Thermal Face Recognition using Local Interest Points and Descriptors for HRI Applications*

G. Hermosilla, P. Loncomilla, J. Ruiz-del-Solar
Department of Electrical Engineering, Universidad de Chile
Center for Mining Technology, Universidad de Chile
jruizd@ing.uchile.cl

Abstract. In this article a robust thermal face recognition methodology based on the use of local interest points and descriptors, is proposed. The methodology consists of the following stages: face segmentation, vascular network detection, wide baseline matching using local interest points and descriptors, and classification. The main contribution of this work is the use of a standard wide baseline matching methodology for the comparison of vascular networks from thermal face images. The proposed methodology is validated using a database of thermal images. This work could be of high interest for HRI applications related with the visual recognition of humans, as the ones included in the RoboCup @Home league, because the use of thermal images may overcome limitations such as dependency on illumination conditions and facial expressions.

Keywords: Face Recognition, Thermal Images, Blood Vessels Matching, SIFT Matching, RoboCup @Home.

1 Introduction

Face analysis plays an important role in building HRI (Human-Robot Interaction) interfaces that allow humans to interact with robot systems in a natural way. Face information is by far the most used visual cue employed by humans. There is evidence of specialized processing units for face analysis in our visual system [1]. Face analysis allows localization and identification of other humans, as well as interaction and visual communication with them. Therefore, if human-robot interaction must achieve the same efficiency, diversity, and complexity that human-human interaction has, face analysis should be extensively employed in the construction of HRI interfaces.

Currently, computational face analysis is a very lively and expanding research field. Face recognition, i.e. the specific process for determining the identity of an individual contained in an image area which has been already identified as containing a face (by a face detection system) and already aligned (by a face alignment process which usually includes eye detection), is a functional key for personalizing robot services and for determining robot behaviors, usually depending on the identity of the

* This research was partially funded by FONDECYT under Project Number 1090250.

human interlocutor. Many different face recognition approaches have been developed in the last few years [1]-[5], ranging from classical Eigenspace-based methods (e.g. eigenfaces [6]), to sophisticated systems based on high-resolution images or 3D models. Many of these methods are well suited to specific requirements of applications such as biometry, surveillance, or security. HRI applications have their own requirements (variable illumination conditions, variable face expressions, just one image per person, in many cases), and therefore some approaches are better suited for them. It would be useful for developers of face recognition systems for HRI to have some guidelines about the advantages of some thermal methodologies over others.

In this context, the aim of this work is to propose and evaluate the usage of vascular networks from thermal face images for recognition purposes, and the use of wide baseline methods for matching both, thermal images and vascular networks.

The main contribution of this work is the use, for the first time, of a standard wide baseline matching methodology for the matching of vascular networks from thermal face images. This approach enables to build recognition systems using just one image per subject, which are robust to variable illumination and variable face expressions.

This paper is organized as follows. The related work is described in section 2. The proposed system is explained in section 3. Description of the experiments and results are outlined in section 4. Finally, conclusions are given in section 5.

2 Related Work

Biometric techniques have attracted increasing interest from both research and industrial communities because of the strong immunity to forgery that they can achieve. Face recognition is one of the main topics in the set of biometric applications as it is one of the most human-used, non-intrusive and user-friendly approaches for biometric recognition and have been a very important issue in the computer vision community, generating a lot of related research on image processing and statistical classifiers and achieving encouraging results [1]-[4][6][7]. Visible-spectrum images have high variability because they are produced by reflection in surfaces, which has strong dependence on luminosity and spatial distribution of the light sources which usually have strong differences in time, and their dependence on reflectivity make possible to fool the system using some simple tricks like photographs or dummy faces. These problems encouraged the interest in the use of thermal images for face recognition because of the great resistance to forgery it can achieve and their high immunity to illumination changes and several other sources of variability in the acquisition of images that enables these systems to operate in real non-controlled environments [8][9]. Face recognition systems using thermal images use techniques adapted from visible-spectrum image face recognition [8]-[10], and new techniques that are specific for thermal images [18][19]. A comparison of several techniques for thermal faces is shown in [20]. Algorithms for obtaining vascular networks from thermal images using segmentation and morphologic operators [18][19] are a very useful tool for building face recognition techniques as they provide a very simple and repeatable imprint representation that is unique for each person, and enable recognition through the matching of these imprints using some specialized or general image matching methodology from the impressive rich variety of techniques that are devoted to this last task.

Wide baseline matching (object recognition) approaches based on local interest points (invariant features) have become increasingly popular and have experienced an impressive development in the last years [14][11][16][21]. Typically, local interest points are extracted independently from both a test and a reference image, and then characterized by invariant descriptors, and finally the descriptors are matched until a given transformation between the two images is obtained. Most employed local detectors are the Harris detector [12] and the Lowe's sDoG+Hessian detector [14], being the Lowe's detector multiscale and the Harris detector single scale. Best performing affine invariant detectors are the Harris-Affine and the Hessian-Affine [17], but they are too slow to be applied in general-purpose applications. The most popular and best performing invariant descriptor [17] is the SIFT (Scale Invariant Feature Transform) [14]. For selecting the local detector and invariant descriptor to be used in a given application it should be taken into account the algorithm's accuracy, robustness and processing speed. Lowe's system [14] using the SDoG+Hessian detector, SIFT descriptors and a probabilistic hypothesis rejection stage is a popular choice, given its recognition capabilities, and near real-time operation. However, Lowe's system main drawback is the large number of false positive detections. This is a serious problem when using it in real world applications.

One of the main weaknesses of Lowe's algorithm is the use of just a simple probabilistic hypothesis rejection stage, which cannot successful reduce the number of false positives. Loncomilla and Ruiz-del-Solar (L&R) propose a system that reduces largely the number of false positives by using several hypothesis rejection stages [15][21]. This includes a fast probabilistic hypothesis rejection stage, a linear correlation verification stage, a geometrical distortion verification stage, a pixel correlation verification stage, a transformation fusion procedure, and the use of the RANSAC algorithm and a semi-local constraints test. In [21] are compared the Lowe's and the L&R systems using 100 pairs of real-world high-textured images (variations in position, view angle, image covering, partial occlusion, in-plane and out-of the-plane rotation). The results show that in this dataset the L&R system reduces the false positive rate from 85.5% to 3.74%, by increasing the detection rate by 5%. For this reason we choose to use this system in this work.

3 Proposed Thermal Face Recognition System

The proposed approach use wide-baseline matching of face vascular networks obtained from thermal images. The vascular networks are obtained through skin segmentation and morphological operators. The image matching stage uses SIFT descriptors and the L&R system for verifying correspondences and generating a final geometrical transformation that relate the vascular networks. A classifier makes a final decision about the recognition. The proposed approach mixes the uniqueness of the representation obtained through vascular networks, the robustness of the recognition based on local descriptors, and a classifier that make the final decision for providing a new methodology for thermal face recognition. A description of the system is provided in the next subsections.

3.1 System Outline

The system is composed by four stages: face segmentation, vascular network extraction, wide-baseline matching and a statistical classifier. A block diagram of the system is shown in figure 1.

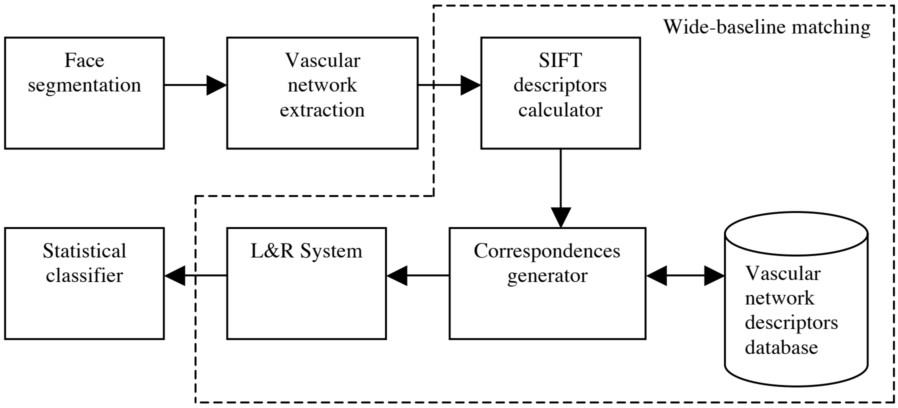


Fig. 1. Diagram of the proposed thermal face recognition system.

3.2 Face Segmentation

A skin detection model in the thermal spectrum is created by modelling both, the skin pixel intensity distribution, and the non-skin distribution, as mixtures of Gaussians (MoG). Both the skin and non-skin intensity distributions are modelled as the mixture of four Gaussians:

$$p_{SKIN}(x) = \sum_{i=1}^{4} \omega_S \frac{1}{\sqrt{2\pi}\sigma_S} \exp\left(-\frac{1}{2} \frac{(x-\mu_S)^2}{\sigma_S^2} \right) \quad (1)$$

$$p_{NON-SKIN}(x) = \sum_{i=1}^{4} \omega_N \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{1}{2} \frac{(x-\mu_N)^2}{\sigma_N^2} \right) \quad (2)$$

A pixel is detected as skin when its intensity has a larger probability of belonging to the skin class than to the non-skin class. The application of this scheme on a face thermal image enables the calculation of a skin mask:

$$I_{MASK}(i,j) = \begin{cases} 1 & if\ p_{SKIN}(I(i,j)) > p_{NON-SKIN}(I(i,j)) \\ 0 & if\ p_{SKIN}(I(i,j)) < p_{NON-SKIN}(I(i,j)) \end{cases} \quad (3)$$

The skin image is obtained by multiplying the thermal image and the skin mask, as it is shown in (4). General parameters for the Gaussians cannot be obtained as they depend on the response of the particular camera to thermal intensity.

$$I_{SKIN}(i,j) = I_{MASK}(i,j)I(i,j) \tag{4}$$

3.3 Vascular network extraction

A smoothed skin image is calculated by filtering the skin image using a 3x3 uniform low-pass filter:

$$I_{SMOOTH} = I_{SKIN} * \frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{5}$$

The smoothed image is eroded and dilated using morphological operations between the image and a 3x3 diagonal cross operator for obtaining a morphological opened image:

$$I_{OPENED} = \left(I_{SMOOTH} \Theta \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \right) \oplus \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{6}$$

Finally, the vascular network image is calculated by subtracting the smoothed image and the opened one:

$$I_{VASCULAR} = I_{SMOOTH} - I_{OPENED} \tag{7}$$

3.4 Wide baseline matching using local interest points

SIFT descriptors are obtained from the vascular network image using the standard Lowe's methodology [14]. Each SIFT descriptor represents a small patch in the image, and it contains a position (x_0,y_0), an orientation (θ), a scale (σ) and a vector of characteristics (v_1,…,v_{128}) calculated from the gradients in the patch. Correspondences between two vascular images are obtained by matching pairs of descriptors with similar vectors of characteristics using the Euclidean distance.

Each correspondence between two descriptors has an associated translation (t_X,t_Y), rotation θ, and scale change s that warps one patch into the other. Coherent groups of correspondences that generate similar warping parameters (t_X,t_Y,θ,s) are found via a Hough transform. For each coherent group of correspondences a probability test is carried out. Following, the L&R methodology [15][21], additional geometrical and statistical verification tests are applied in order to eliminate wrong groups of correspondences, which do not represent similar patches in both images.

In figure 2 are shown two examples of matches between pairs of thermal images and vascular network images, corresponding to the same person, and to different persons.

3.5 Classification

A final coherent group of correspondences is obtained for each test image by matching it to all gallery images. From all the final groups of correspondences, the most numerous one is selected in order to decide the correct gallery image, and to recognize the identity of the person.

4 Results

Results are obtained using the UXX Thermal Faces Database. This database contains 156 320x240 thermal face images that correspond to 6 images per subject and 26 subjects. The images were using a Cedip Jade UC infrared camera.

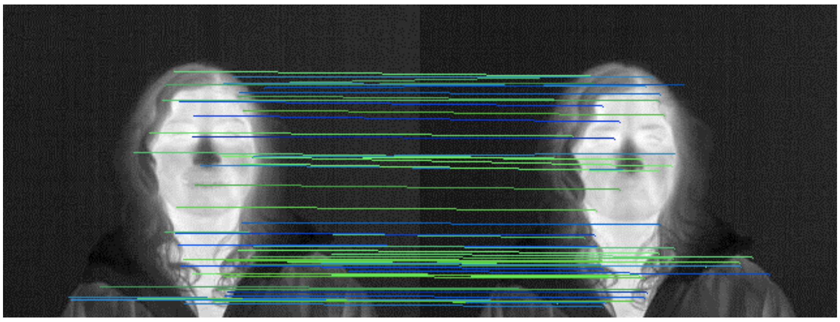
The proposed methodology consisting on the matching of vascular networks of thermal images is compared with the matching of plain thermal images (see examples in fig. 2). In both cases the matching is implemented using Lowe's system and the improved L&R system. Thus, the following four recognition systems are implemented and compared:

- *Lowe-Thermal*: Lowe's matching of plain thermal images.
- *L&R-Thermal*: L&R's matching of plain thermal images.
- *Lowe-VascularNet*: Lowe's matching of vascular networks of thermal images.
- *L&R-VascularNet*: L&R's matching of vascular networks of thermal images.

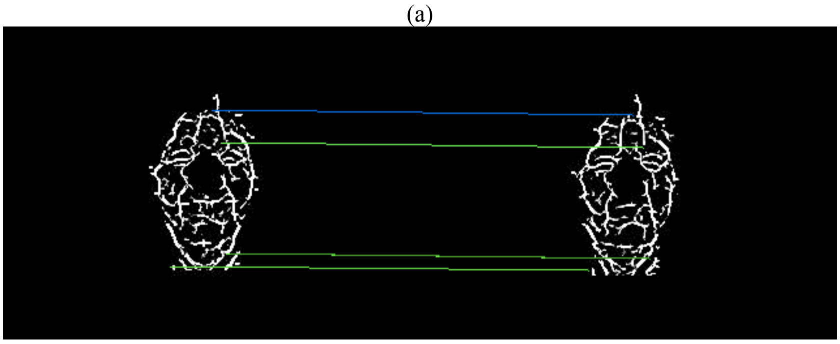
4.1 Database

The UXXX Thermal Faces Database was obtained by using a Cedip Jade UC infrared camera. This camera has a spectral range between 8-12 μm and a resolution of 320x240 pixels [22]. A video was recorded for each of the 26 subjects kept in a fixed position. The subjects gesticulated vowels (2,000 frames captured), and the happy/sad/anger expressions (100 frames captured). From the video for each subject, three frame images containing different vowels, and three frame images containing different expressions were selected to build the database. The images are divided into categories depending on the frame number in the video.

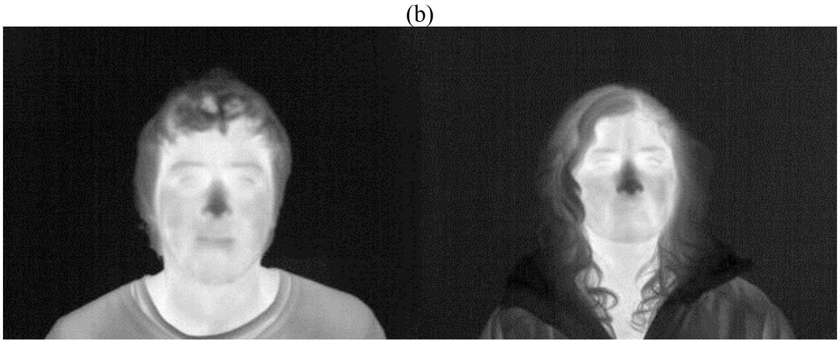
- V1: Frame 100 in the vowel's sequence video, 26 images
- V2: Frame 1,000 in the vowel's sequence video, 26 images
- V3: Frame 1,500 in the vowel's sequence video, 26 images
- E1: Frame 45 in the happy expression video, 26 images
- E2: Frame 45 in the sad expression video, 26 images
- E3: Frame 45 in the anger expression video, 26 images



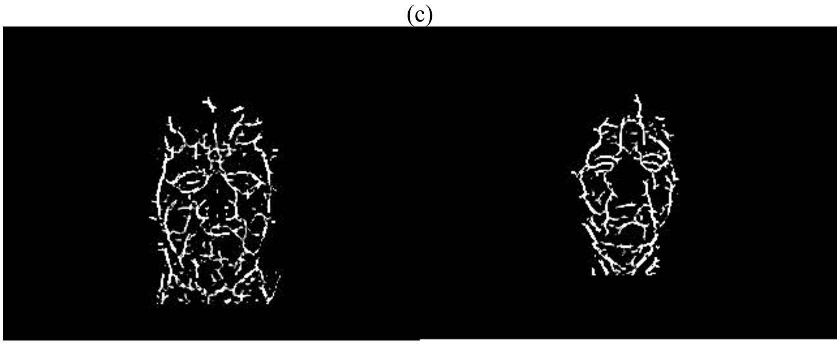
(a)



(b)



(c)



(d)

Fig. 2. Matching of pairs of thermal images ((a)-(b)), and the corresponding vascular network images ((c)-(d)). In (a) and (b) the images corresponds to the same person, while in (c) and (d) they correspond to different persons.

4.2 Recognition Results

Cross-validation is used for evaluating the performance of the different methods in the database. From the set of categories {V1, V2, V3, E1, E2, E3} all the possible pairs of categories (A,B) are formed. For each pair (A,B), the images in A are selected as test images, while the images in B are selected as gallery images. Each image in the test set A is recognized in the gallery B, which generate both correct and incorrect recognitions. The proportion of correct recognitions (recognition rate) for all the pairs of categories is shown in tables 1 to 4. The mean and variance of the overall recognition rates for the four methods are shown in table 5.

Table 1: Recognition rates for *Lowe-Thermal* method.

Test\Gallery	V1	E1	V2	E2	V3	E3
V1	100	84.6	96.1	80.7	100	96.1
E1	92.3	100	92.3	92.3	96.1	100
V2	100	88.4	100	92.3	100	96.1
E2	84.6	96.1	80.7	100	80.7	100
V3	100	92.3	96.1	76.9	100	88.4
E3	96.1	100	96.1	100	96.1	100

Table 2: Recognition rates for *L&R-Thermal* method.

Test\Gallery	V1	E1	V2	E2	V3	E3
V1	100	92.3	100	80.7	100	88.4
E1	96.1	100	96.1	88.4	88.4	100
V2	96.1	100	100	80.7	100	100
E2	88.4	100	80.7	100	84.6	100
V3	92.3	88.4	100	80.7	100	84.6
E3	88.4	100	92.3	96.1	96.1	100

Table 3: Recognition rates for *Lowe-VascularNet* method.

Test\Gallery	V1	E1	V2	E2	V3	E3
V1	100	88.4	100	84.6	96.1	100
E1	96.1	100	96.1	100	96.1	96.1
V2	100	92.3	100	92.3	100	96.1
E2	96.1	100	88.4	100	73.0	100
V3	100	84.6	100	84.6	100	92.3
E3	100	96.1	100	96.1	100	100

Table 4: Recognition rates for *L&R-VascularNet* method.

Test\Gallery	V1	E1	V2	E2	V3	E3
V1	100	88.4	96.1	76.9	92.3	96.1
E1	96.1	100	92.3	92.3	92.3	88.4
V2	96.1	100	100	88.4	100	100
E2	92.3	100	84.6	100	96.1	96.1
V3	88.4	96.1	100	76.9	100	96.1
E3	100	88.4	92.3	96.1	96.1	100

Table 5: Overall recognition rates of the four methods.

Method	**Mean**	**Standard Deviation**
L&R-Thermal	95.7	6.2
Lowe-Thermal	94.3	6.1
L&R-VascularNet	94.2	6.8
Lowe-VascularNet	93.9	6.9

4.3 Recognition Analysis

The L&R system applied directly over the thermal image generate the better results; however, the recognition rates for the four methods have comparable means and variances. The difference between the means of the better and the worst method is around 1.8, which is small compared to the variances around 6.5 obtained in the experiments. These experiments show that the transformation from the original thermal images to the vascular network representations preserve information that permit the recognition of the subjects.

5 Conclusions

The main contribution of this work is the use, for the first time, of a standard wide baseline matching methodology for the matching of vascular networks from thermal face images. This work is a preliminary step in the development of face recognition systems using vascular networks detection in thermal images followed by general image matching algorithms. All the variants compared in this work have similar performance for recognizing faces from thermal images. This results shows that vascular network images preserve important discriminative information about the original thermal images. As the vascular network representation has a simple and particular structure, specialized methods can be developed in future works for increasing dramatically the face recognition rates using thermal imaging systems.

As future work we would like to increase largely the size of the database, and to include outdoor illumination. The proposed recognition methodology will be validated in this extended database.

References

1. P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face Recognition by Humans: 19 Results All Computer Vision Researchers Should Know About, Proc. of the IEEE, Vol. 94, No. 11, Nov. 2006, pp. 1948-1962
2. W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, ACM Computing Surveys, 2003, pp. 399-458.
3. X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, Face recognition from a single image per person: A survey, Pattern Recognition, Vol. 39, pp. 1725–1745, 2006.
4. R. Chellappa, C.L. Wilson, S. Sirohey, Human and Machine Recognition of Faces: A Survey, Proceedings of the IEEE, Vol. 83, Issue 5, May 1995, pp. 705-740.
5. Face Recognition Homepage. Available in January 2008 in: http://www.face-rec.org/
6. M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence, Vol. 3, No. 1, 1991, pp. 71-86.
7. J. Ruiz-del-Solar, R. Verschae, M. Correa, (2009). Recognition of Faces in Unconstrained Environments: A Comparative Study. EURASIP Journal on Advances in Signal Processing (Recent Advances in Biometric Systems: A Signal Processing Perspective), Vol. 2009, Article ID 184617, 19 pages.

8. D. A. Socolinsky and A. Selinger, A comparative Analysis of face recognition performance with visible and thermal infrared imagery, in Proc. ICPR 2002, Quebec, Canada, August 2002.
9. A. Selinger and D. Socolinsky, Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study, Tech. Rep., Equinox Corporation, 2001.
10. S. Desa, S. Hati, IR and Visible Face Recognition using Fusion of Kernel Based Features, ICPR 2008:1-4.
11. Ferrari, V., Tuytelaars, T., Van Gool. L.: Simultaneous Object Recognition and Segmentation by Image Exploration. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 40 – 54. Springer, Heidelberg (2004).
12. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. 4th Alvey Vision Conf., pp. 147-151. Manchester, UK. (1998).
13. Lowe, D.: Local feature view clustering for 3D object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, pp. 682-688. IEEPress, New York (2001).
14. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. Int. Journal of Computer Vision 60 (2), 91-110 (2004).
15. Loncomilla, P., Ruiz-del-Solar, J.: A Fast Probabilistic Model for Hypothesis Rejection in SIFT-Based Object Recognition. In: Martinez-Trinidad, J.F., Carrasco, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol 4225, pp. 696-705. Springer, Heidelberg (2006).
16. Mikolajczyk, K., Schmid, C.: Scale & Affine Invariant Interest Point Detectors. Int. Journal of Computer Vision 60 (1), 63 – 96 (2004).
17. Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. 27 (10), 1615 – 1630.
18. P. Buddharaju, I. Pavlidis, “Multi-Spectral Face Recognition - Fusion of Visual Imagery with Physiological Information”. Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems, Springer, pp. 91-108, January, 2007.
19. P. Buddharaju, I. Pavlidis, C. Manohar, “Face Recognition Beyond the Visible Spectrum”, Advances in Biometrics: Sensors, Algorithms and Systems, Springer, pp. 157-180, October, 2007.
20. G. Hermosilla, J. Ruiz-del-Solar, R. Verschae, M. Correa, (2009). “Face Recognition using Thermal Infrared Images for Human-Robot Interaction Applications: A Comparative Study”, 6th IEEE Latin American Robotics Symposium – LARS 2009, Oct. 29 - 30, Valparaíso, Chile (CD Proceedings).
21. J. Ruiz del Solar, P. Loncomilla, “Robot Head Pose Detection and Gaze Direction Determination Using Local Invariant Features”, Advanced Robotics, Volume 23, Number 3, pp. 305-328(24), 2009.
22. Jade camera link http://www.edevis.de/products/products_ircameras_jade_uc_en.php