



HAL
open science

Near infrared sensing and associated landmark detection for face recognition

Walid Hizem, Yang Ni, Bernadette Dorizzi

► **To cite this version:**

Walid Hizem, Yang Ni, Bernadette Dorizzi. Near infrared sensing and associated landmark detection for face recognition. *Journal of Electronic Imaging*, 2008, 17 (1), pp.011005-1 - 011005-7
10.1117/1.2898556 . hal-01371950

HAL Id: hal-01371950

<https://hal.science/hal-01371950>

Submitted on 26 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Near Infrared Sensing and Associated Landmark Detection for Face Recognition

Walid Hizem, Yang Ni and Bernadette Dorizzi

INT\GET, 9 Rue Charles Fourier, Evry, France;

{Walid.Hizem, Yang.Ni, Bernadette.Dorizzi}@int-edu.eu

ABSTRACT

We present, in this work, a new system of Near-Infrared capture that aims to eliminate ambient light effects. We designed this sensor to be integrated into the mobile platform VINSI. One of the originalities of this sensor is that we can acquire two images at the same time: one in infrared and the other in visible light. We developed a novel method for landmark detection of sufficiently low complexity so that it can be implemented on our mobile platform. We tested this method on three different near-infrared face databases, and we observed stability in the results of localization precision over these databases associated to performance comparable to the state of the art.

Keywords: Near-Infrared sensing, Landmark Detection, illumination suppression

1. INTRODUCTION

The illumination problem is one of the limitations in face recognition, especially in an outdoor environment. In ⁽¹⁾, the influence of the illumination - in the uncontrolled session in FRGC (Face Recognition Grand Challenge) - on the performance of the algorithms can be easily observed as the performance decreases by a factor of 1.6 between the controlled and uncontrolled cases. There are two solutions to tackle this problem. The first and most commonly used one is to eliminate analytically the illumination effect ⁽²⁾⁽³⁾⁽⁴⁾. However, the performance of these algorithms depends on the quality of the original image and the conditions of illumination. The analytical approach has also the disadvantage of being complex and time consuming. These algorithms are therefore not suitable for mobile devices where the processing capacities and memory are limited. The other direction (active approach) is to design active imaging techniques to obtain face images captured under consistent illumination conditions. The VINSI (Vérification d'Identité Numérique Sécurisé Itinérante) project aims at designing and implementing a mobile device for biometric identity verification through face and fingerprints. This device could be used to verify the identity of travellers carrying a biometric visa or passport. In this context of mobility and high illumination variations, active approaches are recommended. We therefore designed a new sensor capable

of acquiring simultaneously a normal-light image and an image in near infrared, thus allowing us to suppress the influence of the ambient light. The accurate finding of landmark position (like position of the eyes, the nose, the mouth) is an important process for image normalization (scale, pose and face expressions), and it is an a priori treatment before any face verification procedure. Landmarks can be well detected on Near InfraRed (NIR) images while their determination is more difficult in saturated normal lightened images because some part of the face could be undetectable. By acquiring both images at the same time, we can detect the facial landmarks in the NIR image and subsequently position them on the normal light image. We propose in this paper a new scheme for accurate landmark detection which we validate on several different databases acquired in our laboratory.

Our paper is divided into four sections. After a brief recall of the state of the art in Near Infrared sensing, we describe in more details the architecture and design of our VINSI sensor. A description of our landmark detection method, suitable for mobile devices, is presented in the third section. Finally, we validate our approach on different near-infrared face databases at our disposal

2. NEAR INFRARED SENSORS

Active sensing consists of eliminating ambient light during the acquisition of the image. In this process, the acquired image depends on an additional and controlled light and therefore requires a dedicated sensor. In this paper we will focus on infrared sensors for ambient lighting suppression. These sensors use the near infrared lighting to illuminate the scene instead of the ambient light. Much recent research ⁽⁵⁾⁽⁶⁾ has been done on this type of image, but using different sensor architectures. The most robust architecture relies on the differential acquisition. It is assumed that the ambient light is a DC component, so it can be suppressed by using a differential process. Two images are then acquired; one with ambient light and the second one with a near infrared light ⁽⁷⁾ and then a subtraction is performed between the two images. The resulting image is in near infrared. The disadvantage of this method is that the subtraction is done outside the sensor, which slows the acquisition flow and in consequence a motion problem can appear. In ⁽⁸⁾ we have shown that the differential operation can be done in the sensor. The idea is to do the subtraction analogically in the sensor. With this procedure we have improved the ambient light suppression resulting in a faster sensor than the method subtracting the two images outside the sensor. Another method uses a visible cut filter in the front of the camera ⁽⁹⁾⁽¹⁰⁾ and adds a near infrared illumination. This solution is optimized for indoor acquisitions or dark environmental conditions but not in situations where ambient light contains too much Near Infrared (sunlight) in which the long acquisition time does not allow any attenuation of ambient light.

In our laboratory, we developed another ambient light suppression sensor based on a CCD sensor with a pulsed flash ⁽¹¹⁾ as a preliminary step forwards the actual VINSI sensor presented in 3. In this approach, we

have shortened the acquisition time and we have used a synchronized flash. This sensor is based on a modified Logitech QuickCam Pro 4000 webcam equipped with a CCD sensor and the exposure time was controlled by the software of the webcam. The resolution of this camera is 640×480 . This solution presents the advantage of being built on a standard sensor similar to the differential sensor, but the acquisition time is set by software and then can't be set freely. We developed then a new sensor (VINSI infrared sensor) that can achieve an acquisition in infrared with programmable acquisition time and in ambient light with a small delay time ($20ms$).

3. VINSI INFRARED SENSOR

This new sensor can therefore acquire successively images in ambient and in infrared light. The acquisition is nearly simultaneous: indeed the delay time is only $20ms$. We combined the video acquisition protocol and the FlashCam principle to perform the dual acquisition. The output of our sensor is an image composed of two frames, the odd one being the visible image and the even one, the near infrared image. During the acquisition of the visible image, the Timing and synchro generator component fix the acquisition time. The integration time of the CCD sensor is set by the lapse of time between the last reset signal of the substrate (V_{xsub}) and the transfert signal (V_{xsg}). During the odd frame, we use the same principle as in the FlashCam. We modify the reset signal V_{xsub} by adding extra ones before the charge transfer from the CCD matrix to the charge lecture circuit (V_{xsg}). By this way, we are reducing the exposure time of the pixels (fig.1). During this time, we activate the flash through the V_{led} signal. We settled the exposure time in the odd frame to $50\mu s$. This exposure time is short enough to suppress the ambient light and long enough to have a good image quality in the near infrared spectrum.

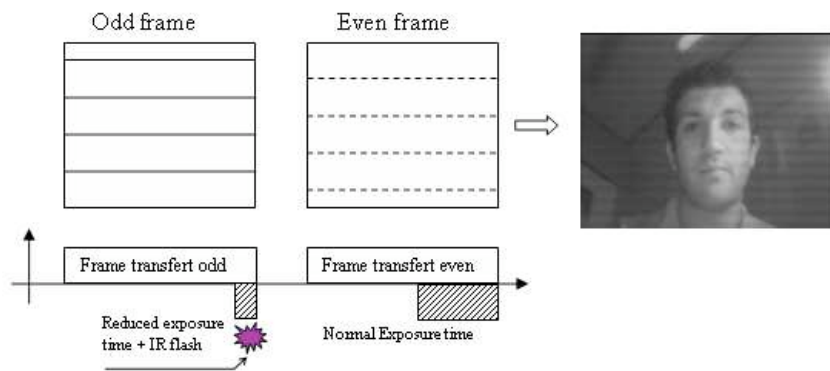


Figure 1. Acquisition Principle

The VINSI infrared sensor is a Black and White CCD of 500×582 (TV resolution) pixels. The architecture

of the module is shown in fig. 2. The module is composed of the CCD, a Samsung component, controlled via the vertical CCD driver with a Timing and synchro generator. We use the outputs of this component to control the acquisition mode and timing. We add to the CCD module a Video Sync Separator LM1881 and an ATMEGA32 microcontroller. To have a good image quality in infrared, we designed a flash composed of 40 LEDs surrounding the lens of the camera (2). We choose a powerful LEDs that can be pulsed and emitting an infrared light of $880nm$. In order to get optimized images, the subject must stand at 50cm from the camera. This distance

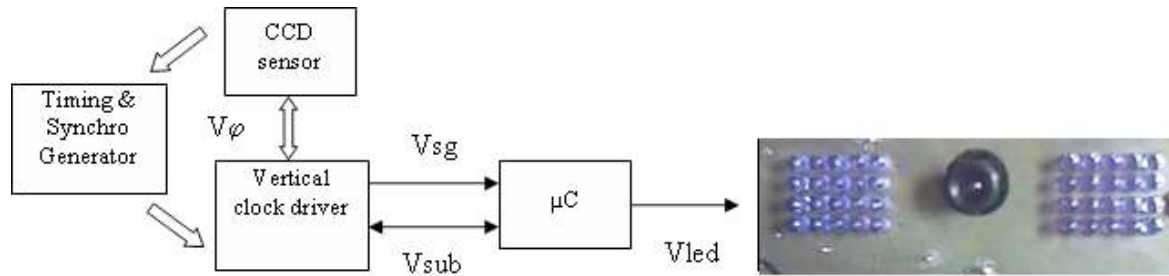


Figure 2. Architecture of the VINSI Module

depends on the radiant intensity and on the optic used in the prototype. After the acquisition phase, we extract the frames from the video and separate the odd and even lines from each frame. A linear interpolation allows the production of 2 images: one in near infrared and a second in visible light. We see in fig.3 that the ambient light has no effect on the near infrared image. The suppression of the ambient light is therefore very efficient.

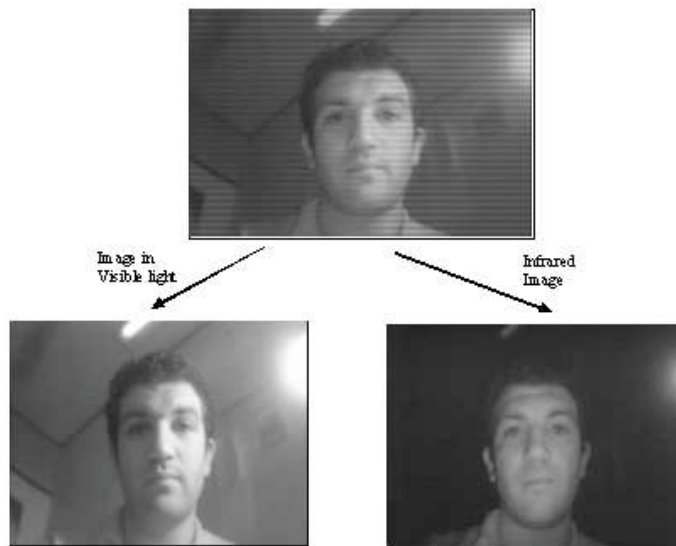


Figure 3. Image subsampling from the VINSI sensor

4. LANDMARKS DETECTION

During the development of the sensor described above, we noticed the remarkable image stability across illumination variation. We therefore know that the landmark detection performed on these images will be of equal quality, independently of the illumination conditions, contrary to what occurs in the case of normal light images. Moreover, we developed a landmark detection algorithm suitable for this type of images with the constraint of an easy integration on the chip. This does not allow us the use of statistical models, neither a high complexity methods. Taking into account that the characteristics of the near infrared images is their high contrast which facilitates the edge detection. In our method, we first apply a height-direction edge filter (figure 4). The resulting

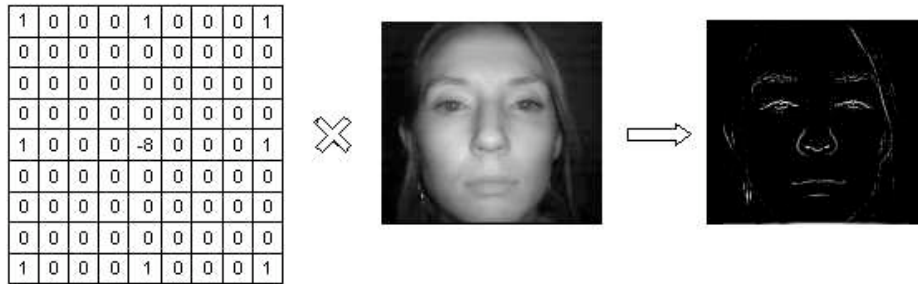


Figure 4. Edge detection in Infrared images

image accentuates the edges of the face and the landmarks (namely eyes, nose, mouth). We have noticed that a horizontal projection allows us to detect the ordinate of the landmarks ⁽⁵⁾. The maxima of the projection coincide with the eyes, the nose, the mouth, the chin and the nostrils. Therefore, we are going to analyze the area of those maxima to find out the position of the landmarks at interest (namely the eyes, the nose and the mouth which are the most stable and common landmarks). The same processing has been done on the visible images.



Figure 5. Samples of acquired images and corresponding edge detection

Due to the sensitivity of the edges to the illumination, the edge image doesn't contain enough information to be treated as in the infrared image. Indeed, we can see in figure 6 that due to the illumination problem we can't get the right part of the face in the edge image.



Figure 6. visible and infrared edge image

Detection of the mouth: The analysis of the three bottom pre-selected areas (by the horizontal projection) will lead us to the ordinate of the mouth. The mouth is present in the area where we can find the longest horizontal line. We apply an opening operation in the pre-selected region with a line structuring element to highlight the line elements. The longest line is selected by a horizontal projection. After that we calculate the barycentre of that line to fix the centre of the mouth.

Detection of the nose: The nose is selected from the zones which were pre-selected during the horizontal projection. The nose is indeed located just above the mouth. So, we analyze the upper areas of the mouth where we can find the model of the nose (presented in figure 7). Using a vertical projection, we will find two important gradients relative to the nose vertical borders.



Figure 7. Model of the nose

Detection of the eyes: The eyes are selected from the upper pre-selected zones, just above the nose. Erosion with a round structuring element in those areas will allow to select the two bigger area. Depending on their size we will know whether the zone contains or not the eyes. The centre of each eye is obtained by calculating the barycentre of the two main objects in that zone.

5. EXPERIMENTATIONS

In order to validate and test our method and the robustness of the VINSI sensor, we have acquired two databases called IV and INT-VINSI. We also tested our algorithm on images acquired with other infrared sensors in order to study its stability relative to the different NIR datasets. In this context, we have tested our method on the Differential sensor (DiffCam⁽⁸⁾) using the BIOMET infrared database ⁽¹²⁾ .

5.1 Databases Description :

Infrared IV^2 face database: We took advantage of a recent acquisition campaign in our institute to acquire this infrared face database using a first version of our sensor. This sensor was equipped with a wide angle lens (fisheye lens), and that led to hemispherical images. The acquisition took place in an indoor environment. We built a booth with a system to control the illumination composed of four halogen lamps: two behind the subject, one at his right and the last one at his left. There are 50 people in this database. For each person, we have at our disposal a video of 40 seconds. The resolution of those videos is 640x480. Every ten seconds we change the illumination conditions: all the four lamps are on, all the four lamps are off, left illumination and right illumination. We have extracted 10 images from videos from which we separate the odd and even frames in order to get the infrared image and the visible one. Some samples of this data base are shown in fig. 8. Note that none of the subjects wear glasses.



Figure 8. Samples of the visible light images in the IV^2 database

INT-VINSI face database: This database contains a set of 25 persons acquired in two sessions over several weeks. Each session contains three videos of ten seconds recorded using the VINSI infrared sensor. Three different illumination conditions are considered for each video. The difference with the previous database lies in the lens used in the sensor (6mm); we also modified the flash to a more powerful one. The first session corresponds to an ambient office lighting. In the second session we illuminate the face with a light source directed to the left. The last one is taken near an open window. The face is illuminated laterally from the right. The intensity of the illumination depends on the daytime and on the weather. Some samples of this database are illustrated in figure 9.

BIOMET: This NIR face database was acquired during the BIOMET campaign acquisition (¹²) with the differential sensor(⁸). The resolution of the images is 160x120 i.e: lower then in the two previous database. We have excluded from our tests people who wear glasses, in order to avoid the problem of reflection.

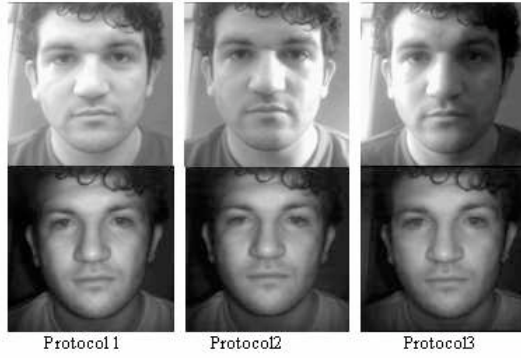


Figure 9. Sample of the INT-VINSI database

5.2 Performance:

The criteria that we use for landmark detection evaluation is the one originally proposed by Jesorsky and al.⁽¹³⁾. This criterion is based on the intraocular distance as follows:

$$d_{eye} = \frac{\max(d_l, d_r)}{\|C_l - C_r\|}$$

Where C_l , C_r are the real centre of the left(respectively right eye) and d_l , d_r are the distances between the detected eye left (respectively right) centre and the real left (respectively right) eye coordinate. Note that the real centre positions are detected manually and can be subject to some error.

To measure the performance of the nose and mouth detection, we used the same principle as for the d_{eye} :

$d_{nose} = \frac{d_n}{\|C_l - C_r\|}$, where d_n is the distance between the detected position of the nose and the one manually determined.

$d_{mouth} = \frac{d_m}{\|C_l - C_r\|}$, where d_m is the distance between the detected position of the mouth and the one manually determined.

In ⁽¹⁴⁾, it was established experimentally that in order to perform a good verification the d_{eye} localization precision has to be lower than 0.05. In other work⁽¹³⁾, it is shown that for an eye localization accuracy of 0.25, the centre of the eye can still be detected inside the eye region. The results of the tests on the IV² database are shown in figure 10: Our method has a ratio of 75% of good eye localization at factor of 0.05 on the IV² database. The performance of our method in the INT-VINSI database is shown in figure 11: On this database the precision of the eye localization is better than on the IV² database. Indeed, even if we use the same capture method, images in INT-VINSI are of better quality (no hemispherical distortion and availability of a better flash). We notice that we have a ration of 80% for $d_{eye} = 0.05$. The performance on the BIOMET database is shown in

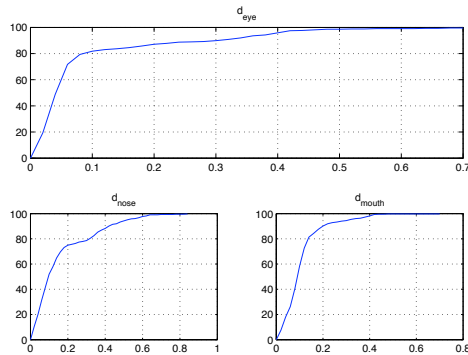


Figure 10. Cumulative histograms of d_{eye} , d_{mouth} and d_{nose} (IV^2)

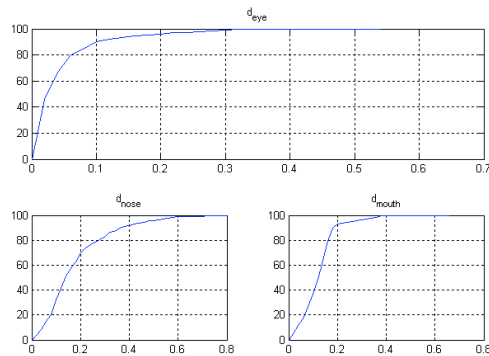


Figure 11. Cumulative histograms of d_{eye} , d_{mouth} and d_{nose} (INT-VINSI)

figure 12: We just measured the error of eye localization for BIOMET, because we do not have at our disposal the mouth and the nose position. We reach a localization rate of 80% for $d_{eye}=0.1$ in this database. This is less than the previous results, because of the limited size of the face in the images (the average size is 50×70 pixels).

6. CONCLUSION

In this work we presented a new sensor for ambient light suppression. This sensor is to be imbedded in the VINSI mobile platform to control the identity of citizens who hold biometric passports. This new sensor is capable of acquiring two images at the same time: the first one is taken in visible light; the second in near-infrared. We also developed a method of landmark detection which we tested on near infrared images acquired with different types of sensors. The best result, in terms of precision of the localization, was found on the images acquired with our new VINSI sensor, which was to be expected because they have the best quality. However, our experiments show that this method can be used on different types of infra-red images and has produced good results. Better performance can be obtained with learning-based methods (such SVM in⁽¹⁵⁾), but these methods can not be

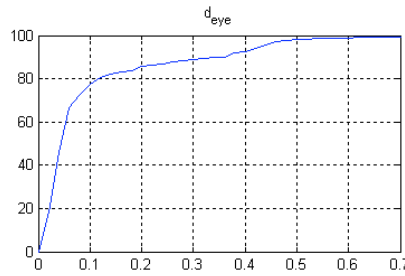


Figure 12. Cumulative histograms of d_{eye} (BIOMET)

simply implemented on a mobile device having limited resources.

Our future works will aim to use these NIR landmarks to provide eye localization on the associated visible image for further processing. Another direction concerns the realization of a verification algorithm for NIR images suitable for analogical architecture.

ACKNOWLEDGMENTS

This work has been founded by the French project VINSI (Vérification d'identité Numérique Sécurisé Itinérante)

REFERENCES

1. P. Phillips and Al., "Overview of the face recognition grand challenge," *Conference on Computer Vision and Pattern Recognition* **1**, pp. 947 – 954, 2005.
2. S. Shan and Al., "Illumination normalization for robust face recognition against varying lighting conditions," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures* , pp. 157–164, 2003.
3. D. Jobson, Z. Rahman, and G. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing* **6**, pp. 965–976, 1997.
4. X. Xie and K. Lam, "An efficient illumination normalization method for face recognition," *Pattern Recognition Letters* **27**, 2005.
5. J. Dowdall, I. Pavlidis, and G. Bebis, "Face detection in the near-ir spectrum," *Image and Vision Computing* **21**, pp. 565–579, 2003.
6. D. Li and W. Liao, "Facial feature detection in near infrared images," *Proc. Fifth Int'l Conf. Computer vision Pattern Recognition and Image Processing* , pp. 26–30, 2003.
7. X. Zou, J. Kittler, and K. Messer, "Ambient illumination variation removal by active near-ir imaging," *Proceedings of IAPR International Conference on Biometric* , pp. 19–25, 2006.

8. Y. NI and X.-L. YAN, “Cmos active differential imaging device with single in-pixel analog memory,” *Proceeding of ISSCC02* , pp. 359–362, 2002.
9. S. Z. Li and Al., “Illumination invariant face recognition using near-infrared images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(4), pp. 627–639, 2007.
10. W.-H. Liao and D.-Y. Li, “Homomorphic processing techniques for near-infrared images,” *Proceedings of Acoustics, Speech, and Signal Processing. (ICASSP '03)* **3**, pp. III– 461–4, 2003.
11. W. Hizem and Al., “Specific sensors for face recognition,” *Int’l Conf. Audio- and video-Based Biometric Person Authentication* , pp. 47–54, 2006.
12. S. Garcia-Salicetti and Al., “Biomet : A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities,” *AVBPA Proceedings* **2688**, p. 1056, 2003.
13. O. Jesorsky, K. Kirchberg, and R. Frisholz, “Robust face detection using the hausdorff distance,” *Int’l Conf. Audio- and video-Based Biometric Person Authentication* , pp. 90–95, 2001.
14. M. Sadeghi, J. Kittler, A. Kostin, and K. Messer, “A comparative study of automatic face verification algorithms on the banca database,” *Proc. Fourth Int’l Conf. Audio- and Video-Based Biometric Person Authentication* , pp. 35–43, 2003.
15. X. Zou, J. Kittler, and K. Messer, “Accurate face localisation for faces under active near-ir illumination,” *Proceedings of the 7th Int’l Conference on Automatic Face and Gesture Recognition (FGR’06)* , pp. 369–374, 2006.

List of Figures

| | | |
|----|--|---|
| 1 | Acquisition Principle | 3 |
| 2 | Architecture of the VINSI Module | 4 |
| 3 | Image subsampling from the VINSI sensor | 4 |
| 4 | Edge detection in Infrared images | 5 |
| 5 | Samples of acquired images and corresponding edge detection | 5 |
| 6 | visible and infrared edge image | 6 |
| 7 | Model of the nose | 6 |
| 8 | Samples of the visible light images in the IV^2 database | 7 |
| 9 | Sample of the INT-VINSI database | 8 |
| 10 | Cumulative histograms of d_{eye} , d_{mouth} and d_{nose} (IV^2) | 9 |

| | | |
|----|---|----|
| 11 | Cumulative histograms of d_{eye} , d_{mouth} and d_{nose} (INT-VINSI) | 9 |
| 12 | Cumulative histograms of d_{eye} (BIOMET) | 10 |