Res. Lett. Inf. Math. Sci., 2009, Vol. **13**, pp. 55–67
Available online at http://iims.massey.ac.nz/research/letters/

1

Face Tracking Using a Hyperbolic Catadioptric Omnidirectional System

A. L. C. Barczak[1], J. Okamoto Jr.[2] and V. Grassi Jr.[3]
[1] *IIMS, Massey University, Auckland, New Zealand*
*[2] *Dept. of Mechatronics, Escola Politécnica, University of São Paulo, Brazil*
**[3] *Dept. of Electric Engineering, São Carlos, University of São Paulo, Brazil*

In the first part of this paper, we present a brief review on catadioptric omnidirectional systems. The special case of the hyperbolic omnidirectional system is analysed in depth. The literature shows that a hyperboloidal mirror has two clear advantages over alternative geometries. Firstly, a hyperboloidal mirror has a single projection centre [1]. Secondly, the image resolution is uniformly distributed along the mirror's radius [2].

In the second part of this paper we show empirical results for the detection and tracking of faces from the omnidirectional images using Viola-Jones method. Both panoramic and perspective projections, extracted from the omnidirectional image, were used for that purpose. The omnidirectional image size was 480x480 pixels, in greyscale. The tracking method used regions of interest (ROIs) set as the result of the detections of faces from a panoramic projection of the image. In order to avoid losing or duplicating detections, the panoramic projection was extended horizontally. Duplications were eliminated based on the ROIs established by previous detections. After a confirmed detection, faces were tracked from perspective projections (which are called *virtual cameras*), each one associated with a particular face. The zoom, pan and tilt of each virtual camera was determined by the ROIs previously computed on the panoramic image.

The results show that, when using a careful combination of the two projections, good frame rates can be achieved in the task of tracking faces reliably.

Keywords: Omnidirectional cameras; Catadioptric Systems; Object Detection; Face Tracking;

1 Introduction

True omnidirectional cameras should be able to obtain views from all directions. Some authors make a distinction between omnidirectional and panoramic sensors (for example [3]), but it is widely accepted that the term omnidirectional is applied to any system that has a 360° field of view on the horizontal, 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, but there is now a renewed interest in computer vision, robotics, surveillance and in other applications where multiple users share one camera. Recently google maps used omnidirectional cameras to store images for their Street View technology [4]. Omnidirectional systems have been used to help with robot navigation [5] and for robot soccer [6].

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 1 shows some of the types of omnidirectional systems encountered in the literature and it

[1] Author for Correspondence: a.l.barczak@massey.ac.nz

is an attempt to classify them, although there are systems that do not fit perfectly in a single category.

The special case of the catadioptric system, formed by a single camera associated with a single mirror, is the focus of this review. 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 1). Several curves for the mirror were evaluated in the past. A good survey of different mirror geometries are described in [1], [7] and [8].

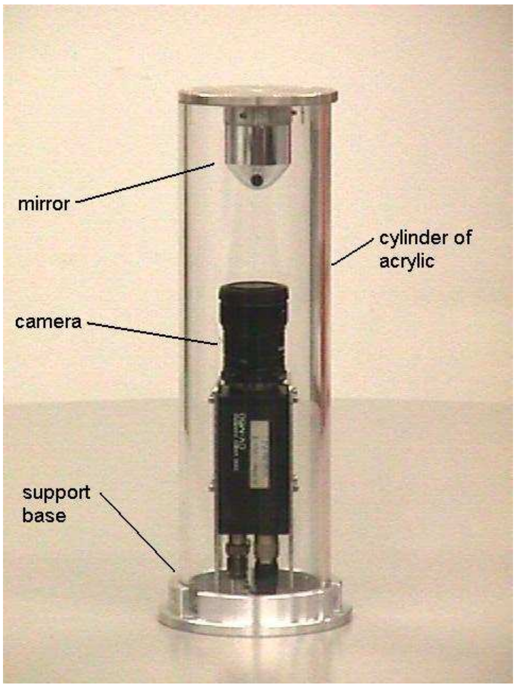

Figure 1: A catadioptric system with a hyperboloidal mirror, used to collect data for this work.

Table 1: Omnidirectional Systems

Cameras	**Type**	**Sub-type**	**Image geometry**	**Typical Resolution**	**Pre-processing**
multiple cameras	aligned	joined CCDs	cylindrical	high	stitch
	non-aligned		spherical	high	stitch/affine
single camera	dioptric	wide view lens	deformed	low/medium	projection
		rotating ([9])	cylindrical	high	stitch/unblur
	catadioptric	planar spherical conical paraboloidal hyperboloidal ellipsoidal	deformed	low/medium	projection

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 perspective images with minimum distortion [1]. Secondly, hyperbolic mirrors have a good balance in terms of the resolution at different points of the mirror [2].

Spherical mirrors are widely available for other purposes, as they are relatively easy to manufacture. However, this shape does not have a single projection centre. [10] have proposed an approximation and successfully built a low resolution catadioptric system using a single sphere with the camera embedded on it.

Face tracking in omnidirectional systems has been used by Wallhoff et al. [11]. They used Rowley's method to detect faces, which is slower than Viola-Jones method [12]. 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 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 [13] 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.

The structure of this paper is the following: in the section 2 we describe how the projections are computed in our system, and explain the concept of virtual cameras. In section 3 we describe how our method can achieve real-time omnidirectional tracking of multiple faces. In section 4 we show the results, focusing on the performance of the system. In section 5 we summarise the paper and present ideas for future work.

2 Virtual Cameras in Omnidirectional Vision Systems

One advantage of the catadioptric systems is that they do not need to stitch images from different cameras. However, the image is usually very deformed (see figure 10), and therefore one needs to compute the appropriate projection before using the image for further processing. The mirror's geometry affects two important properties of the system: resolution and focusing [7].

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.

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, [2] studied the conditions for constant resolution on omnidirectional cameras.

2.1 Implementation of Hyperbolic Systems

Svoboda [14] implemented a catadioptric system using a perspective camera associated with a hyperboloidal mirror. Grassi and Okamoto [15, 16] 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 [15] the mirror is located on top of the camera. In this way the resolution is distributed conveniently for a mobile robot. Figure 1 shows the acquisition system.

Based on [17] and [7], Grassi and Okamoto [15] 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. The two basic types of projections, panoramic and perspective, are shown in figure 2.

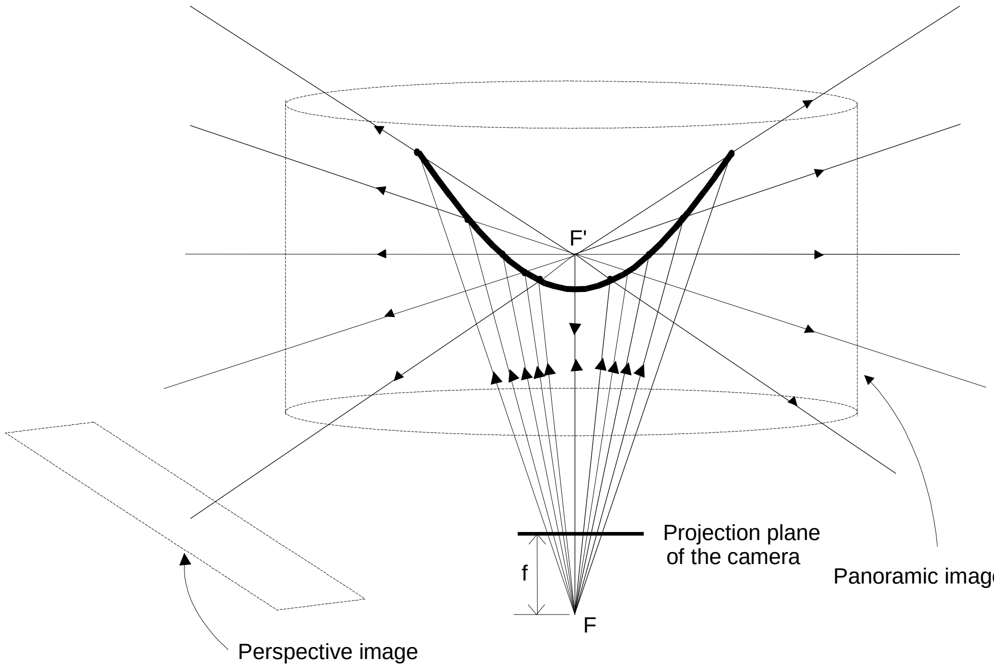

Figure 2: Perspective and panoramic projections [15].

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 [15], one can map the positions of pixels from the omnidirectional image to a panoramic image (see figure 3):

$$u = r_p \, cos\left(\frac{2\pi u_{pn}}{H_{pn}}\right) \tag{1}$$

and

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

The positions are stored in a look-up table, so the computation can be done in real-time.

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. The images are the equivalent of images acquired by a

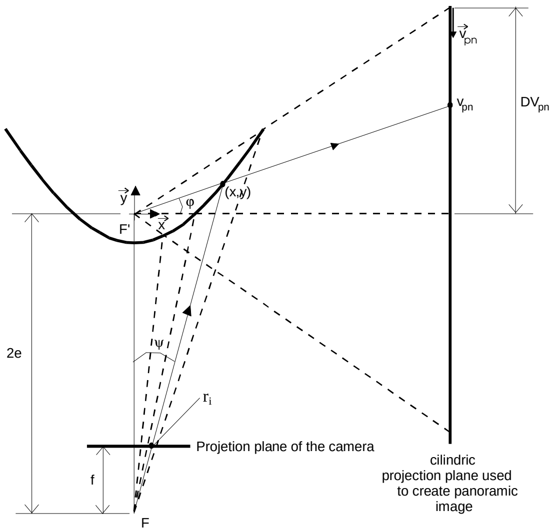

Figure 3: Panoramic projection: [15].

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 4-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 4-b). The following equations express the relationship between the angles and the geometry of the mirror [15]:

$$tan(\phi) = \frac{f_p sin(\phi_0 + v_p cos(phi_0))}{f_p cos(\phi_0)} \tag{3}$$

$$tan(\theta) = \frac{(f_p cos(\phi_0) - v_p sin(\phi_0))sin\theta_0 - u_p cos(\theta_0)}{(f_p cos(\phi_0) - v_p sin(\phi_0))cos\theta_0 + u_p sin(\theta_0)} \tag{4}$$

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:

$$u = \frac{x(2e + y_t)r_{pixel}}{xtan(\phi) + 2e} cos(\theta) \tag{5}$$

$$v = \frac{x(2e + y_t)r_{pixel}}{xtan(\phi) + 2e} sin(\theta) \tag{6}$$

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

3 Implementation of the Tracking Algorithm

In our system we used Viola-Jones to detect faces in images with panoramic projection or with perspective projection. We used the classifier provided by Lienhart et al. [18], slightly modified to fit in 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.

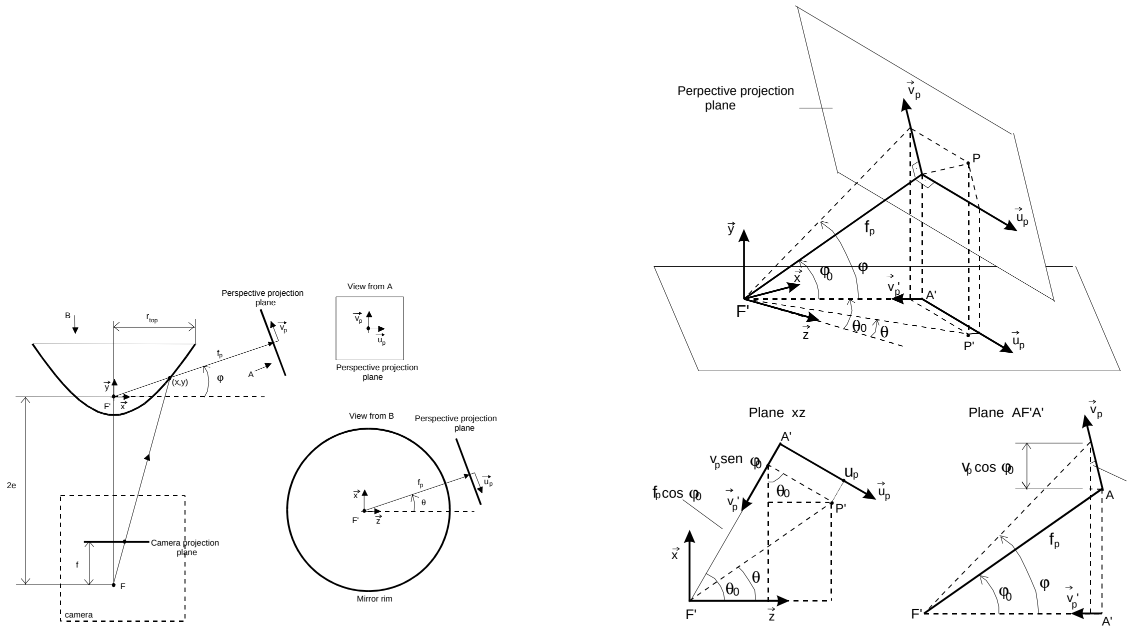

Figure 4: Perspective projection: a) Mirror-camera system b) Perspective projection plane [15].

It is relatively straightforward to use Viola-Jones method [12] 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. Using the hyperboloidal mirror there was no need to calibrate the system in order to detect faces using the OpenCV 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 5).

For the perspective projections, two approaches 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, 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 only possible to compute faces in real-time from perspective projections if ROIs have been located using the panoramic projection.

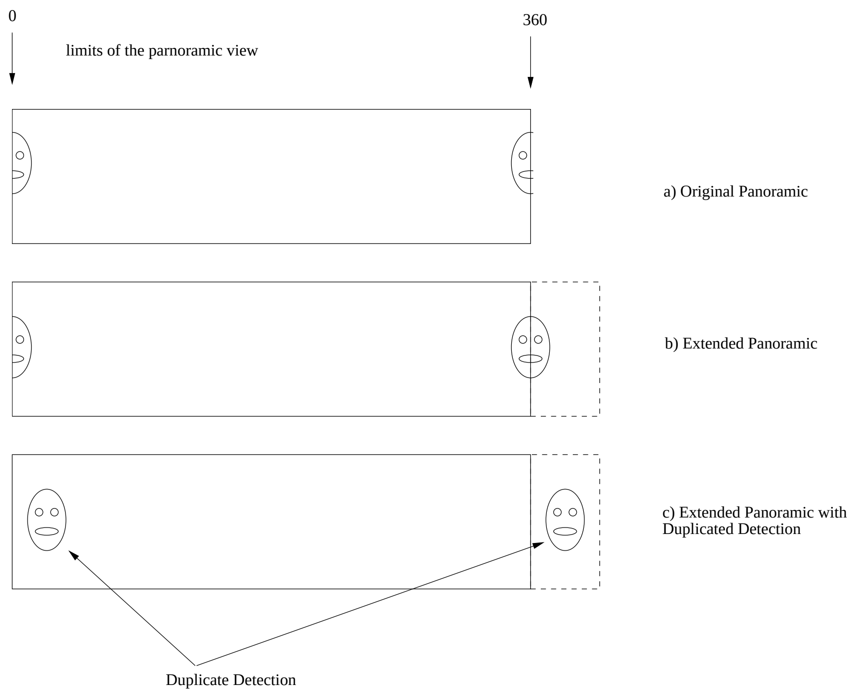

Figure 5: Extending the panoramic projection.

3.1 Detecting and Tracking Faces

In the system developed for this work, 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 image 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 perspective images 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 6. 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.

4 Experimental Results

The experiments were carried out using our omnidirectional vision system equipped with a camera with a resolution of 640x480 pixels (see figure 1), which delivers greyscale images only. The images acquired were sent via a wireless network and received via CORBA in the processing computer. The processing computer was a Intel(R) Pentium(R) 4 CPU 3.00GHz, with a cache size 1024 KB and 6002.23 bogomips. The OS was Linux (kernel 2.6.17). The detection parameters were: scaling factors between 1.1 and 1.5, and a kernel size of 20x20 pixels.

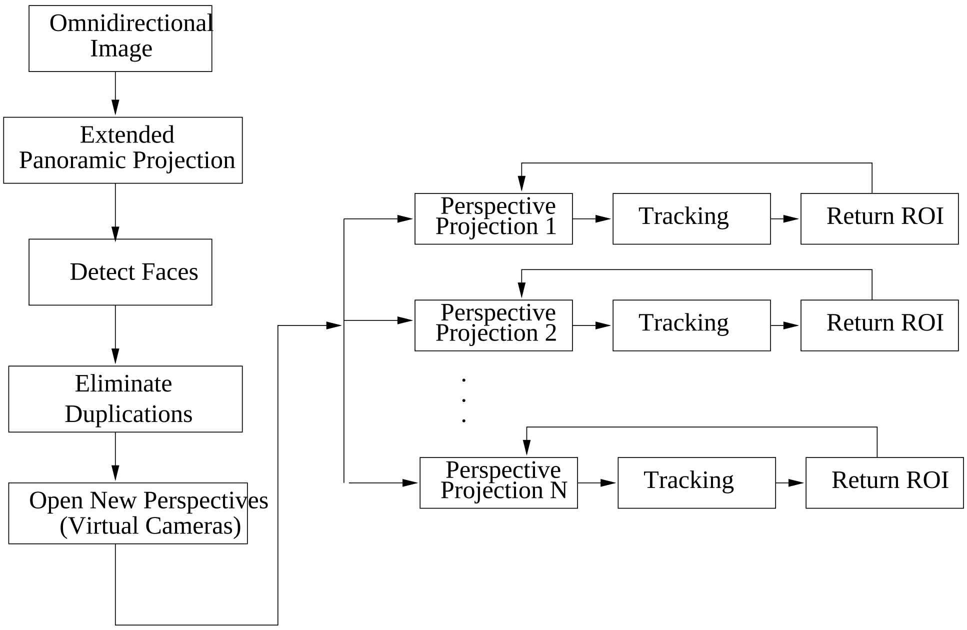

Figure 6: Task flow.

4.1 Perspective Projections

As figure 7 shows, it is not possible to compute too many perspectives in real-time due to the extra computational effort. To show the effects on the system, we computed several perspective projections real-time and measured the fps rates as a function of the number of projections computed. Initially, the system was able to deliver around 30 fps without any projection or classification computations. When computing a single panoramic image per frame without running the detection process, the system was able to achieve 25 fps. The rate dropped rapidly as we compute perspective images. With 200 perspective projections per frame the rate dropped to less than 2 fps, showing that it is unfeasible to try to compute enough perspective projections to sweep the entire omnidirectional image. The initial size for the perspective projections used in this experiment was the same as the size of the classifier's initial kernel, 20x20 pixels.

4.2 Typical Frame Rates

We tested the system with several people entering and exiting the view of the camera. Figure 8 shows typical frame rates. The figure shows that the rate tends to drop slightly when more faces are detected. This was due to the extra effort needed for the perspective projections and for the classification.

Figure 9 shows the variation of the frame rates as a function of the scaling factor s. For scaling factor beyond 1.4 the system was no longer able to detect faces as accurately, missing many potential detections.

An example of an image where several faces are tracked in real-time is shown in figures 10, 11 and 12.

4.3 Hit rates

The hit rates for Viola-Jones method is known to be sensitive to the scaling factor [19]. The characteristic ROC curve (hit rates against false detections) is standard measurement tool used for

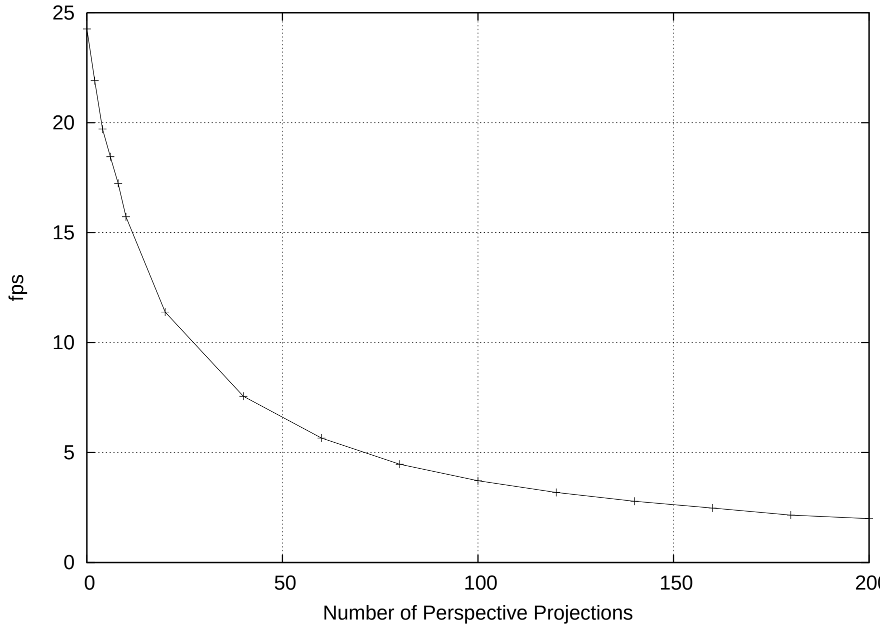

Figure 7: Frame rates per n perspective projections.

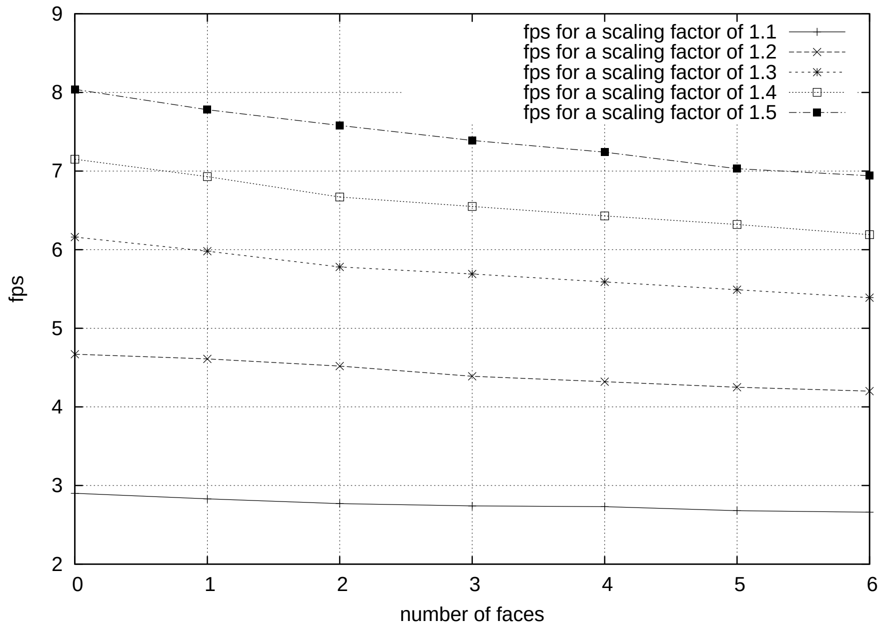

Figure 8: Typical frame rates.

classifiers. Viola and Jones [12] as well as Lienhardt and Maydt [19] varied the number of layers when running classifiers in order to plot ROC curves.

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 13 shows hit rates based on a sequence of test images acquired by the same omnidirectional system. A total of 472 frontal faces can be found

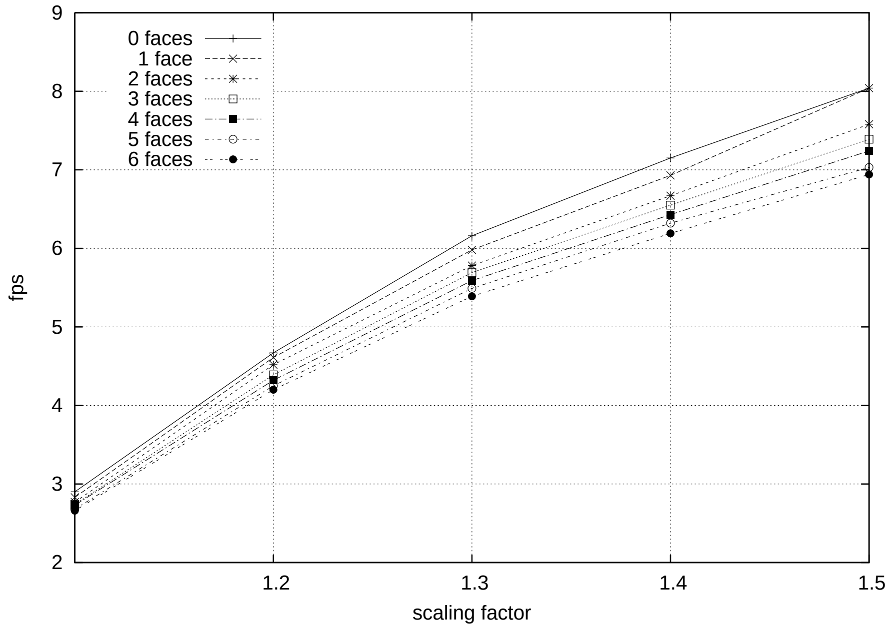

Figure 9: Frame rates for different scaling factors.

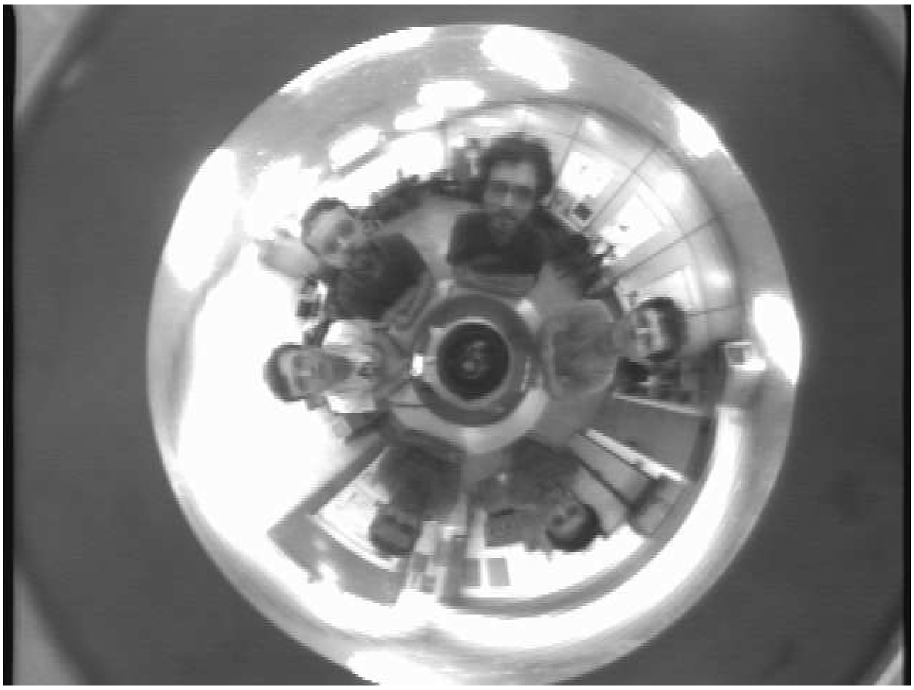

Figure 10: Example of an omnidirectional image with 6 faces.

in the test sequence. For scaling factors 1.1 and 1.2, the detection was 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.

Figure 11: Example of an extended panoramic projection of figure 10. The position of the detected faces are shown with a circle.

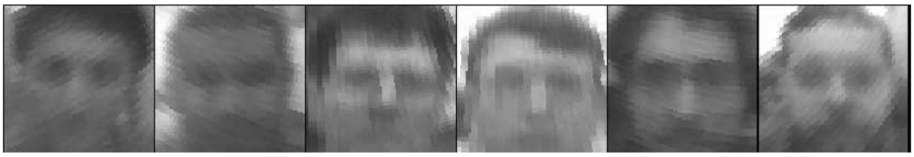

Figure 12: Example of various perspective images for the faces detected in figure 11.

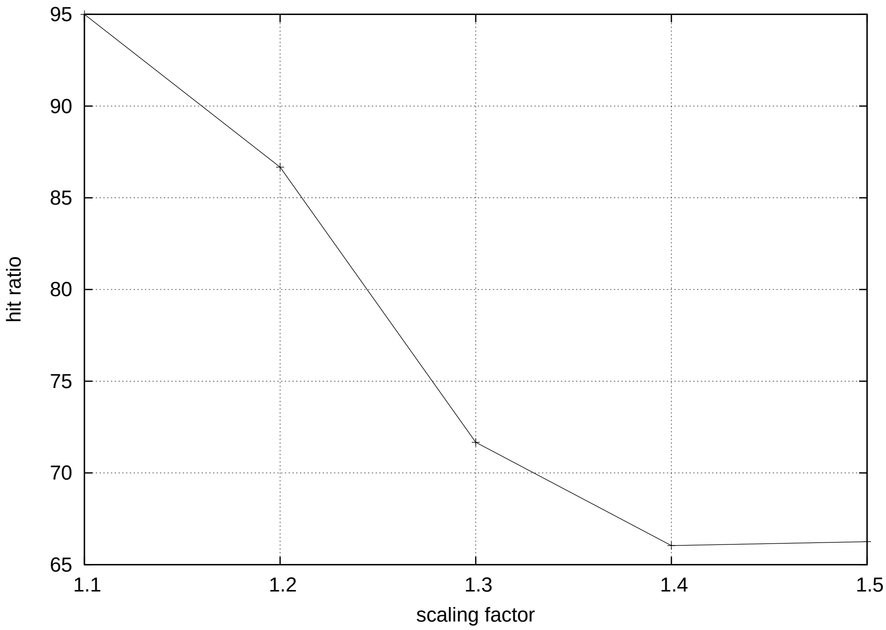

Figure 13: Hit rates (%) for different scaling factors.

4.4 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 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 3 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 (e.g., [20]). All these solutions impose more computational effort, so there is a trade-off to what can be achieved.

5 Conclusions

In this paper an omnidirectional system was used to track multiple faces. Systems with single projection centre are more suitable for applications such as object detection because it is possible to get projections free from distortions. Panoramic projections can be used directly by detection algorithms, although some additional steps described in the article have to be carried out. Perspective projections can also be used directly by the detection algorithm. However, due to the extra effort to compute and cover all areas in a way that objects are not missed, it is more efficient to establish a ROI before detection. This can be done using the panoramic projection. It is straightforward to detect faces from the perspective projections if adequate ROIs are computed. In order to keep the system running real-time, the number of perspective projection need to be kept at a minimum, which in our system was around 7.

Acknowledgements

The authors would like to thank Fabiano R. Correa of the Laboratory of Advanced Perception (LPA) for the interface for the remote image acquisition written in CORBA. This work was partially supported by FAPESP (grant 03/08132-1), by CnPq (grant 472727/2003-6) and a Sabbatical grant from Massey University (2008).

References

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

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

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

[4] Google, 2008. [Online]. Available: http://en.wikipedia.org/wiki/Google_Street_View

[5] V. C. Guizilini and J. Okamoto, "Solving the online slam problem with an omnidirectional vision system," in *ICONIP (1)*, ser. Lecture Notes in Computer Science, M. Köppen, N. K. Kasabov, and G. G. Coghill, Eds., vol. 5506. Auckland, New Zealand: Springer, 2009, pp. 1110–1117.

[6] J. Silva, N. Lau, A. J. R. Neves, J. Rodrigues, and J. L. Azevedo, "Obstacle detection, identification and sharing on a robotic soccer team," in *EPIA*, ser. Lecture Notes in Computer Science, L. S. Lopes, N. Lau, P. Mariano, and L. M. Rocha, Eds., vol. 5816, Aveiro, Portugal, October 2009, pp. 350–360.

[7] S. Baker and S. K. Nayar, "A theory of catadioptric image formation," in *ICCV98*, Bombay, India, 1998, pp. 35–42.

[8] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 175–196, November 1999.

[9] R. Pless, "Using many cameras as one," in *CVPR03*, vol. II, June 2003, p. 587593.

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

[11] 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, 2004, pp. 933–936.

[12] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[13] 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, Jun 2007, pp. 1–10.

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

[15] 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.

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

[17] T. Svoboda and T.Pajdla, "Epipolar geometry for central catadioptric cameras," *International Journal of Computer Vision*, vol. 49, no. 1, pp. 23–37, Jan 2002.

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

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

[20] A. Barczak, C. Messom, and R. Chemudugunta, "Real-time rotationally invariant features for environmental feature detection by mobile robots and sensor networks," in *ROSE2007*, Ottawa, Canada, 2007.