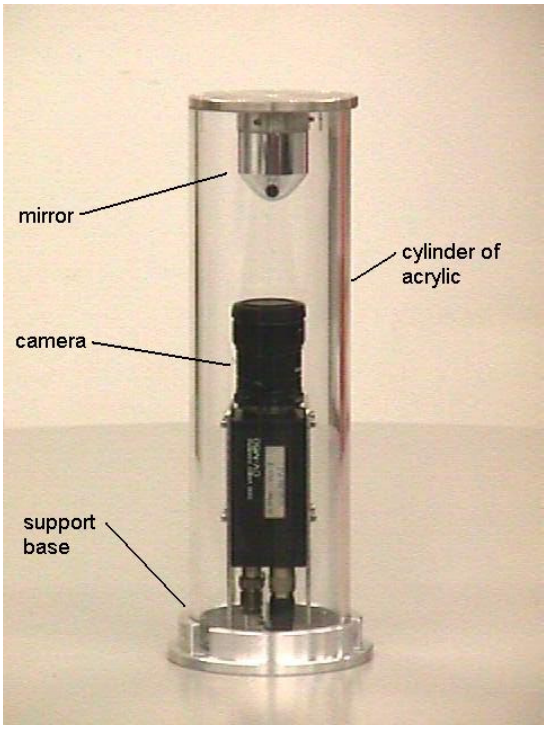

Figure 10.1: A catadioptric system with a hyperboloidal mirror, used to collect data for this work.

than a certain limit in order to work without any significant distortions.

The conics are attractive curves to use in optics because they have such special properties. Conical mirrors are also popular because they are easy to manufacture. Paraboloidal and Ellipsoidal mirrors have been used as well, although they present difficulties in the manufacturing. The limitation of these three curves is the fact that they do not present a single projection centre. Hyperboloidal mirrors, on the other hand, can be fitted in a way that they have a single projection centre. Hyperboloidal mirrors are easier to calibrate and limit distortions of the omnidirectional image.

10.2 Virtual Cameras in Omnidirectional Vision Systems

One important advantage of the catadioptric systems is that they do not need to stitch images from different cameras. However, the image is still very deformed, so usually before it can be used for further processing it needs to be projected appropriately. The mirror's geometry affects two important properties of the system: resolution and focusing [4].

The resolution of the catadioptric system can be different than the resolution of the sensor in the camera. For example, when using ellipsoidal and hyperboloidal mirrors, the resolution is highest around the periphery of the mirror. Also, as cameras are usually rectangular, part of the area covered is not utilised because the omnidirectional image is circular.

The single projection centre (also called single view point) allows the system to obtain perspective images which are relatively free from distortions. This is achieved by computing the position of pixels that are in a plane perpendicular to a ray passing thought the focus of the mirror. This produces an image that is the equivalent of another image acquired by a perspective camera in that position. Also, the entire image can be projected on a cylinder, producing a panoramic image. The panoramic image is free from vertical distortions.

Table 10.1: Omnidirectional Digital Camera Systems

Cameras	**Type**	**Sub-type**	**Image geometry**	**Typical Resolution**	**Pre-processing**
multiple cameras	aligned non-aligned	joined CCDs	cylindrical spherical	high high	stitch stitch/affine
single camera	dioptric	wide view lens rotating ([6])	deformed cylindrical	low/medium high	projection stitch/unblur
	catadioptric	planar spherical conical paraboloidal hyperboloidal ellipsoidal	deformed	low/medium	projection

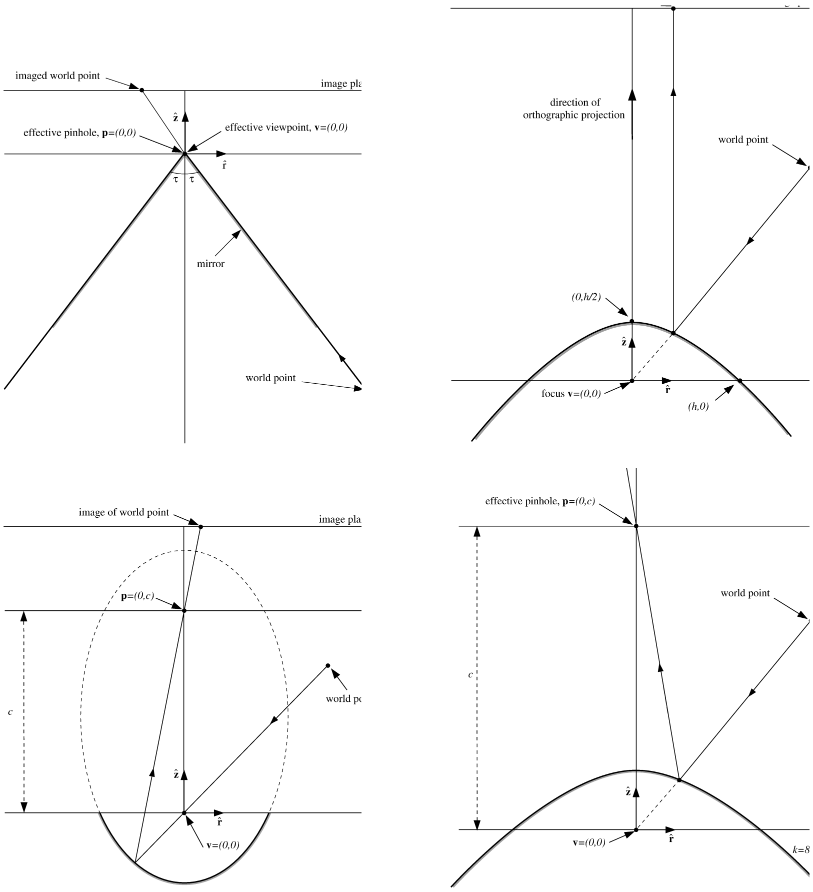

Figure 10.2: Schematics of mirrors based on conic curves: a) Conical. b) Paraboloidal. c) Ellipsoidal. d) Hyperboloidal. [3]

Besides the single projection centre condition, the mirror's curve can be computed according to its application. Depending on the hyperbola's parameters, the distribution of the resolution over the omnidirectional image can change. This allows for the customisation of the mirror. For example, [8] studied the conditions for constant resolution on omnidirectional cameras.

10.2.1 Implementation of Hyperbolic Systems

Svoboda [9] implemented a catadioptric system using a perspective camera associated with a hyperboloidal mirror. Grassi and Okamoto [10, 11] implemented a similar system, manufacturing the mirror with a special aluminium alloy and an ultra-precision CNC lathe machine.The difference between the two systems is that in [10] the mirror is located on top of the camera. In this way the resolution is distributed conveniently for a mobile robot. Figure 10.1 shows the acquisition system used by [10].

Based on [12] and [4], Grassi and Okamoto [10] computed the equation for a hyperboloidal mirror according to resolution constraints and according to the physical characteristics of the perspective camera adopted to be used in the system. The omnidirectional image can be projected according to the needs of the application. It is possible to compute two basic types of projections, panoramic and perspective (figure 10.3).

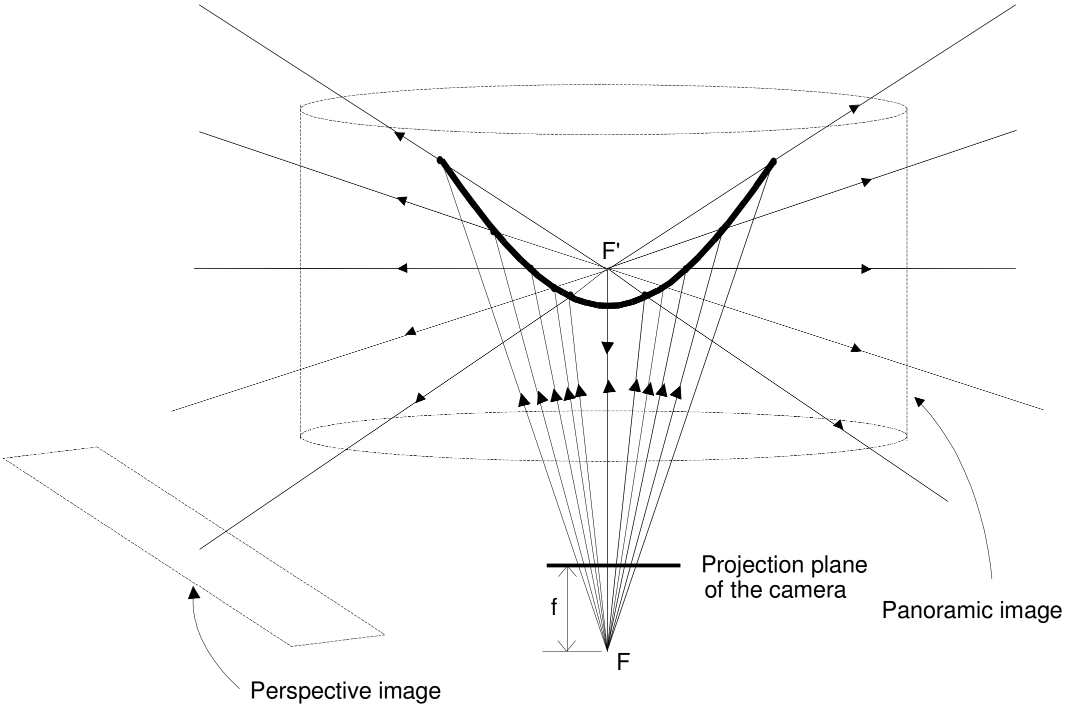

Figure 10.3: Perspective and panoramic projections [10].

10.2.2 Panoramic Projections

The panoramic projection is computed taking the shape of the specific hyperbola into consideration. By doing the computation in this way, the panoramic projection is free from vertical distortions. From Grassi and Okamoto [10], one can map the positions of pixels from the omnidirectional image to a panoramic image (see figure 10.4):

$$u = r_p \, cos\left(\frac{2\pi u_{pn}}{H_{pn}}\right) \tag{10.1}$$

and

$$v = r_p \, sin\left(\frac{2\pi v_{pn}}{H_{pn}}\right) \tag{10.2}$$

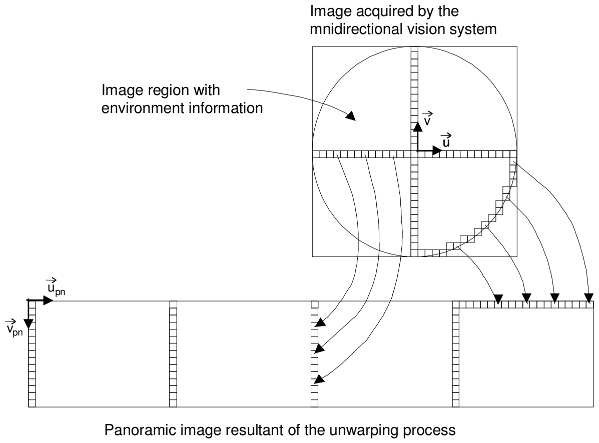

Figure 10.4: Panoramic projection [10].

The factor r_p reflects the shape of the hyperbola, and is a function of the eccentricity (r_p is a function of $2e$ in figure 10.5) The positions are stored in a look-up table, so the computation can be done in real-time.

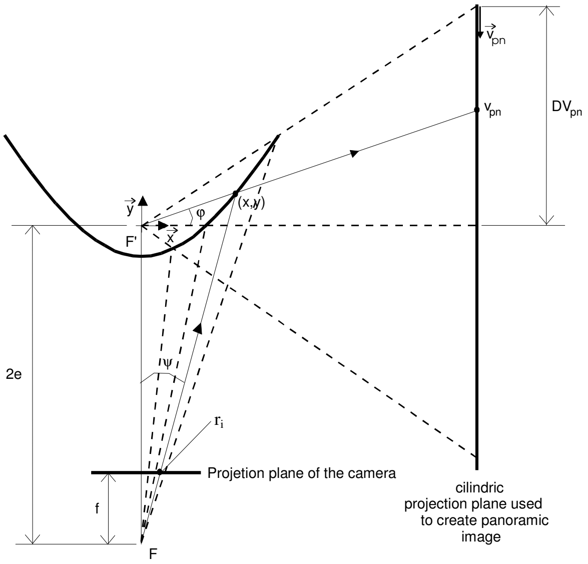

Figure 10.5: Panoramic projection [10].

10.2.3 Perspective Projections

The property of single projection centre allows to compute perspective projections free from distortions. The single projection centre is located at the focus of the hyperbola. One can define a projection plan that is perpendicular to a straight line passing through the focus of the mirror in order to map the pixels correctly. These images are the equivalent of images acquired by a perspective camera with a focus located at the same position of the focus of the hyperbola. The projection plan is defined by three parameters, f_p, θ_0 and ϕ_0 (see figure 10.6-a). The coordinate of a pixel $I(u_p, v_p)$ can be a function of the direction given by the angles θ and ϕ (figure 10.6-b). The following equations express the relationship between the angles and the geometry of the mirror [10]:

$$tan(\phi) = \frac{f_p sin(\phi_0 + v_p cos(phi_0))}{f_p cos(\phi_0)} \tag{10.3}$$

$$tan(\theta) = \frac{(f_p cos(\phi_0) - v_p sin(\phi_0))sin\theta_0 - u_p cos(\theta_0)}{(f_p cos(\phi_0) - v_p sin(\phi_0))cos\theta_0 + u_p sin(\theta_0)} \tag{10.4}$$

For a beam of light with angle ϕ, one can find a point (x, y) in the mirror's surface where it is reflected. Given (x, y), the pixel $i(u, v)$ in the camera (defined by the direction (θ, ϕ)) is:

$$u = \frac{x(2e + y_t)r_{pixel}}{xtan(\phi) + 2e} cos(\theta) \tag{10.5}$$

$$v = \frac{x(2e + y_t)r_{pixel}}{xtan(\phi) + 2e} sin(\theta) \tag{10.6}$$

It is possible to use look-up tables for fixed set of f_p, θ_0 and ϕ_0, and then compute perspective projections in real-time.

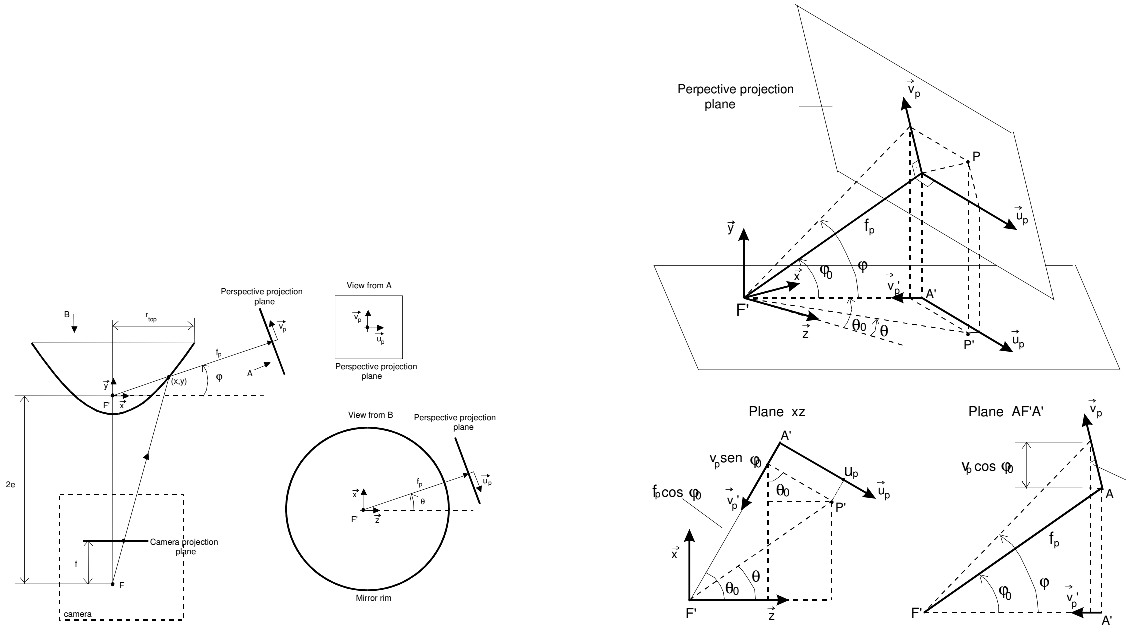

Figure 10.6: Perspective projection: a) Mirror-camera system b) Perspective projection plane [10].

10.2.4 Implementing Omnidirectional System Using Web-cams and a Customised Mirror

A web camera with a resolution of 640 x 480 was mounted with a customised hyperboloidal mirror. The system does not have a true single projection point, as the camera is not a real pinhole camera. However, for practical purposes, it is close enough.

The mirror was manufactured using a special aluminium alloy with a very homogeneous internal structure. The manufacturing process used an ultra-precision lathe with a diamond cutting tip, capable of turning within very narrow tolerances, of the order of tens of nanometres.

The lathe can only cut straight lines, so the curve is a only a good approximation of a hyperbola, formed by thousands of straight lines. This is not a problem due to the limited resolution of the camera. However, for higher resolution cameras the curve would require a better interpolation. Examples of images obtained by the system are shown in figures 10.7, 10.8, 10.9.

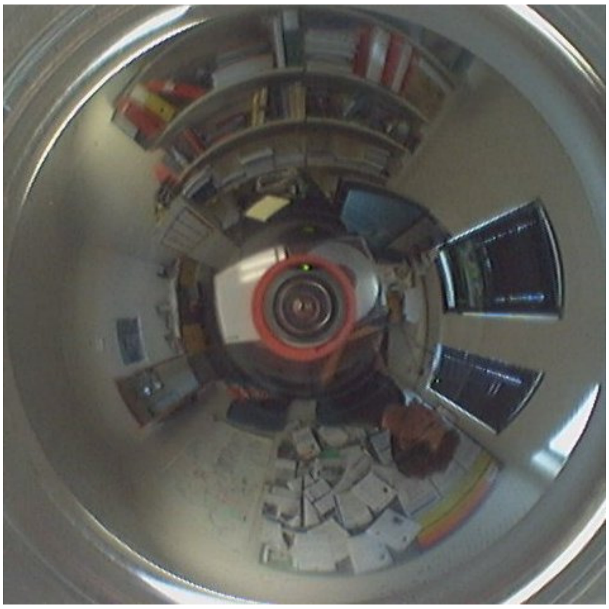

Figure 10.7: Omnidirectional image.

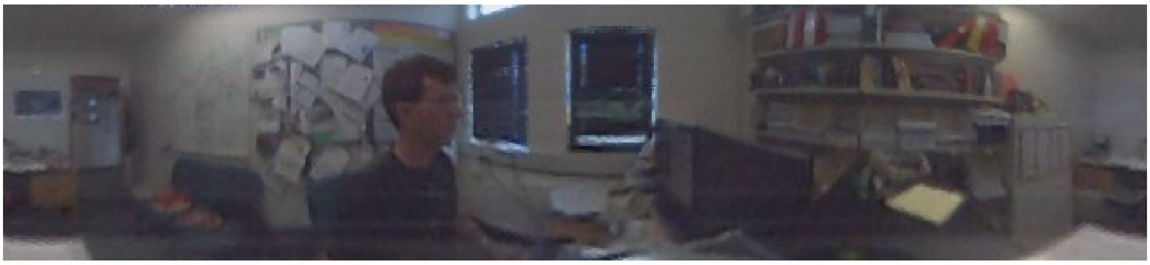

Figure 10.8: Panoramic projection.

10.3 Object Detection in Omnidirectional Systems: an example

In this omnidirectional system, we used Viola-Jones method to detect faces in images with panoramic projection or with perspective projection. We used the classifier provided by Lienhart et al.[13], slightly modified to fit our system. Preliminary tests showed that for the lighting conditions faced by the system a scaling factor between 1.1 to 1.5 resulted in a reasonable hit rate, while the false detections were adequate for the proposed approach. We used full resolution panoramic and perspective images, rather than the half-resolution images used in the original OpenCV implementation.

Face tracking in omnidirectional systems has been used by Wallhoff et al. [14]. They used Rowley's method to detect faces, which is slower than Viola-Jones method [15]. They did not address the problems of missing detections when the faces are located on one of the borders of the panoramic images. They used the PETS-ICVS 2003 database to gather results. We did

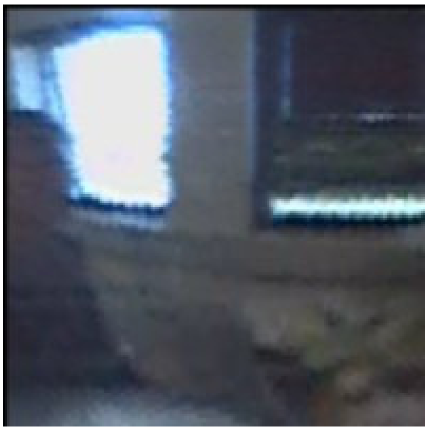

Figure 10.9: Perspective projection.

not use the PETS-ICVS database due to the differences in the acquisition of the images. In the PETS-ICVS images the camera is on top of the mirror, so the faces are located on the lower resolution part, in the centre of the images. In our system the mirror is located on top, so the high the resolution areas lie in the external part of the omnidirectional image.

Douxchamps and Campbel [16] used Viola-Jones with colour images (although this is irrelevant to the classifiers, as the classifiers only work with grayscale images) and higher resolution images (1000x1000 pixels). They do not give details about the mirror or the camera used in the system, focusing in showing that the system works with different illumination conditions.

10.3.1 Implementation of the Tracking Algorithm

It is relatively straightforward to use Viola-Jones method [15] with panoramic or perspective projections from omnidirectional images. However, there are specific issues that needed to be addressed. Objects in the image cannot be partitioned, otherwise the SATs cannot be used efficiently.

The panoramic projection can, in principle, be used directly for detection, as the deformations caused by the transformation of the images are small, or at least they do not affect the classifiers. However, if the face is to be found in one of the horizontal borders of the panoramic image, it is not going to be detected even though the face is continuous in the omnidirectional image. It is necessary to copy part of the panoramic image in order to make sure that the image of the objects, situated at any part of the field of vision, are continuous. Ideally, the portion to be copied should be at most half of the size of the largest subwindow, which is a function of the kernel size and the scaling factor used during the detection. Although this solves the problem of missing detection, it creates duplicated detections. An object at a smaller scale can be found in two different places in the extended panoramic projection. These duplications can be merged into a single coherent one based on their angular positions, as the angle would be the same for both detections (figure 10.10).

For the perspective projections, two solutions would be possible. The first would be to adopt a certain focal distance (zoom) and compute a perspective with the desired size for the scalable kernel to be used once. In other words, each perspective projection would make the same role of a sub-window. This would imply in computing several thousand projections in order to cover the whole omnidirectional image. Besides the problem of computing so many perspectives in real-time, it is also disadvantageous in the sense that the SATs cannot be computed in advance, making it impossible to compute the Haar-like features in real-time. It is, therefore,

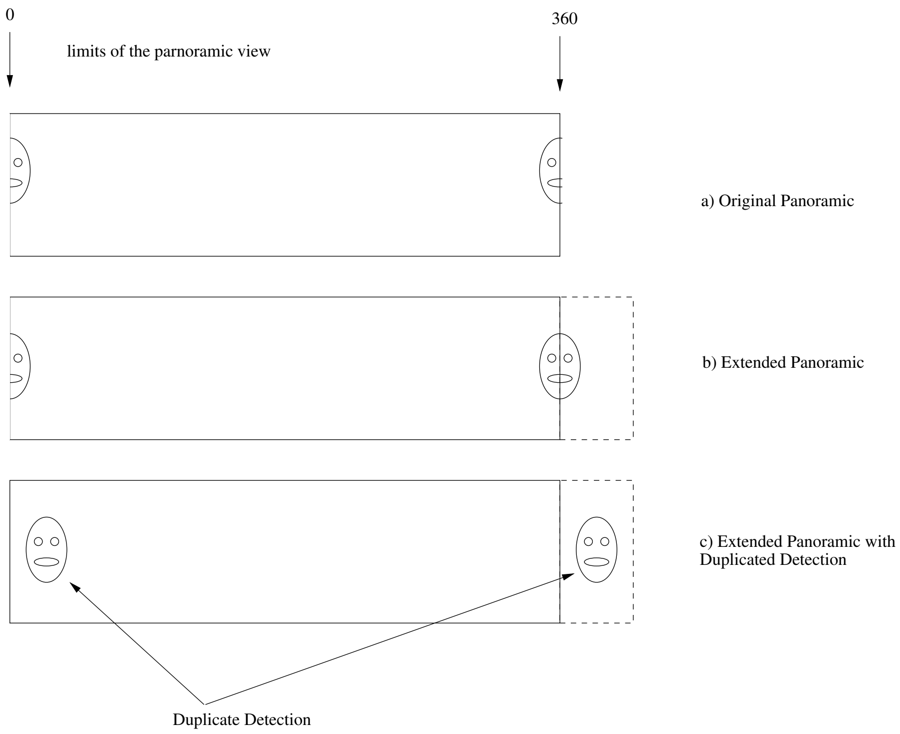

Figure 10.10: Extending the panoramic projection.

computationally too expensive to cover the entire area of view using only perspective projections.

The other possible approach would use fewer perspective projections, each projection covering a larger area. However, an object image can be located between projections, and to avoid missing objects, large areas of overlap would have to be used. It is more convenient that once ROIs have been located using the panoramic, then one or more perspective images are used only for tracking.

10.3.2 Detecting and Tracking Faces

In the system developed for this example, one extended panoramic per frame is computed. The face classifier detects all the faces in the panoramic, and a data structure of the existing faces keeps the positions and one parameter to store a number of frames to live (f_{tl}). If the face continue to be detected in the vicinity of the previous detection, then the system adds to f_{tl} up to a top limit. If the face is not detected temporarily, then one is subtract from f_{tl}. If the f_{tl} of a certain face is smaller than a minimum m (typically in our system $m = 3$), then the entry for that face is deleted. In parallel, for every existing detection a perspective is computed according to its size and position. The position and existence of all faces at a certain time is confirmed running the classifier against each perspective projection. The perspective image of each face is shown in a separate window. In essence, we use the panoramic image for initial detection and perspectives to track the faces and keep them in storage. When the person leaves the area, 1 is subtracted from f_{tl} until the entry is deleted and the correspondent window is destroyed. A simplified task flow can be seen in figure 10.11. The idea is to keep the number of perspective projections at a minimum, so the system can deliver a reasonable frame rate. In the next section we show the results of this approach.

The fps rates were measured as a function of the number of projections computed. Initially, the system is able to deliver around 30 fps without any projection or classification computations.

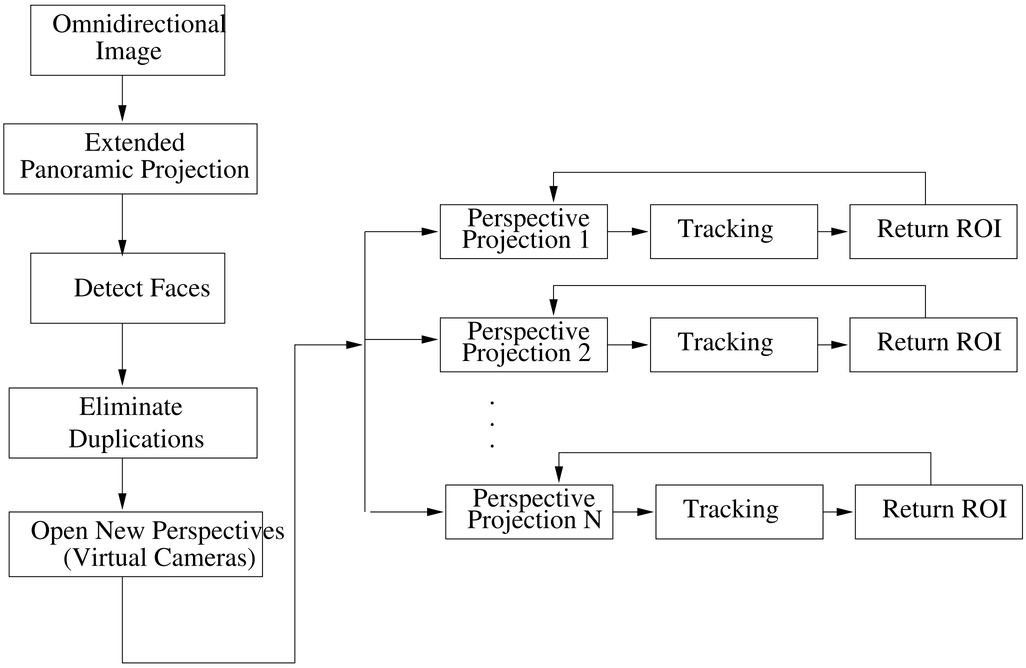

Figure 10.11: Task flow.

When computing a single panoramic image per frame without running the detection process, the system is able to achieve about 25 fps. The rate drops rapidly as we compute only a few perspective images. With 200 perspective projections per frame the rate drops to less than 2 fps, showing that it is unfeasible to try to compute enough perspective projections to sweep the entire omnidirectional image. The size for the perspective projections used in the experiments was the same as the size of the classifier's initial kernel of 20x20 pixels.

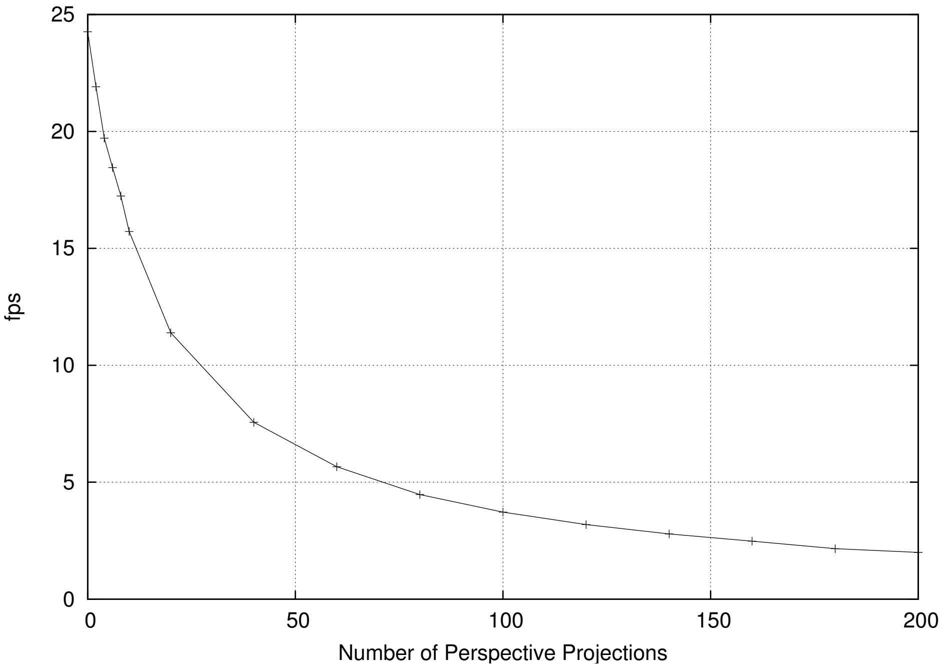

Figure 10.12: Frame rates per n perspective projections.

We tested the system with several people entering and exiting the view of the camera. Figure 10.13 shows typical frame rates. The figure shows that the rate tends to drop slightly when more faces are detected. This is due to the extra effort needed for the perspective projections and for the classification.

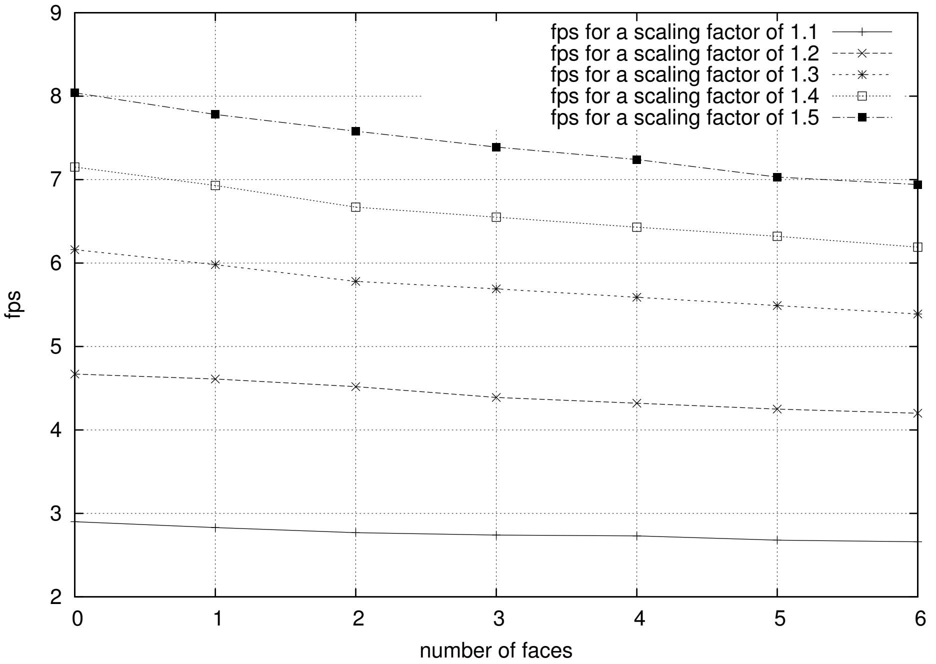

Figure 10.13: Typical frame rates.

Figure 10.14 shows the variation of the frame rates as a function of the scaling factor s. For scaling factor beyond 1.4 the system is no longer able to detect faces as accurately, missing many potential detections.

An example of an image where several faces are tracked is shown: the extended panoramic with a circle showing the detection in figure 10.15 and the perspective projections for each face based on the ROI of the first detection (figure 10.16.

10.3.3 Hit rates

The hit rates for Viola-Jones method is known to be sensitive to the scaling factor [17]. The characteristic ROC curve (hit rates against false detection) is standard measurement tool used for classifiers. Viola and Jones [15] as well as Lienhardt and Maydt [17] varied the number of layers when running classifiers in order to plot ROC curves. For the scope of this paper, we plotted hit rates against the scaling factor to demonstrate that a trade-off needs to be chosen in terms of performance and accuracy. Figure 10.17 shows hit rates based on a sequence of images acquired by the omnidirectional system. A total of 472 frontal faces can be found in the sequence. For scaling factors 1.1 and 1.2, the detection is sufficient to keep track of the face during the whole sequence. For the other scaling factors, the algorithm may loose the track of some faces temporarily.

10.3.4 Limitations of the current system

In practise, due to the area covered and the resolution of the camera and set number of perspective projections, the system was limited to track a maximum of 10 faces. We tested it with

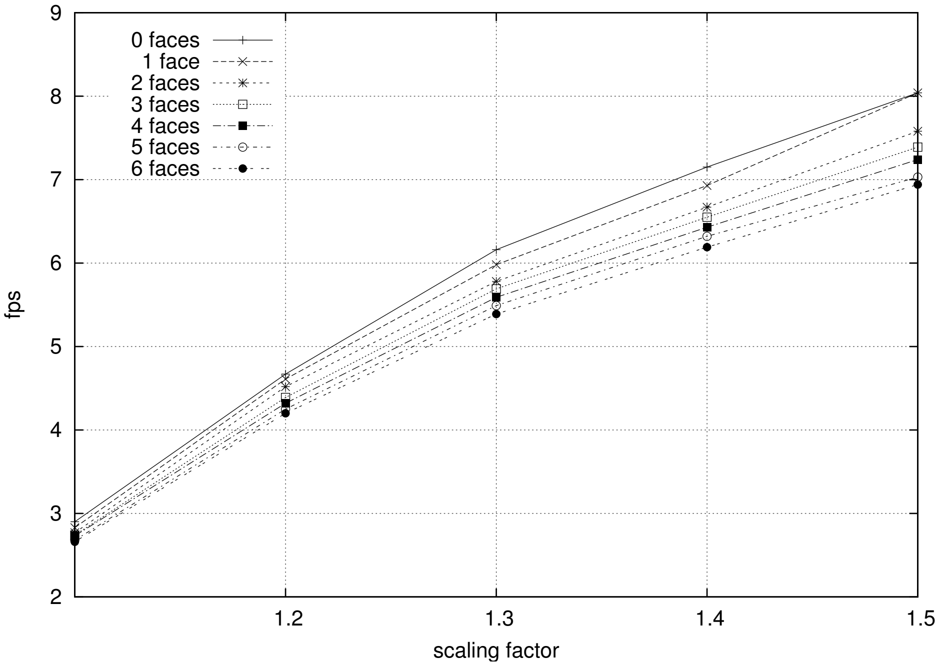

Figure 10.14: Frame rates for different scaling factors.

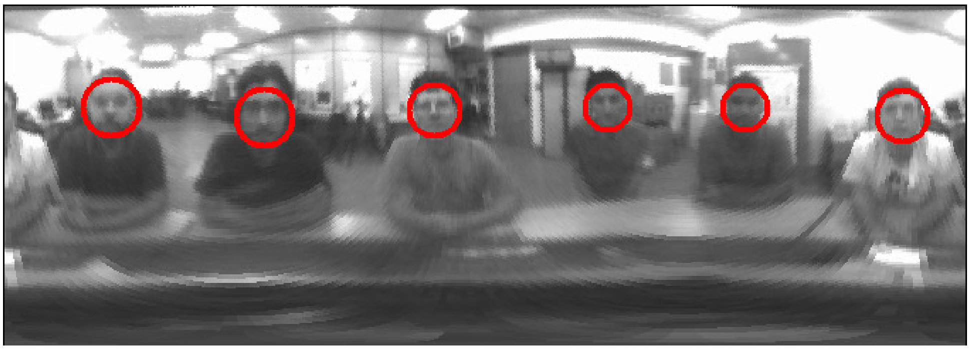

Figure 10.15: Example of an extended panoramic projection. The position of the detected faces are shown with a circle

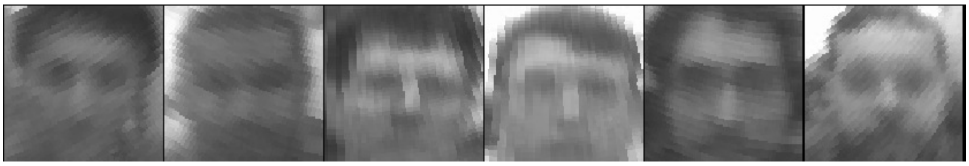

Figure 10.16: Example of various perspective images for the faces detected in figure 10.15.

7 faces in real-time. The classifier used was trained with frontal faces only. If the person turns, the system keeps track of the face for a limited time and it deletes the face after 5 frames. The tracking process restarts as soon as the person looks directly to the camera again. There are different alternatives to solve this problem, including the use of multiple classifiers. All these solutions impose more effort to the computation, so there is a trade-off to what can be achieved.

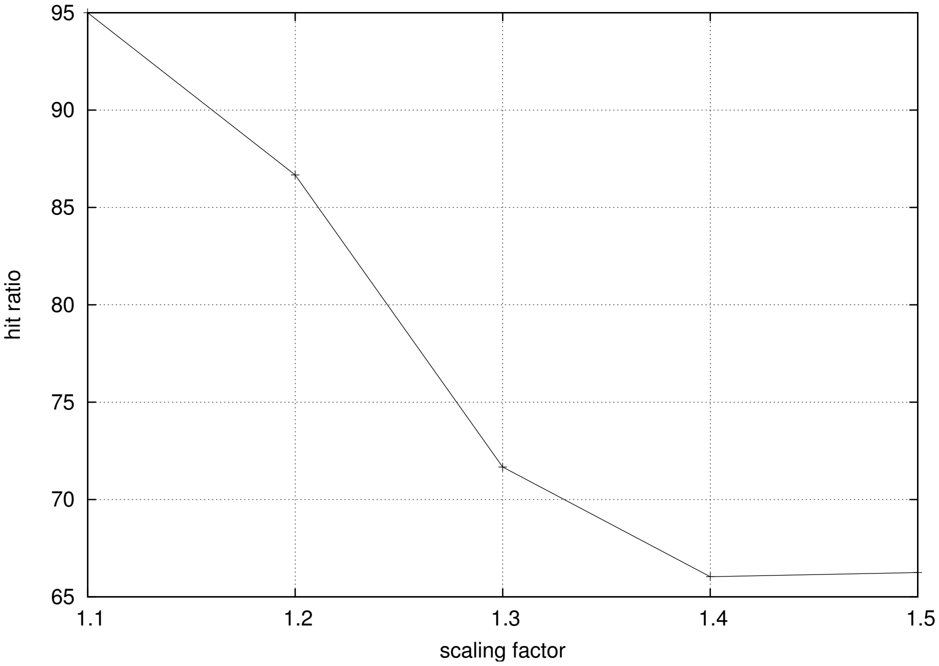

Figure 10.17: Hit rates (%) for different scaling factors.

10.4 Exercises

10.4.1 Exercise 1

1. Download the code for the hyperboloidal catadioptric system. Compile and run it with static images (downloadable from the same URL).

2. Move the mouse on the panoramic image in order to get new perspective images. Note that for every new perspective position, a new LUT needs to be computed.

3. Click left or right buttons to zoom in and out.

Bibliography

[1] R. A. Hicks and R. Bajcsy, “Reflective surfaces as computational sensors,” in *Second Workshop on Perception for Mobile Agents*, (Fort Collins, Colorado), pp. 82–86, 1999.

[2] “Google street view,” 2008.

[3] S. Baker and S. K. Nayar, “Single viewpoint catadioptric cameras,” in *PV01*, pp. 39–71, 2001.

[4] S. Baker and S. K. Nayar, “A theory of catadioptric image formation,” in *ICCV98*, (Bombay, India), pp. 35–42, 1998.

[5] S. Baker and S. K. Nayar, “A theory of single-viewpoint catadioptric image formation,” *International Journal of Computer Vision*, vol. 35, pp. 175–196, November 1999.

[6] R. Pless, “Using many cameras as one,” in *CVPR03*, vol. II, pp. 587–593, June 2003.

[7] S. Derrien and K. Konolige, “Approximating a single viewpoint in panoramic imaging devices,” in *IEEE Workshop on Omnidirectional Vision 2000*, pp. 85 – 90, 2000.

[8] J. Gaspar, C. Decco, J. Okamoto, J., and J. Santos-Victor, “Constant resolution omnidirectional cameras,” in *Third Workshop on Omnidirectional Vision*, pp. 27 – 34, June 2002.

[9] T. Svoboda, T. Pajdla, and V. Hlaváč, “Central panoramic cameras: Geometry and design,” Tech. Rep. Research report K335/97/147, Czech Technical University, Faculty of Electrical Engineering, Center for Machine Perception, December 1997. Available at ftp://cmp.felk.cvut.cz/pub/cmp/articles/svoboda/TR-K33597-147.ps.gz.

[10] V. Grassi Jr. and J. Okamoto Jr., “Development of an omnidirectional vision system,” *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. XXVIII, no. 1, pp. 56–68, 2006.

[11] V. Grassi Jr., C. Decco, J. Okamoto Jr., and A. Porto, “Development of an omnidirectional vision system,” in *XVI COBEM2001 Robotics and Control*, pp. 28–37, 2001.

[12] T. Svoboda and T.Pajdla, “Epipolar geometry for central catadioptric cameras,” *International Journal of Computer Vision*, vol. 49, pp. 23–37, Jan 2002.

[13] R. Lienhart, L. Liang, and A. Kuranov, “A detector tree of boosted classifiers for real-time object detection and tracking,” in *ICME2003*, pp. 277–280, IEEE, 2003.

[14] F. Wallhoff, M. Zobl, G. Rigoll, and I. Potucek, “Face tracking in meeting room scenarios using omnidirectional views,” in *17th International Conference on Pattern Recognition (ICPR’04)*, vol. 4, pp. 933–936, 2004.

[15] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, May 2004.

[16] D. Douxchamps and N. Campbell, "Robust real-time face tracking for the analysis of human behaviour," in *4th Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI 2007)*, vol. LNCS 4892, (Brno, Czech Republic), pp. 1–10, Jun 2007.

[17] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *ICIP02*, (Rochester, NY), pp. I: 900–903, September 2002.