Fusion of Omnidirectional and PTZ Cameras for Face Detection and Tracking

Amine Iraqui, Yohan Dupuis, Rémi Boutteau, Jean-Yves Ertaud, Xavier Savatier

Fusion of Omnidirectional and PTZ cameras for face detection & tracking

Amine Iraqui.H, Yohan Dupuis, Rémi Boutteau, Jean-Yves Ertaud and Xavier Savatier
Institut de Recherche en Systèmes Electroniques Embarqués (IRSEEM, France)

Abstract

Many applications for mobile robot authentication require to be able to explore a large field of view with high resolution. The proposed vision system is composed of a catadioptric sensor for full range monitoring and a pan tilt zoom (PTZ) camera leading to an innovative sensor, able to detect and track any moving objects at a higher zoom level. In our application, the catadioptric sensor is calibrated and used to detect and track regions of interest (ROIs) within its 360 degree field of view (FOV), especially face regions. Using a joint calibration strategy, the PTZ camera parameters are automatically adjusted by the system in order to detect and track the face ROI within a higher resolution.

1. Introduction

With rapidly growing demands in monitoring mobile robot applications for nomad biometric authentication, substantial developments have been released with multiple-camera sensor systems. One of the intelligent examples is the use of an omnidirectional camera in conjunction with pan tilt zoom (PTZ) camera, referred to as a dual camera system. Omnidirectional cameras are able to explore a wide field of view (FOV) within its 360 degrees full range monitoring. However, due to a low and non-uniform resolution, these catadioptric sensors are unable to provide close observations of particular targets, especially in biometric authentication applications. This is where PTZ cameras fit in. With high mobility and zoom ability, PTZ cameras compensate the deficiencies of omnidirectional cameras.

Based on a unified model projection introduced by Geyer [1], the catadioptric sensor is calibrated, in order to generate correct perspective images. By using Viola & Jones algorithm [6] in the resulting images, the program detects the face ROI, and then applies the tracking algorithm based on a correlation approach. Finally, a joint calibration method is performed to localize the face ROI in order to generate a zoomed in face image with high resolution.

This paper is organized as follows. Section 2 describes the vision system proposed. Section 3 is dedicated to the modeling and the calibration catadioptric sensor. Section 4 focuses on the proposed face detection and tracking algorithms. The fusion procedure of the omnidirectional and PTZ cameras is described in section 5. Section 6 illustrates our experiment results, and section 7 concludes this paper.

2. An overview of the system

The set-up of the dual camera system is shown in Fig.1. The proposed vision system, dedicated to mobile robot applications for biometric authentication, fuses catadioptric sensor with a dome PTZ camera.

Fig.1. Prototype of the vision system

Among all configurations of central catadioptric sensors existing in the literature, the combination of a hyperbolic mirror and a camera is preferable for the sake of compactness since a parabolic mirror needs a bulky telecentric lens. The hyperbolic mirror assures only one point of projection in order to create a perspective image. The chosen reflector is directly placed over the camera CCD sensor (1.3 Megapixel) and its field of view extends from 16degrees over the horizon to 90 degrees below. By using such a sensor, the obtained resolution image is non-uniform and low.

To obtain images with a better resolution, we use a Dome Axis Network PTZ camera placed in parallel to the azimuth axis of the omnidirectional camera. This camera offers wide zoom ability (optical zoom 35x), a pan range of 360°, a tilt range of 180°, a high mobility and speed movement until 450°/second.

3. The catadioptric sensor

3.1. Fixed view point constraint

The catadioptric sensor architecture adheres to the Single-View-Point theory [2]. The SVP constraint enables to generate correct perspective images. In fact, the optical center of the camera has to coincide with the second focus **F'** of the hyperbola located at distance 2e from the mirror focus as illustrated in Fig.2. The eccentricity *e* is a parameter of the mirror given by the manufacturer.

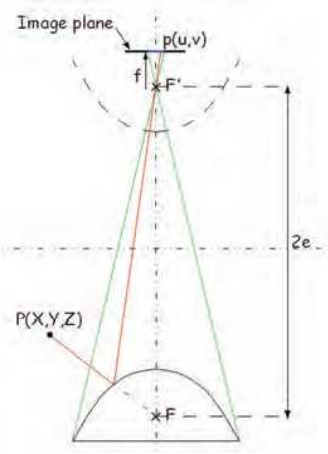


Fig.2. Image formation with a hyperbolic mirror.

To realize this task, we first calibrate our camera with a standard calibration tool to determine the central point and the focal length. Knowing the parameters of both the mirror and the camera, the image of the mirror on the image plan can be easily predicted if the SVP constraint is taken into consideration, as illustrated in Fig.2. The expected mirror boundaries are superposed on the image and the mirror has then to be moved manually to fit this estimation as shown in Fig.3.

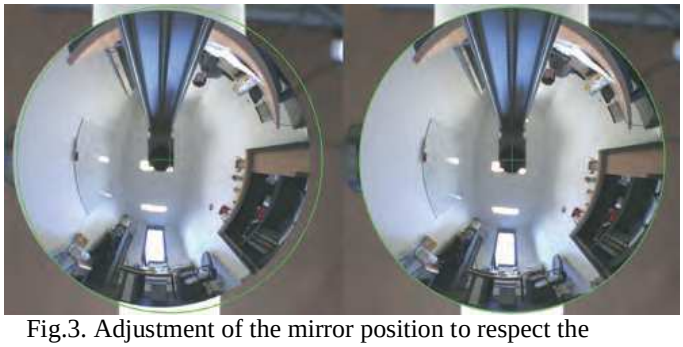


Fig.3. Adjustment of the mirror position to respect the SVP constraint. The mirror border has to fit the estimation (green circle).

3.2. Sensor calibration

The sensor calibration is a necessary step to establish the relationship between the 3D points of the scene and their projection into the image (pixel coordinates). Although there many calibration methods, they can be classified in two main categories: parametric and non-parametric methods. The first category consists in finding an appropriate model for the projection of 3D point onto the image plane. Non-parametric approaches associate one projection ray to each pixel [11] and provide a "black box model" of the sensor. They are well adapted for general purposes but minimization algorithms (gradient descent, Gauss-Newton, Levenberg-Marquardt, etc) are more complex. As a consequence, we used a parametric method, which requires the choice of a model.

Several models are available for catadioptric sensors: Ad hoc model, polynomial approximation and generic model. The Ad hoc model was introduced in [9]. Mirror parameters, camera parameters and the intrinsic matrix of rigid transformation between them, are used to model the image formation. The large number of parameters to be estimated leads to an error function which is difficult to minimize because of local minima [10]. The polynomial approximation of the projection function was introduced by Scaramuzza [12], who proposed a calibration toolbox for his model. The generic model, also known as the unified model, was introduced by Geyer [2] and Barreto [4] who proved its validity for all central catadioptric sensors. This model was, then, modified by Mei [10], who generalized the projection matrix and also took into account the distortions. We choose to work with the unified model that can be applied with a reasonable number of parameters to be estimated. Figure 4 represents the projection process.

As described in [5] and [8], the projection of a 3D point can be done as follow:

(1) The 3D point X[w y z] is projected onto the unit sphere centered on C_m:

$$(X_s)_{C_m} = \frac{X}{\|X\|} \quad (1)$$

(2) Points $X_s = [x_s \; y_s \; z_s]^T$ are, then, projected onto the new frame with the origin $C_p = [0 \; 0 \; \xi]^T$:

$$(X_s)_{C_p} = (x_s \quad y_s \quad z_s + \xi)^T \quad (2)$$

(3) The obtained point $(X_s)_{C_p}$ onto a normalized plane:

$$m = \left(\frac{x_s}{(z+\xi_s)} \quad \frac{y_s}{(z+\xi_s)} \quad 1 \right)^T \quad (3)$$

(4) The last step enables us to find the camera projection matrix K expressed according to γ_u and γ_v, which are respectively, the generalized horizontal and vertical focal length, and (u_0, v_0) the coordinates of the principal point on the image point and the skew α:

$$p = K.m = \begin{pmatrix} \gamma_u & \gamma_u.\alpha & u_0 \\ 0 & \gamma_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}.m \quad (4)$$

In our model, we consider that the impact of the parameter α, often null, is irrelevant. Parameters to be estimated in that model are: $\xi, \gamma_u, \gamma_v, u_0$ and v_0.

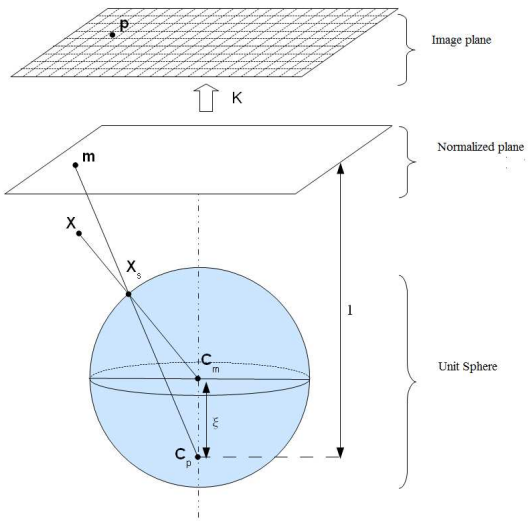


Fig.4. Unified projection model.

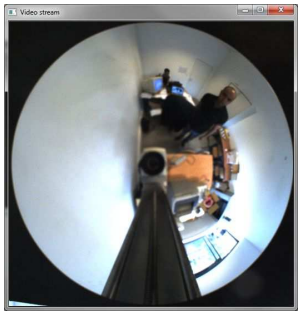
To estimate the previous parameters, Boutteau, in [3], has already implemented freely available software to estimate the parameters from several pictures. With the tool developed, calibration is achieved by observing a planar pattern at different positions. The pattern can be freely moved (the motion does not need to be known) and the user needs to select the four points corners pattern. This calibration process is similar to that of Mei [10]. It consists of a minimization over all the model parameters of an error function between the estimated projections of the pattern corners and the measured projection using Levenberg-Marquardt algorithm [13]. This minimization process enables to find the parameters combination that reduces the error of pattern retro projection.

4. Face detection & tracking on catadioptric pictures

Most of image processing techniques are performed on conventional images, i.e. perspective images. Actually, deformations caused by the catadioptric system do not give us the opportunity to perform the existing face detection algorithms on raw images (Fig.5, (a)). As a consequence, geometrical transformations should be performed to obtain a panoramic image close to perspective images, where the face detection algorithm will be performed.

4.1. Panoramic images unwrapping

The existence of the model of the unit sphere, as described in paragraph 3.2, simplifies the unwrapping problem. In fact, under the fixed view point constraint and by performing a retro projection, we are able to project the pixels of the panoramic pictures onto the unit sphere used in the unified model. Then, these pixels are projected onto the image plane. Thereby, we obtain the mapping between the pixels on the panoramic image and their corresponding on the camera retinal plane. Figure 5 (Fig.5, (b)) shows the result of the spherical transformation of the considered catadioptric image.
Face detection algorithm is applied to these unwrapped images.



(a)

(b)

Fig.6. Image transformation outcome: (a) Original catadioptric image (b) Unwrapped image obtained by spherical transformation

4.2. Face detection & tracking algorithms

Face detection techniques have been researched for years and much progress has been proposed in literature. However, in 2001, Paul Viola and Michael Jones [6] achieved a robust real time method for face detection, which is fifteen times quicker than the existing methods.
The technique relies on the use of simple Haar-like features that are evaluated quickly through the use of a new image representation called "integral image" that allows fast feature evaluation. Figure 8 represents two of the 60000 Haar-like features available. The difference of pixel intensities between the white and black regions enables us to define a certain threshold, during the training stage of the detector, characteristic to the face region.
To reduce the number of the computations required to achieve the rectangular filtering, we use the integral image at location p(x,y), which is the the sum of the pixel values above and to the left of the pixel including itself (Fig.7, (a)). As shown in figure 7 (Fig.7, (b)), let a, b, c and d be the values of the integral image at the corner of the region of interest. Then, the sum of the original image values within the ROI is:

$$\sum pixels = a + d - b - c \quad (5)$$

As a consequence, the sum is reduced to three operations. Moreover, the integral image is used for each Haar-like feature. It avoids repeated operations as large as the number of pixels in the ROI.

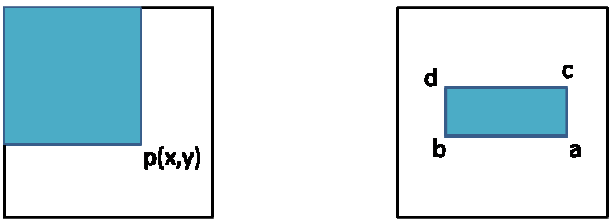


Fig.7. Integral Image: (a) Computation process (b) ROI pixels sum.

To select the best filtering feature, Adaboost, the machine learning introduced in [7], is used. In fact, given a set of weak classifiers, not much better than random, if we iteratively combine their output, the training error will quickly converge to zero. The best threshold for each filter is determined over the training test. Then, the best filter is selected. Misclassified examples cost is then reweighed and repeat. The learning process stops when all the examples are well classified. The filters are, then, ranked and linearly combined to create the final classifier.
Finally, to speed up the face detection, a set of cascade classifiers are used. The overall form of the detection process is that of a degenerate decision tree, what we call a "cascade". As explained before, the rectangular filters are ranked and weighted. A positive result from the first classifier triggers the evaluation of a second classifier which has also been adjusted to achieve very high detection rates, and so on (Fig.9). A negative outcome at any point leads to the immediate rejection of the sub-window.

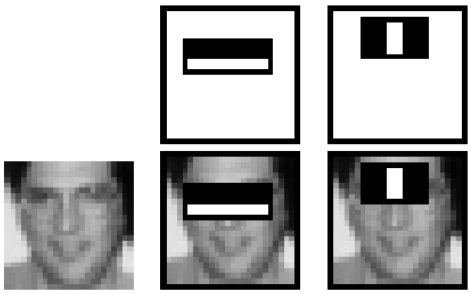


Fig.8. Haar like features (Courtesy of [6])

Once the face determined from the unwrapped panoramic image by Viola & Jones algorithm, we use the center of the corresponding ROI to perform the tracking algorithm. Actually, the new location of the face ROI is determined with correlation coefficients between the ROI and the region delimited by the rectangle (Fig.7, (b)). The maximum corresponds to the new location of the ROI center. The face regions are updated and the process is repeated from frame to frame. In the catadioptric face detection, only visual detection is allowed because of the low resolution of the resulting images. This is the principal reason of using a PTZ camera capable to obtain a zoomed in face ROI with a high resolution.

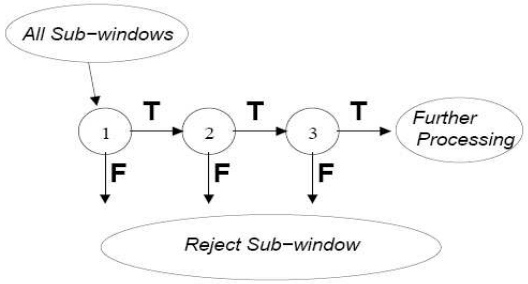


Fig.9. Nested Classifiers.

5. Fusion of Omnidirectional and PTZ cameras

5.1. Joint Calibration Strategy

The joint calibration method is based on defining a reference position (x_0, y_0) for the PTZ camera on the catadioptric 360 degree image is defined. This position has to coincide with the default orientation of the PTZ

camera and is chosen as the starting point for pan angle evaluation (Fig.10.). In fact, given a point *p(x, y)* and the image catadioptric image center $C(x_0, y_0)$, we can compute the pan angle θ_p in the omnidirectional referential:

$$\theta_p = \arctan\left(\frac{y - y_0}{x - x_0}\right) \quad (6)$$

The corresponding tilt angle is evaluated using the catadioptric sensor calibration by performing a retro projection of the corresponding catadioptric image pixel onto the unit sphere, as explained in paragraph 3.2.

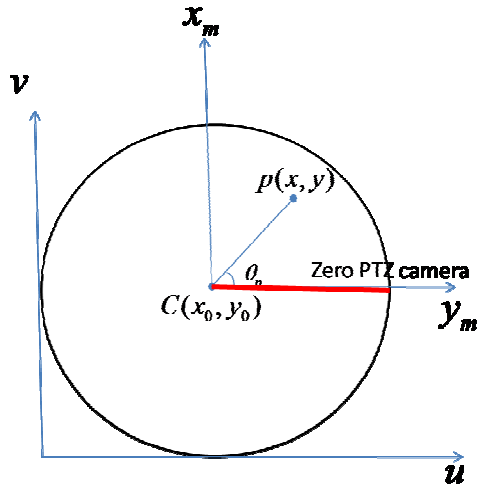


Fig.10. Determination of the pan angle

Due to the distance (baseline) between the catadioptric and PTZ sensors and a non-perfect alignment of their pan axis (non-accurate zero position reference), we introduce a transformation matrix between the Omnidirectional and PTZ cameras that provides the relationship between a given 3D point $(X,\ Y,\ Z)$ in the omnidirectional camera referential and its corresponding point $(X',\ Y',\ Z')$ in the PTZ camera referential (equation (7)).

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_X \\ t_Y \\ t_Z \end{pmatrix} \quad (7)$$

In the equation above, the rotation parameters (a_{ij}) represent the default alignment between both cameras and the translation vector $(t_X,\ t_Y,\ t_Z)^T$ corresponds to the baseline.

In our case, only the pan angle required for the PTZ camera has to be computed. As we have the pan angle θ_P determined from the catadioptric image, the associated vector *u* is given in the following expression:

$$u = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} = \begin{pmatrix} \cos(\theta_P) \\ \sin(\theta_P) \\ 0 \end{pmatrix} \quad (8)$$

The corresponding vector $v(v_x \quad v_y \quad v_z)^T$ in the PTZ camera referential can provide the calculation of the required pan angle α_{pan} of the PTZ camera by the following equation:

$$\tan(\alpha_{pan}) = \frac{v_y}{v_x} \quad (9)$$

Due to equations (8) and (9), only the parameters $(a_{11}, a_{12}, a_{21}, a_{22})$ have to be identified.

5.2. Face detection with PTZ camera

As explained in the previous sections, the catadioptric sensor is able to detect and track face ROIs. By using the ROI data localization from the catadioptric images, the PTZ camera can detect and also make a zoom in the face ROI. A brief description of the different steps for the face ROI localization are as follow:

1. Performing Viola & Jones face detection algorithm on the unwrapped catadioptric image to identify and localize the face ROI.
2. Using the pan angle of the ROI center, we calculate the pan angle to be directed to the PTZ camera. We use a constant tilt angle and a minimum zoom value. This enables us to point the PTZ camera on a large area where the face is probably located.
3. Performing, again, Viola & Jones face detection algorithm to detect the face ROI in the obtained PTZ image.
4. Commanding the PTZ camera in order to center the face ROI detected in the PTZ image (pan and tilt calculation) and then applying a zoom factor computed as expressed below:

$$Zoom = Min\left(\frac{\text{Im}age\_width}{ROI\_width}, \frac{\text{Im}age\_height}{ROI\_height}\right)$$

This last step enables us to obtain a zoomed in face image with high resolution, which can be useful for face recognition processing.

6. Results

In order to test the proposed system, the omnidirectional camera has been placed at 20 cm under the dome Axis PTZ camera. For the omnidirectinal camera, we opted for a 1609 per 396 pixels image. The PTZ camera ensures an image resolution of 704 x 576.

For the current tests, the architecture system used for measurement is a laptop based on Intel Core Duo 2.1GHz. For the implementation, C language and the Open CV library are used to develop the program. As shown in Figure 11, the first step consists in initializing the catadioptric acquisition process taking and image acquisition. The acquisition frame is 10 fps and the processing time period is about 50 ms. The resulting pictures are unwrapped according to the spherical projection. The processing time of this step is about 30 ms. Then, the program runs the face detection process in these unwrapped images on its own thread. This operation takes a time of 200 ms.
Once the face ROI localization is determined, the program sends the corresponding pan angle via http sockets. Then, a separated process is created to detect the face in the current PTZ image by applying Viola & Jones algorithm. By centering and zooming the PTZ camera on the face localization, we obtain an image of a high resolution face picture. The resulting face images can be registered and updated when running the face tracking algorithm.

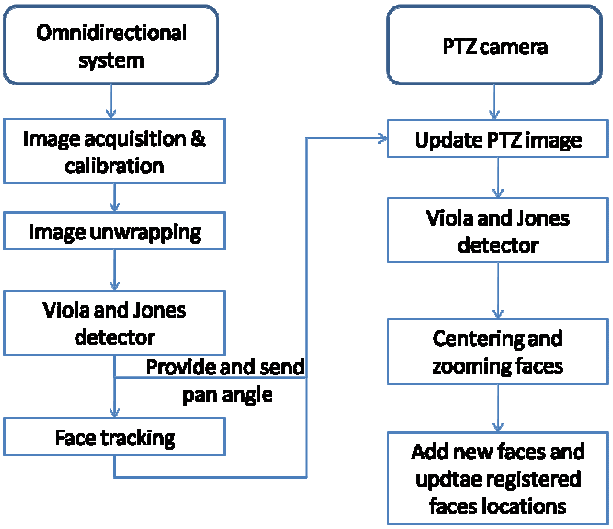


Fig.11. System architecture

Figure 12 illustrates the different processing steps performed in our program, enabling us to have an image of a zooming face with high resolution.

To further optimize the processing time of the application, possible improvements in the system architecture are objectives of the upcoming studies, notably in face detection and recognition algorithms, human gait and iris analysis.

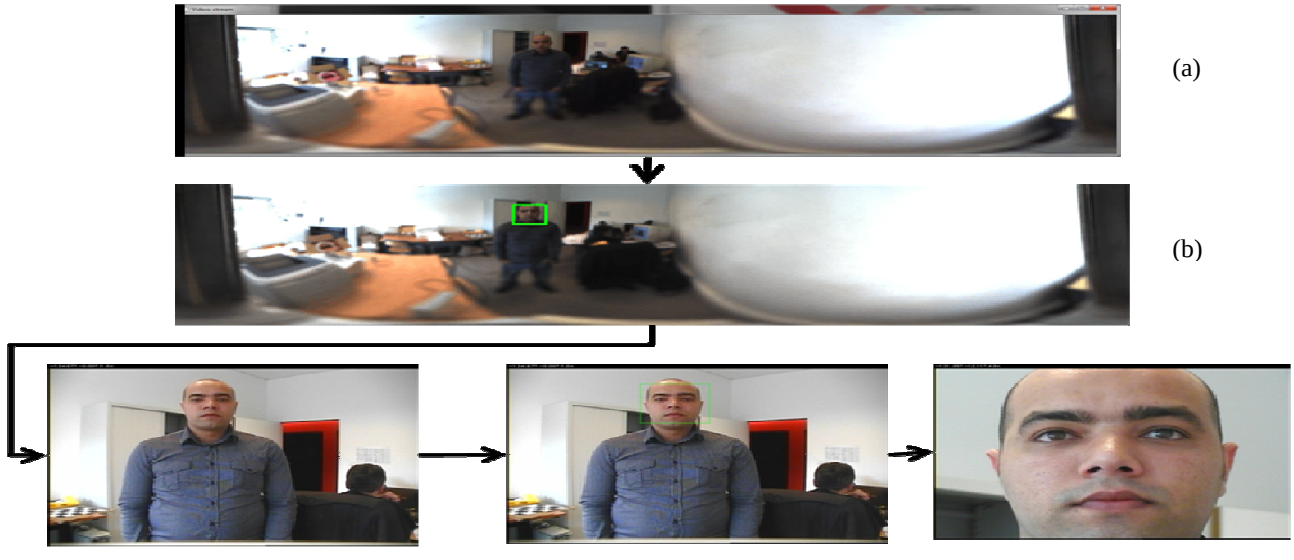


Fig.12. Illustration of face detection & zooming procedure using dual camera
(a) The unwrapped picture (b) Face detection in the unwrapped image (c) PTZ image corresponding to the provided pan angle (d) Face detection in the PTZ image (e) PTZ image with a centered and zoomed face

7. Conclusion

In this paper, a dual camera vision system, capable to automatically detect and track regions of interest at a higher zoom level, has been described.
Experiment results using robust calibration methods and real-time detection and tracking algorithms demonstrates a significantly improved accuracy in providing a closer look of the target for recognition purposes. Our future work focus on the improvement of the current prototype system architecture in order to reduce the processing time of our application.

8. Acknowledgment

This work is part of the Nomad Biometric Authentication (NOBA) project funded by ERDF under the Interreg IVA program (Ref. No. 4051) in collaboration with ESIGELEC/IRSEEM.

9. References

[1] Geyer, C., & Daniilidis, K., « A Unifying Theory for Central Panoramic Systems and Practical Implications », *Proceedings of the European Conference On Computer Vision, June 2000, pp 445-461.*

[2]Baker, S. & S.K, « A Theory of Single-Viewpoint Catadioptric Image Formation», *International Journal of Computer Vision, Vol. 103, No. 3, Setempber 2006, pp pp 175-196.*

[3] Boutteau.R, « Reconstruction tridimensionnelle de l'environnement d'un robot mobile, à partir d'informations de vision omnidirectionnelle, pour la préparation d'interventions », *Phd thesis, Université de Rouen, 2009.*

[4] Barreto, J.P, « A Unifying Geometric Representation for Central Projection Systems », *Computer Vision and Image Understanding, Vol. 103, No.3, September 2006, pp 208-217.*

[5] Mei, C. and Rives, P., « Single View Point Omnidirectional Camera Calibration from Planar Grids», *IEEE International Conference on Robotics and Automation (ICRA), 2007.*

[6] Paul Viola and Michael Jones, « Robust Real-time Object Detection», *International Journal of Computer Vision, 2001*

[7] Freund, Y., Schapire, R.E, « A Decision-Theoritic Generalization of on-Line Learning and an Application to Boosting», *Journal of Computer and System Science, 1997.*

[8] Dupuis, Y., « Nomad Biometric Authentication », *Master thesis, ESIGELEC, 2009.*

[9] Ragot, N., « Conception d'un capteur de stéréovision omnidirectionnelle: architecture, étalonnage et applications à la reconstruction 3D», *Phd thesis, Université de Rouen, 2009.*

[10] Mei, C., Rives, P., « Single View Point Omnidirectional Camera Calibration from Planar Grids», *Proceedings of the International Conference on Robotics and Automation (ICRA), April 2007, pp 3945-3950.*

[11] Ramalingram, R., Sturn, P. Lodha, S.K, « Towards Complete Generic Camera Calibration », *Proceedings of the International Conference on Computer Vision and Pattern Recognition, June 2005, pp 767-769.*

[12] Scaramuzza, D., Martinelli, A., Siegwart, R., «A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure From Motion», *Proceedings of the International Conference on Computer Vision, January 2006, pp 45-52.*

[13] Levenberg, K. «A method for the solution of certain probless in least squares», *Quarterly of Applied Mathematics, 1944, Vol.2, 164-168.*