J Intell Robot Syst (2012) 66:223–243
DOI 10.1007/s10846-011-9612-2

Human Detection and Identification by Robots Using Thermal and Visual Information in Domestic Environments

Mauricio Correa · Gabriel Hermosilla · Rodrigo Verschae · Javier Ruiz-del-Solar

Received: 11 December 2010 / Accepted: 30 May 2011 / Published online: 12 July 2011

Abstract In this paper a robust system for enabling robots to detect and identify humans in domestic environments is proposed. Robust human detection is achieved through the use of thermal and visual information sources that are integrated to detect human-candidate objects, which are further processed in order to verify the presence of humans and their identity using face information in the thermal and visual spectrums. Face detection is used to verify the presence of humans, and face recognition to identify them. Active vision mechanisms are employed in order to improve the relative pose of a candidate object/person in case direct identification is not possible. The response of the different modules is characterized, and the proposed system is validated using image databases of real domestic environments, and human detection and identification benchmarks of the RoboCup@Home research community.

Keywords Human detection · Human identification · Service robot · Thermal image

M. Correa · G. Hermosilla · R. Verschae ·
J. Ruiz-del-Solar
Department of Electrical Engineering,
Universidad de Chile, Av. Tupper 2007,
Santiago, Chile

M. Correa
e-mail: macorrea@ing.uchile.cl

G. Hermosilla
e-mail: ghermosi@ing.uchile.cl

J. Ruiz-del-Solar
e-mail: jruizd@ing.uchile.cl

M. Correa · G. Hermosilla · R. Verschae (✉) ·
J. Ruiz-del-Solar
Advanced Mining Technology Center,
Universidad de Chile, Av. Tupper 2007,
Santiago, Chile
e-mail: rodrigo@verschae.org

1 Introduction

There is increasing interest in domestic service robots in the robotics community. A domestic service robot is a subclass of mobile service robots designed to interact with humans in a home-environment, and to provide different kinds of services (cleaning, cooking, entertainment, companionship, and surveillance, to name just a few). The home environment is defined as ’any place where people live their daily lives’, which can include, for example, a kitchen, a bedroom, or a garden. Although some special-purpose domestic robots are already popular (e.g. vacuum robots [35]), we are still far from having general-purpose domestic robots.

Among the basic skills of domestic service robots are the ability to move autonomously in domestic environments, the ability to recognize and manipulate ‘home’ objects (cups, books, glasses, medications, chairs, door handles, etc.), and the capability of identifying humans and interacting with them using intuitive interfaces such as speech, gestures, and facial information. The following analysis will focus on human's detection and identification using visual information, which, from our point of view, is required in domestic robots that need to be general purpose and used by non-expert users.

The robust detection of humans in real-home environments is a challenging task, mainly because of variable illumination conditions, cluttered backgrounds, and variable poses of a human body with respect to the robot's camera. In fact, a human body is a complex and deformable object with several degrees of freedom, whose appearance can change greatly when mapped onto a 2D image. Thus, the problem of the detection of a human body or a human body-part using standard CCD and CMOS cameras that work in the visible spectrum is far from being solved! Depending on the specific circumstances, humans can be detected by using information about their faces, silhouettes, skin, or movement, as well as by using depth information. None of these methods is all-purpose and any of them can fail depending on the specific circumstances. For instance, face and silhouette detection depend on the specific relative pose of humans (e.g. a face can not be detected when the human is observed from the back); skin detection depends largely on the illumination conditions and on the background (e.g. human skin can easily be confused with other materials such as wood); human movement detection depends largely on the illumination conditions and the relative movement of humans (e.g. a human in a static position can not be detected); and human detection using depth information requires further analysis in order to distinguish between human-body parts and other objects.

The robust recognition of humans using visual information [1] is also dependent on environmental conditions such as illumination, cluttered backgrounds, and relative pose of the person to be identified. When restricted to visual interaction, face recognition is the most natural and frequently used clue for identifying people. In [4], four requirements that should be fulfilled by face recognition methods to be used in service robotics applications are identified: (1) *Full online operation*: No training or offline enrollment stages of the face recognition module. All face recognition processes must be run online. The robot has to be able to build the database of faces to be recognized from scratch, and incrementally; (2) *Real-time operation*: The recognition process should be fast enough to allow real-time interaction; the whole face analysis process, which includes detection, alignment and recognition, should take no more than 300 ms (~3 fps), with 200–250 ms being a recommended value; (3) *One single image per person problem*: A two-dimensional face image of an individual should be enough for his/her later identification. Databases containing just one face image per person should be considered. The main reasons are savings in storage and computational costs, and the impossibility of obtaining more than one face image from a given individual in certain situations; and (4) *Unconstrained environments*: It is required that there are no restrictions on environmental conditions such as scale, pose, lighting, focus, resolution, facial expression, accessories, make-up, occlusions, background, and photographic quality. High demanding HRI (Human Robot Interaction) applications—for example, robot interaction with known and unknown people in unconstrained domestic environments—have these requirements. Currently, state-of-the-art methods designed to fulfill these requirements are highly dependent on illumination conditions; for instance, most of them fail when used in mixed indoor-outdoor conditions, and on the face pose [4].

Thermal sensors, however, allow the robust detection of human bodies independently of the illumination conditions (i.e. no light is required) and of the pose (the thermal radiation of a human body can be detected in any pose), and its detection range is up to several meters, which is enough for domestic environments. In addition, humans can also be identified by analyzing their faces in the thermal spectrum [2, 3, 22]. Taking all of these properties into consideration, it seems natural to include thermal cameras in current and future domestic service robots. The price of thermal cameras is no longer a factor for not using them in domestic robots, since the price has fallen significantly in recent years, now being comparable to the price of middle-range laser sensors and time-of-flight cameras, both commonly used in domestic robots. In the current work, the robot is powered with a FLIR TAU 320 thermal camera [37]. This camera has a resolution of 324 × 256 pixels, and it is sensitive in the 7.5–13.5 μm long-wave infrared range.

Given this context, the goal of this work is the proposal of a robust system for robot detection and identification of humans in domestic environments. Robust human detection is achieved through the use of thermal and visual information sources that are integrated to detect human-candidate objects, which are further processed in order to verify the presence of humans and their identity using face information in the thermal and visual spectrums. Face detection is used to verify the presence of humans, and face recognition to identify them. Active vision mechanisms are employed in order to improve the relative pose of a candidate object in case direct identification is not possible, e.g. the object is too far away and the robot must approach it, or the view angle is not appropriate for identifying the human so the robot must find a better view angle.

In conditions of bad or variable illumination, the system relies mainly on the use of thermal information. But, in conditions of good illumination, thermal and visual information complement each other. For instance, visual information allows a better analysis of the textures and a more robust detection of eyes, which is used for face alignment before identification. Thermal information allows an easier differentiation of human bodies and faces in complex backgrounds.

It is important to mention that in the implemented system, state-of-the-art methods for detecting and recognizing faces in the visible and thermal spectrums are used. In the case of face detection in the thermal spectrum, boosted cascade classifiers are used for the first time to solve this problem.

This paper is organized as follows: Related work is described in Section 2. The proposed human detection and identification system for robots is explained in Section 3. Descriptions of experiments and results are presented in Section 4. Finally, conclusions of this work are given in Section 5.

2 Related Work

Research activities in domestic service robotics have increased greatly in recent years. Some of the main drivers of this phenomenon are the projected future use of domestic robots for improving elderly people's quality of life, childcare applications, entertainment and education, and providing specific services such as housekeeping. In addition, interesting initiatives, such as the RoboCup@Home [36], whose aim is to provide benchmark tests and methodologies for evaluating the abilities and performance of domestic service robots in realistic, non-standardized home environment settings, are expected to accelerate and focus technological and scientific progress in the domain of domestic service robots [34].

The robust detection and identification of humans by robots in domestic environments is a challenging open problem. For instance, in the RoboCup@Home, even the best teams are not able to achieve robust human detection and identification in competitions designed to test these kind of abilities (e.g. ‘Who is Who?’ [36]).

There has been a large number of papers in the recent literature that address human detection and identification. In terms of sensor technology, several works are based on the use of stereo vision [27, 30], monocular vision [29, 32], sonar and vision [28], laser and vision [31], and thermal vision [6, 8–10]. For instance, in [27] a stereovision system uses a dense depth image for the detection and tracking of people; [28] uses a sonar in combination with a skin color detector to detect faces; and [31] uses a laser to detect the legs of humans and a vision system to find faces. One of the main benefits of using thermal vision is simplifying the segmentation of human bodies or human body-parts from the background.

In the case of detecting human bodies, the problem that has been studied the most is the problem of pedestrian detection (see [43] for a Survey). Some approaches are based on the use of far-infrared images [6–10]. These approaches use either probabilistic templates [9], warm symmetrical objects of specific size and aspect ratios [10], or temporal filtering (e.g. the Kalman filter) [7]. That a head detector works better than a body detector when using statistical classifiers is shown in [6].

In the case of visual pedestrian detection, recent works have focused on pedestrian detection for in-car pedestrian emergency braking, including: (1) a comparison of using different features (global and local features (PCA coefficients) [41], Haar wavelets, and local receptive fields), and different Classifiers (support vector machines, feed-forward neural networks, and k-nearest neighbor classifiers); (2) the use of a system based on stereo-based ROI generation, shape-based detection, texture-based classification and stereo-based verification [42]; (3) the use of a cascade detection algorithm for a general class of models defined by a grammar, in which the models can represent each part recursively as a mixture of other parts [44]. In terms of feature types, other relevant work include the use of new features such as Histograms of Oriented Gradient (HOG) [45], HOG features of variable-size [49], the use of region covariance matrices [47, 51], the joint use of motion and appearance features [46, 52], and the comparison of different features types [48, 50]. In terms of classifiers, relevant work includes the use of a Viola & Jones [24] like cascade classifiers [46, 48, 49, 51], the use of covariance matrices together with a Riemannian manifold [47, 51], and the comparison of existing methods [50]. In addition, new databases have recently been proposed (e.g. the Caltech pedestrian dataset [53] and the DaimlerChrysler pedestrian dataset [43, 47]) for this particular problem. However, it should be stressed that pedestrian detection is a completely different application than human detection and identification in a domestic environment, and that both applications have different challenges to be met.

For detecting and identifying people, face information is one of the most popularly used clues. Existing work on face detection using machine learning algorithms has been almost exclusively applied to visual images, with little work devoted to the use of thermal images [5]. The best face detection methods are based on the use of machine learning algorithms such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Boosting Classifiers [15, 26]. The most popular face detection paradigm is based on the use of cascades of boosted classifiers, allowing robust and efficient detection of faces [24]. The visible-spectrum and thermal face detectors implemented in this work are based on this paradigm. To the best of our knowledge this kind of detector has not been used before with thermal images, and so we use it for the first time in this work.

Several different face recognition approaches have been developed in the last few years [17–20], ranging from classical Eigenspace-based methods (e.g. Eigenfaces [21]), to sophisticated systems based on high-resolution images and 3-D models. Several methods have been developed for the recognition of faces using thermal images, and most of these methods are based on the same kind of approaches used on visible images [2–5, 11–14, 16]. In [3] a comparison of face-recognition methods using thermal images (long wave infrared, 8–12 μm) is presented. The study considers the previously mentioned HRI requirements of online and real-time operation, one image per person and unconstrained environments, and it focuses on the three methods that obtained the best results in the visible spectrum [4]: Local Binary Pattern (LBP) Histograms, Gabor Jet descriptors, and Scale-Invariant Feature Transform (SIFT) Descriptors. In general terms the results presented in [3] indicate that LBP-based methods are able to obtain very high recognition rates and present computational and memory requirements that are adequate for HRI use. For this reason LBP Histograms are used to implement the visible spectrum and thermal face recognition modules used in the proposed system.

Thus, one of the main contributions of this work is the proposal of a robust system for robot detection and identification of humans in domestic environments, based on state-of-the-art face detection and recognition methods, which work in the visible and thermal spectrums.

3 Human Detection and Identification System

3.1 System Overview

The design of the proposed system for robot detection and identification of humans takes into account the complementary advantages of using thermal and visual information, and it considers

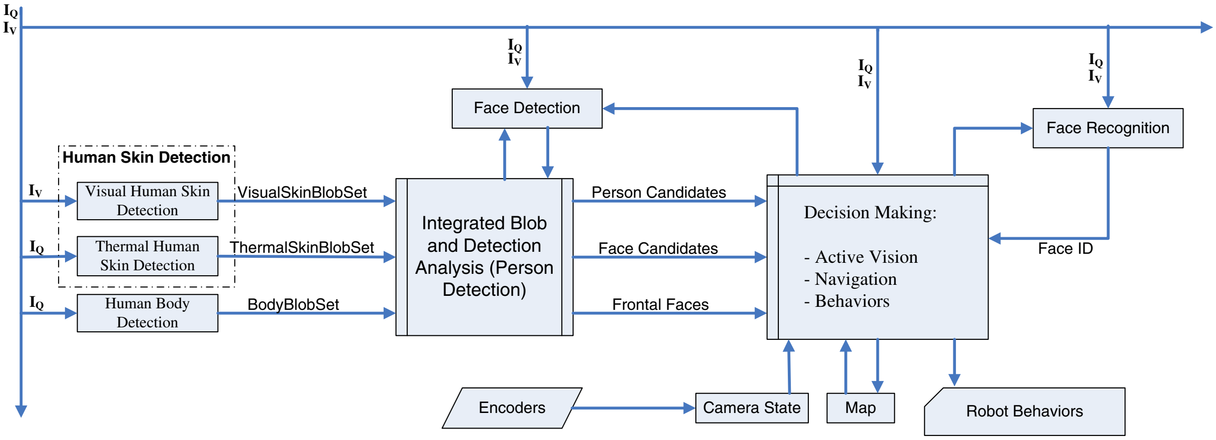

Fig. 1 Block diagram of human detection and identification system. I_V visual image, I_Q thermal image. See text for details

that both cameras have a similar field of view and depth of field. The proposed system fulfills the four requirements mentioned in Section 1. Figure 1 presents the diagram of the proposed system. The main modules can be grouped into the following categories: human skin detection, human body detection, person detection (*Integrated Blob and Detection Analysis* module), face detection, face recognition, and decision-making. Each of these modules works with visual images (I_V in Fig. 1), thermal images (I_Q in Fig. 1), or information extracted from both sources.

First, human skin detection and human body detection are used in order to detect sets of skin blobs in the thermal and visible spectrum (*ThermalSkinBlobSet* and *VisualSkinBlobSet*), and a set of body blobs in the thermal spectrum (*BodyBlobSet*). Then, in the *Integrated Blob and Detection Analysis* module, the information contained in these sets is integrated and analyzed using the *Face Detection* module. The *Integrated Blob and Detection Analysis* module generates person, face and frontal face candidates that are further analyzed in the *Decision Making* module. This last module performs the key task of guiding the search for humans, as well as finding appropriate views of human faces in order to detect and recognize humans with high accuracy, while at the same time minimizing the movements of the robot. Among other tasks, the *Decision Making*

Table 1 List of modules and methods

Module name	Sub module	Output	Method
Human skin detection	Visual human skin detection	VisualSkin BlobSet	Dynamically updated Skindiff algorithm [23]
	Thermal human skin detection	ThermalSkinBlobSet	Mixture of Gaussians (MoG)
Human body detection		BodyBlobSet	MoG + Heuristics
Integrated blob and detection analysis	–	Person Candidates, Face Candidates, and Frontal Faces	Heuristics
Decision making	Active Vision; Navigation; Behaviors	Faces and People position	Heuristics
Face detection	–	Detected Faces (Visible and Thermal)	Nested Cascades of Boosted Classifiers [15]
Face recognition	–	Face ID	Histograms of LBP features [33]

module generates the movement commands to the robot's body and head, and controls the speech interaction with humans; it is in charge of interacting with the face recognition module and navigating through the environment. Table 1 lists the submodules and the methods used in each module. The modules are detailed in the sections below.

Figure 2 presents an example of a visible and thermal image, as well as the output of some modules. It is important to note that all modules work on grey-scale images (thermal or visual), with the exception of the visual skin detection module which works on RGB images.

3.2 Human Skin Detection

The *Visual Human Skin Detection* module determines image regions that contain human skin (in the visible spectrum) using the *Skindiff* skin segmentation algorithm [23]. Skindiff is a fast skin detection algorithm that uses neighborhood information (local spatial context) to achieve robustness. It has two main processing stages, pixel-wise classification and spatial diffusion. The pixel-wise classification uses a non-parametric skin model [25], G_t, implemented using histograms, and the spatial diffusion takes neighborhood information into account when classifying a pixel, and starts from pixels that have a large likelihood of being skin pixels [23]. The skin probability model can be adapted continuously using information of the face-area's pixels (visual skin pixels) detected in previous frames:

$$G_t = G_{t-1}\alpha + \hat{G}_{face(t)}\,(1-\alpha)\,, \tag{1}$$

where $\hat{G}_{face(t)}$ is estimated using the currently detected face, and G_o is the initial model, which can be initialized from a previously stored model, and α a constant. If no face was detected in the previous frame, α is set to one, otherwise $0 < \alpha < 1$. Although our experience shows that updating the skin model with the detected faces greatly improves the results [40], in the current work the system works on individual frames and the model is not updated. The obtained set of visual skin blobs is called *VisualSkinBlobSet.*

The *Thermal Human Skin Detection* module is based on a parametric probability model of the

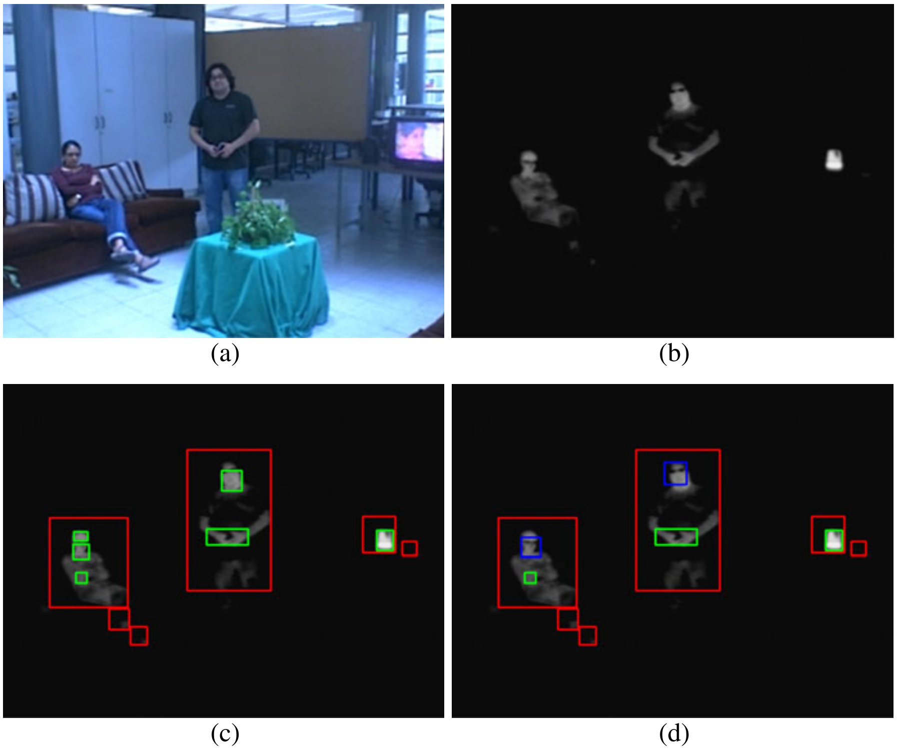

Fig. 2 Output of selected modules: **a** Visible image. **b** Thermal image. **c** Human Skin detection: human body blob in *red* and thermal skin blobs in *green*. **d** Person detection: Person Candidates in *red*, Face Candidates in *green*, and Frontal Faces in *blue*. Some false detections are observed

distribution of temperature of skin. *Mixture of Gaussian* (MoG) models the skin ($P[x_q|skin]$) and non-skin ($P[x_q|non\text{–}skin]$) distributions. A Bayes classifier determines skin pixels as the ones that fulfill the following relationship:

$$r\left(x_q\right) = P\left[x_q|skin\right] / P\left[x_q|non - skin\right] > u_q, \quad (2)$$

with x_q the observed temperature of pixel x, and u_q a previously fixed threshold. Then, a diffusion operation is applied to group the thermal skin pixels into thermal skin blobs. The obtained set of thermal skin blobs is called *ThermalSkinBlobSet*. All parameters of the skin probability model were obtained using a training database.

3.3 Human Body Detection

The *Human Blob Detection* is in charge of detecting human bodies and human-body parts using thermal information. The detection also includes body parts covered by clothes. The same probability ratio $r(x_q)$ used for the detection of thermal human skin is used here, but the threshold is adapted to account for changes in the temperature of the environment, and changes in the response of the camera. This is done by applying a linear mapping to the values of $r(x_q)$—taking the maximum and minimum observed values and mapping them to 0 and 1—and by using a fixed decision threshold in this range. This allows adapting for situations where there is a large difference in the temperature of the bodies (clothes) and faces; the body and clothes temperature can vary greatly from summer to winter. After each pixel has been classified, an opening morphological operation is applied to fill holes that appear in the body, but seldom appear in face regions. The set of (thermal) human body blobs is called *BodyBlobSet*.

It is important to clarify that thermal human skin detection allows segmenting the human skin regions that are not covered by clothes, such as arms, faces, hands, and legs, while the (thermal) human body detection allows detecting larger body parts, which might be partly covered by hair or clothes.

3.4 Face Detection

Human face detection is based on the use of a state-of-the-art multi-scale object detection framework (see the block diagram in Fig. 3) previously developed by our group [15], which uses boosted cascade classifiers. The same framework is used to build face detectors able to detect faces in the visible and in the thermal spectrums. However, although both kinds of detectors share the same structure, the process required for building each detector is different, mainly because of the use of different training images in each case. To the best of our knowledge, statistical classifiers, and in particular boosted cascade classifiers, have not been used for detecting human faces in thermal images to date.

The detection works as follows: First, to detect the face at different scales, a multi-resolution analysis of the images is performed by downscaling the input image by a fixed scaling factor—e.g. 1.2—(*Multi-resolution Analysis* module). Afterwards, windows of 24 × 24 pixels are extracted in the *Window Extraction* module for each of the scaled versions of the input image. Then, the windows are analyzed by the nested cascade classifier (*Cascade Classification Module*). Finally, in the *Final Detections* module, the windows classified as positive (i.e. containing a face) are fused (normally a face will be detected at different scales and positions) to obtain the size and position of the final detections.

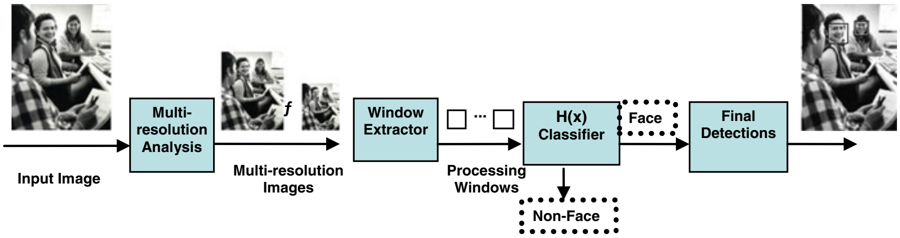

Fig. 3 Block diagram of the face detection framework

The key concepts used in the framework are nested cascades, boosting, and domain partitioning classifiers. Cascade classifiers consist of several layers (stages) of classifiers of increasing complexity for obtaining fast processing speed together with high accuracy. The main idea of cascade classifiers is to process most non-object windows as fast as possible, and to process the object windows and the object-like windows carefully. Real Adaboost is employed to find and combine several weak hypotheses, and for feature selection. Nested cascades allow higher classification accuracy and processing speed by reusing in each layer the confidence given by its predecessor, and the cascade is composed of several integrated (nested) layers, each one containing a boosted classifier. A nested cascade, composed of M layers, is defined as the union of M boosted classifiers H_C^k each one defined by:

$$H_C^k(x) = H_C^{k-1}(x) + \sum_{t=1}^{T_k} h_t^k(x) - b_k \tag{3}$$

with $H_C^0(x) = 0$, h_t^k the weak classifiers, T_k the number of weak classifiers in layer k, and b_k a threshold (bias) value that defines the operation point of the strong classifier. The output of H_C^k is a real value that corresponds to the confidence of the classifier, and its computation makes use of the already evaluated confidence value of the previous layer of the cascade, and the class assigned corresponds to its sign. We use domain partitioning weak hypotheses, each one giving self-rated confidence values that estimate the reliability of each prediction. The weak classifiers prediction depends only on which block a given sample instance falls into for a given feature:

$$h(f(x)) = c_j \ni f(x) \in F_j \tag{4}$$

For each classifier, the value is associated with each partition block (c_j). The outputs, c_j from each of the weak classifiers, obtained during training, are stored in an LUT to speed up its evaluation.

For the training and validation of the face detectors the following datasets were used:

- Visible spectrum. Training set: 5,000 frontal face images and 3,500 non-face images. Validation set: 5,000 frontal face images and 1,500 non-face images. The images were obtained from many different sources, and all of them were made under real world conditions, including variations in illumination conditions, backgrounds, races, etc.
- Thermal spectrum: Training set: 20,000 frontal face images (generated from 800 frontal face images) and 15,000 non-face images (generated from 200 images not containing faces). Validation set: 20,000 face images and 15,000 non-face images. The thermal images were obtained using a similar camera to the one used in this work.

The training procedures are described in [15].

3.5 Face Recognition

3.5.1 Recognition of Human Faces in Visible-Spectrum Images

A comparative study of state-of-the-art face recognition methods that are suitable to work in unconstrained environments using visible information is presented in [4]. The analyzed methods were selected by taking into account their performance in former studies, in addition to being real-time, having just one image per person, and being fully online. The evaluation considered real-world conditions that included variations in scale, pose, lighting, focus, resolution, facial expression, accessories, make-up, occlusions, background and photographic quality. One of the main conclusions of that study is that the use of the histograms of LBP features methodology is an excellent choice when real-time operation and high recognition rates are required, and the faces are captured in uncontrolled environments. For these reasons, this methodology was selected to implement face recognition in visible images in our system.

Face recognition using histograms of LBP features was originally proposed in [33], and has been used by many groups since then. In the original approach, three different levels of locality are defined: pixel level, regional level, and holistic level. The first two levels of locality are realized by dividing the face image into small regions from which LBP features are extracted, and histograms are used for efficient texture information representation. The holistic level of locality, i.e. the global description of the face, is obtained by concatenating the regional LBP extracted features. The recognition is performed using a nearest neighbor classifier in the computed feature space using one of the three following similarity measures: histogram intersection, log-likelihood statistic, or Chi-square.

We implemented this recognition system without considering preprocessing (cropping using an elliptical mask and histogram equalization are used in [33]), and by choosing the following parameters: (1) images divided in 80 (4 × 20) regions, instead of using the original divisions which range from 16 (4 × 4) to 256 (16 × 16), and (2) histogram intersection as a similarity measure. Before recognition, faces images are aligned using an eye detector built using the previously mentioned object detection paradigm. The eye detector is described in [15].

3.5.2 Recognition of Human Faces in Thermal Images

In [3], a comparative study of state-of-the-art face recognition methods for HRI applications using thermal images was presented. The results obtained in that study also show that the Histograms of LBP features methodology is robust and efficient in the recognition of faces in the thermal spectrum. Therefore, the LBP-histogram methodology was selected to implement face recognition in thermal images in our system.

As in the case of the visible-spectrum images, the images are divided into 80 regions, and histogram intersection is used as the similarity measure. However, faces are not aligned before recognition, mainly because eye detection in thermal images is more inaccurate than in visual images [5]. Non-aligned images can be used because LBP Histograms can handle inaccurate alignment [4], which was an additional reason to select this method. In the case of the visual face recognition module, the faces were aligned using an eye detector because a more accurate alignment is needed in order to remove the background that could affect the recognition process considerably in unconstrained environments.

3.6 Integrated Blob and Detection Analysis (Person Detection)

Person candidates, face candidates, and frontal faces are determined by integrating the information contained in the sets of body and skin blobs, and using the face detection module. The following procedure is used: (1) First, all body blobs are selected as *Person Candidates*. (2) Then, face detection is applied inside each body blob. All detected frontal faces that are inside a body blob are marked as frontal faces (a frontal face detector is used), and the corresponding body blob is marked as containing a face. The body blobs that do not contain a face are marked as person candidates that are not facing the camera. (3) Finally, all skin blobs that do not overlap with a detected frontal face are marked as face candidates.

3.7 Decision Making

The decision-making module is in charge of actively searching for humans and human-faces, in addition to HRI, which will not be described here. Given a map M and a set of map positions P_i, $i = 1,..., N$ to be visited, the procedure followed by the robot to search for human candidates inside the map is as follows:

1. The robot moves to P_1.
2. The robot actively searches for human candidates by looking in three directions (0°, 45° and −45°) by moving its head. The obtained visual and thermal images are analyzed by the *Human Skin Detection*, *Human Body Detection,* and *Integrated Blob and Detection Analysis* modules, and two disjoint sets are obtained: B(bodies) and F(faces). B contains the detected objects that were classified as person candidates and do not contain any detected frontal faces, and F contains the detected frontal faces and frontal face candidates (found in either visual or thermal images). The largest object is selected from set B. This object will be called MB, and it corresponds to the main body candidate.
3. For each element in F and for MB, the distance from the robot to the object is estimated using the face information in the case of the

elements belonging to set F, and the blob's width in the case of object MB. The width of MB is considered to be a good estimation of the width of the torso of a person. The position of each object/person represented by the elements in F or by MB is stored in the map only if this position is within the map. If MB's width is small, such as for lamps, computers, and other objects, its position is considered to be outside the room, and thus discarded. Thus no additional rules are needed to remove small-size false positives.

4. The robot tries to reach each unrecognized person in the map by starting with the nearest person. To do this, the robot approaches the person and speaks, asking the person to look at the robot's face. If the robot cannot detect a face, it makes a local search by moving the head towards four different directions (top-right, top-left, bottom-right, bottom-left), and tries to detect faces in each of these positions. If the robot cannot detect a face, the object is removed from the map and the robot continues visiting the remaining candidates. If a person is detected and its face is recognized or set as unknown, the person's position in the map is set as visited. After all objects in the map have been visited, the robot moves to P_{i+1}, and steps (2) to (4) are repeated.

4 Social Robotics Platform

The proposed system for human detection and identification has been incorporated into *Bender*, a domestic service robot. One of the most interesting features of Bender is its ability to interact with humans using human-like modalities (face, hand gestures, speech, facial expressions, etc.)

4.1 Hardware Components

The main hardware components of the robot are: (see Fig. 4)

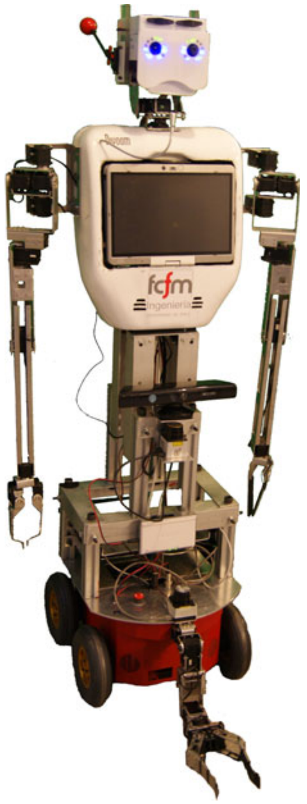
Fig. 4 Picture of the Bender robot

- *Chest* The robot's chest incorporates a tablet PC as the main processing platform, an HP 2710p, powered with a 1.2 GHz Intel Core 2 Duo with 2 GB DDR II 667 MHz, and running Windows XP Tablet PC edition. The tablet includes 802.11 bg connectivity. The screen of the tablet PC allows: (1) the visualization of relevant information for the user (a web browser, images, videos, etc.), and (2) entering data thanks to the touch-screen capability.
- *Head* The robot's head incorporates two CCD cameras (Philips ToUCam III-SPC900NC), pan-tilt movement of the whole head, and the capability of expressing emotions. This is achieved by several servomotors that move the mouth, eyebrows, and the antenna-like ears, and RGB LEDs placed around each eye. In addition, it has RGB LEDs in the forehead to simulate the robot's breathing. The head movements and expressions are controlled using dedicated hardware (PIC18F4550-based), which communicates with the Tablet PC via USB. The cameras are connected to the Table PC using USB ports. The head's weight is about 1.6 Kg.

- *Thermal Vision* The thermal camera is a FLIR 320 TAU Thermal Camera [37], with sensitivity in the range 7.5–13.5 µm (long wave thermal range) and a resolution of 324 × 256 pixels. It has a full frame rate of 30 Hz (NTSC) and 25 Hz (PAL), the sensitivity is lower than 75 mK, and the scene range is from −40°C to +600°C. We use a 9 mm lens with 48° × 37°FOV. The camera is placed in the robot head and is calibrated manually (contrast and brightness) at each session in order to improve the contrast between human-body parts and other objects and facilitate the detection of people.
- *3D Vision* The robot is powered with a PMD CamCub2.0 TOF (Time-Of-Flight) camera [38]. The camera, with a resolution of 204 × 204 pixels, is placed in the robot chest, and used for object detection while grasping.
- *Arms* The two arms of the robot are designed to allow the robot to manipulate objects. They are strong enough for raising a large glass of water or a cup of coffee. Each arm has six degrees of freedom, two in the shoulder, two in the elbow, one for the wrist, and one for the gripper. The actuators are eight servomotors (six RX-64 and two RX-28). The arms are controlled directly from the Tablet PC via USB. The arm's weight is about 1.8 Kg.
- *Mobile Platform* All described structures are mounted on a mobile platform. The platform is a Pioneer 3-AT, which has four wheels, provides skid-steer mobility, and is connected to a Hokuyo URG-04LX laser for sensing. This platform is endowed with a Hitachi H8S microprocessor. A Tablet PC (HP tc4200) is placed on the top of the mobile platform with the task of running the navigation software. This Tablet PC is connected to the chest Tablet PC by means of an Ethernet cable.

4.2 Software Architecture

The main components of the software architecture are shown in Fig. 5. Speech synthesis and analysis, as well as vision tasks (general object recognition, face, hand and gesture recognition), take place in the Tablet PC HP 2710p (running Windows XP Tablet PC edition), while the Navigation and Mapping Modules reside in the Tablet PC HP tc4200 (running Linux), and the low-level control modules run in dedicated hardware (head and arm control). Both Tablet PCs are connected using URBI (see Fig. 5). All the modules running in

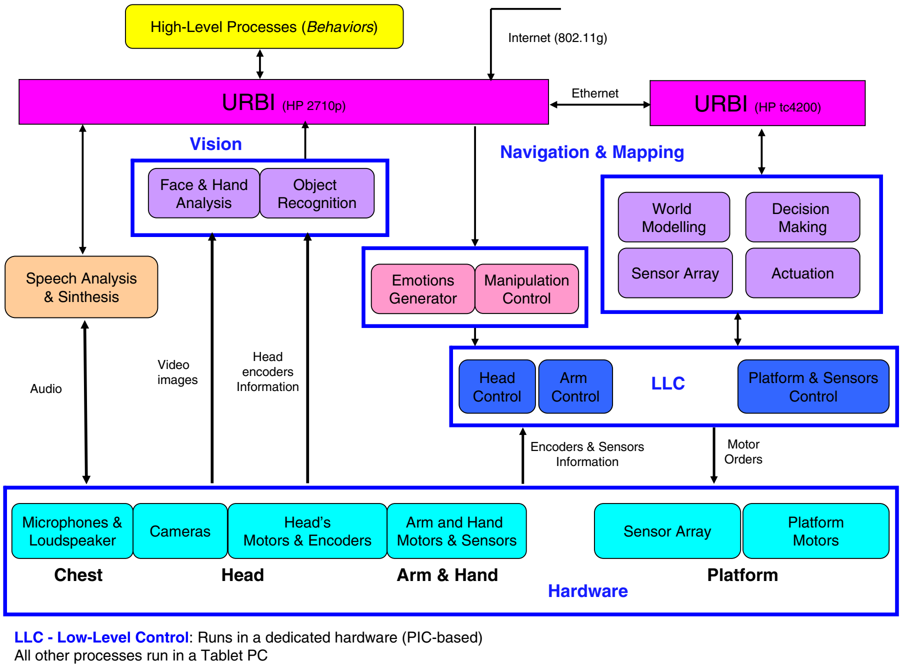

Fig. 5 Modular organization of Bender's software. The HP tc4200 runs the Navigation and Mapping and the rest of the high level processes run in the HP 2710p

the HP 2710p are controlled through URBI using UObjects. The Navigation and Mapping Modules are implemented using the CARMEN Navigation Toolkit [39], which provides localization, simulation, collision avoidance and logging, among other functionalities. CARMEN has the added advantage of being open source and providing specific support for the hardware at our disposal: while the "pioneer" module sends movement commands to the Pioneer base and reads odometry information, the "laser" module reads data from the Hokuyo laser. CARMEN allows interfaces with other programs, thus enabling high-level processes (behaviors) to send commands.

The different software modules are explained in [40].

5 Experiment Results

The ability of the system to detect and identify humans in real domestic environments using a standard benchmark of RoboCup@Home is presented. In addition, the ability of single modules for detecting human-bodies and human-faces is carried out using image databases created in domestic environments under variable illumination and view conditions. The detection results are presented in terms of Detection Rate (DR), True Positives (TP), and Number of False Positives (FP). A detection is considered correct if the detection window and the ground-truth window overlap each other in at least 50% of their areas, otherwise the detection is considered to be a false positive detection.

5.1 Evaluation Databases

Four different image databases were generated in order to evaluate the different modules, and to compare the use of visual and thermal information for human-body and human-face detection in domestic environments. The databases were captured using a Philips ToUCam III-SPC900NC visual camera and a FLIR TAU 320 thermal camera. These two cameras are the same ones used in our domestic robot Bender (see description in Section 4). The visual and thermal images were captured simultaneously, with the cameras 2 cm apart, and at a height of 120/160 cm depending on the experiment. Figure 6 presents some examples of images captured with this camera configuration.

The following databases were created:[1]

- *Illumination Database* Goal: To verify how much the kind of illumination affects the different modules when using visible and thermal images, and to test the skin detection modules and how they are affected by lighting conditions. Setup: 16 people, five images per illumination and per person. The distance between the camera and the observed person was varied from 1 to 3 m with a step of 50 cm. Three subsets were created, each one corresponding to different illumination conditions:
    - *Indoor light*, which corresponds to the standard illumination conditions of indoor rooms, incandescent and natural light
    - *Lamp light*, which case a floodlight lamp was placed 1 m behind the camera.
    - *Night light*, case where the images were taken without any artificial indoor illumination, only natural light was used.
- *Rotation Database* Goal: To evaluate how much the illumination affects human and face detection, to test whether it is possible to detect blobs of the person even if the camera is facing the back of the person, and to evaluate the maximum rotation where the visible skin detection works. Setup: 18 people, five images with two kinds of illumination per person (150 images). 15 different yaw rotation angles were considered: 0°, 5°, 10°, 15°, 20°, 25°, 30°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, 180° (clockwise). The images were captured with the camera placed at a fixed distance of 2 m from the subject. The kinds of illumination are indoor light, and a direct lamp located with an orientation of 90° with respect to the subject.
- *Distance Database* Goal: To evaluate how the visual and thermal face detectors are affected by the size of the face in the image, i.e. to characterize the response of the detectors to

[1] These databases are available at http://vision.die.uchile.cl

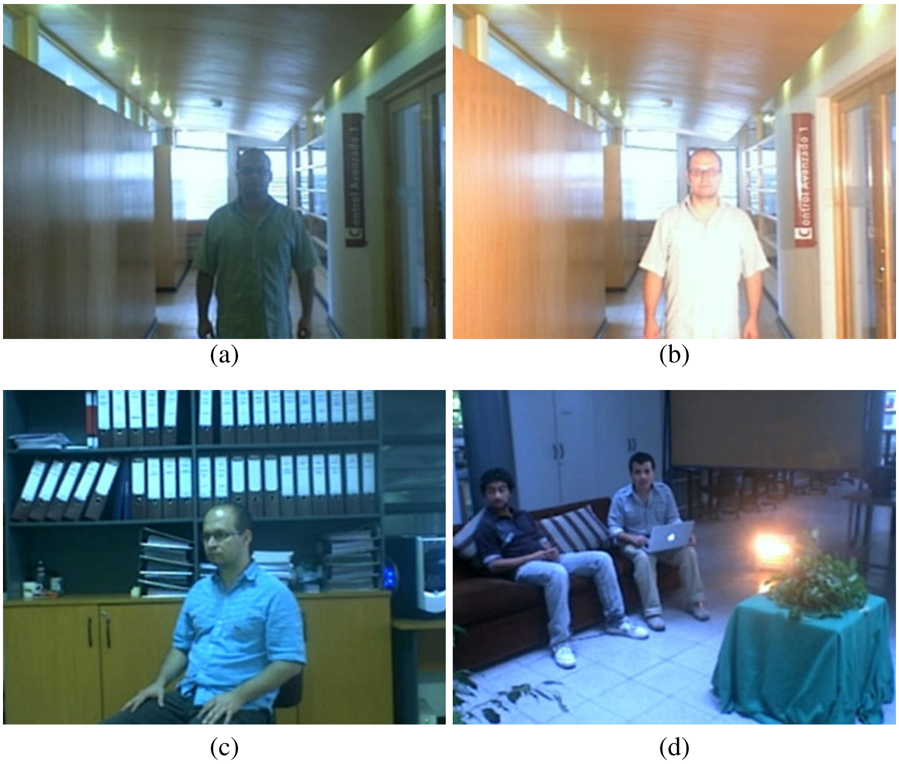
(a) (b) (c) (d)

Fig. 6 Examples of test images: **a** indoor light, illumination set, **b** lamp light, illumination set, **c** indoor light, rotation set, **d** arena set

different distances. Setup: 18 people, 11 images per person taken with steps of 50 cm, going from 1 to 6 m.

- *Arena Database* Goal: To evaluate the human-body and human-face detection abilities of different modules in real domestic environments. Description: 101 images containing 171 people and 104 faces, of which 37 are frontal faces. This database was built in a home environment with humans involved in real life situations. This database contains humans walking, sitting, and talking to each other, as well as humans lying on the floor. Unlike the other two databases, the number of humans per image is variable (from 0 to 4). This database also contains hot devices such as heaters, which are included for evaluating the capability of thermal-based detection modules for discriminating between this kind of device and human bodies.

5.2 Evaluation of Single Modules

Three kinds of evaluations are performed: human skin detection (Tables 2 and 3), human body detection (Tables 4 and 5), and face detection

Table 2 Thermal and visual human skin detection using the *illumination* DB

Distance	Indoor light				Lamp light				Night light			
	Thermal		Visual		Thermal		Visual		Thermal		Visual	
[mt]	DR %	FP	DR %	FP	DR %	FP	DR %	FP	DR %	FP	DR %	FP
1	96.77	50	6.45	212	96.77	49	51.61	133	90.32	43	0	135
1.5	85.71	25	7.14	232	81.82	30	15.91	151	75	33	0	103
2	80.44	14	6.52	257	78.26	18	0	166	78.26	13	0	289
2.5	68.75	11	4.17	255	72.92	9	2.08	170	68.75	8	0	108
3	63.83	5	14.89	289	66.67	6	2.08	162	64.58	4	0	107

DR detection rate (out of 16 Subjects); *FP* false positives (for the complete DB)

Table 3 Thermal and visual human skin detection using the *rotation* DB

Rotation	Indoor light				Lamp light			
	Thermal		Visual		Thermal		Visual	
[degrees]	DR %	FP	DR %	FP	DR %	FP	DR %	FP
0	96.55	22	65.52	68	98.25	24	12.28	61
5	96.61	22	71.19	66	98.28	22	10.34	63
10	96.61	29	62.71	65	98.28	27	12.07	49
15	96.61	28	59.32	66	98.28	32	8.62	52
20	96.61	27	62.71	66	96.61	29	8.47	49
25	98.28	27	58.62	66	96.67	26	6.67	48
30	98.31	29	55.93	70	98.31	29	1.69	42
45	96.61	29	52.54	75	96.55	25	5.17	47
60	98.21	22	41.07	72	98.25	25	5.26	53
75	89.47	28	36.84	71	87.72	28	7.02	53
90	89.80	36	26.53	71	88.00	30	10.00	42
105	89.13	34	17.39	70	90.91	32	4.55	30
120	94.87	37	12.82	68	94.74	34	2.63	31
135	97.22	35	11.11	63	94.44	41	0.00	26
180	74.19	22	16.13	57	68.75	31	0.00	27

DR detection rate (out of 18 Subjects); *FP* false positives (for the complete DB)

(Table 6). For these evaluations, the databases described in Section 5.1 are used.

5.2.1 Human Skin Detection

The detection of human skin carried out by the thermal and visible human skin detectors is evaluated using the two data sets (*Illumination* and *Rotation*) described in Section 5.1. The obtained results, displayed in Tables 2 and 3, clearly show that the thermal skin detector is not dependent on the illumination conditions, while the visible skin detector is highly dependent on the illumination conditions, giving good results only when there is enough (and non-saturated) light, and backgrounds that do not contain skin-like colors. The thermal skin detector works better at 2 m from the subjects, having a smaller number of detections (both true positives and false positives) when the subject is far away from the camera, and a smaller number of detections (both true and false positives) when the subject is closer to the camera, as can be observed in Table 2. The visible skin detector does not show different performances for different distances (*Illumination*

Table 4 Thermal human body detection using the *illumination* DB

Distance	Indoor light		Lamp light		Night light	
[mt]	DR %	FP	DR %	FP	DR %	FP
1	100	44	100	47	100	46
1.5	100	57	100	56	100	54
2	100	56	100	56	100	56
2.5	100	56	100	57	100	56
3	100	53	100	54	100	53

DR detection rate (out of 16 Subjects); *FP* false positives (for the complete DB)

Table 5 Thermal human body detection using the *rotation* DB

Rotation	Indoor light		Lamp light	
[degrees]	DR %	FP	DR %	FP
0	100	30	100	29
5	100	31	100	28
10	100	31	100	27
15	100	32	100	32
20	94.44	32	94.44	33
25	94.44	32	100	32
30	94.44	32	100	32
45	94.44	31	100	29
60	100	30	100	30
75	100	28	100	27
90	100	26	100	27
105	100	27	100	25
120	100	24	100	24
135	100	25	100	24
180	100	27	100	24

DR detection rate (out of 18 Subjects); *FP* false positives (for the complete DB) database), and has very bad performance in all subsets of the database. The main reason is the large number of skin-like colors in the background of the database images, which is a common issue in most domestic environments containing wood and other material with skin-like colors (see Fig. 6 for an example). It is important to mention that the thermal human skin detector is able to work with night light (almost no illumination), which is a very important capability for a domestic robot.

It can be observed in Table 3 that the thermal skin detector is very robust, and responds rather well for different rotations, being able to locate skin areas even if the person is not facing the camera. The visible skin detector has about three times more false positives than the thermal skin detector, and the detection of skin parts decreases with the rotation angle. As in the illumination dataset, in the rotation dataset the thermal skin detector is invariant to the illumination conditions and the visible skin detector has a lower detection rate when a (strong) lamp is used.

5.2.2 Human Body Detection

It can be seen in Tables 3 and 4 that the implemented human body detector is very robust and able to detect human bodies and human-body parts under different illumination conditions, and view angles in the two databases (*Illumination* and *Rotation*) described in Section 5.1. Detection in night light conditions works well also.

Nevertheless, it should be stressed that the system has few false positives (approximately 1.6 per each image), and that human body candidates need to be further analyzed (e.g. using a face detector) in case decisions about the presence of humans need to be made.

5.2.3 Frontal Face Detection

As previously mentioned (Section 3.4), we implemented two frontal face detectors, one for thermal images and one for visual images. Table 6 presents the obtained face detection results in the *Distance* database described in Section 5.1. As can be observed, the detectors work robustly for faces that are close to the camera (<3 m). It is important to stress that at these distances the thermal face detector has a higher detection rate than the visual face detector, and that it is able to detect more than one face at the same time. Thus, our hypothesis is confirmed that a thermal face detector based on the use of cascades of boosted classifiers, as implemented in this work for the first time, is able to detect human faces robustly.

Table 6 Thermal and visual frontal face detection in the *distance* DB

Distance	Thermal		Visual	
[mt]	DR %	FP	DR %	FP
1	100	0	100	1
1.5	100	0	100	2
2	100	1	88.89	0
2.5	100	1	77.78	1
3	83.33	1	66.67	0
3.5	55.56	1	77.78	1
4	38.89	1	83.33	1
4.5	27.78	0	77.78	0
5	16.67	0	83.33	0
5.5	16.67	0	55.56	0
6	0	0	16.67	0

DR detection rate (out of 16 Subjects); *FP* false positives (for the complete DB)

For both detectors, the detection rate decreases when the faces move away from the camera, but the decrease is much greater for the thermal face detector. This is caused by the resolution of the used thermal camera (324 × 256) being half that of the visual camera (640 × 480). Thus, the implemented thermal frontal face detector is not capable of robustly detecting frontal faces that are more than 3 m away from the camera. (This result is also observed in Section 5.3).

5.3 Person Detection in Domestic Environments

Table 7 presents detection results for humans, faces and frontal-faces in the *Arena* database (described in Section 5.1) using five different methods:

- *Visual Frontal Face Detector:*
  The frontal faces detected in the visible spectrum are the only information used to detect humans. This is one of the standard methods

used for detecting humans in an unknown scene.

- *Thermal Frontal Face Detector:* The frontal faces detected in the thermal spectrum are the only information used to detect humans.
- *Human Body Detector:* The human blobs detected, using the thermal human detector (see Section 3.3), are used for detecting humans.
- *Thermal Skin Blob Detector:* Skin blobs detected, using the thermal human skin detector (see Section 3.2), are used for detecting faces.
- *Person Detector using the visual face detector:* People and faces are detected using the *Integrated Blob and Detection Analysis* module described in Section 3.6. In this case a visual face detector is used.
- *Person Detector using the thermal face detector:* People and faces are detected using the *Integrated Blob and Detection Analysis* module described in Section 3.6. In this case a thermal face detector is used.

Note that in the *Arena* database most faces are far from the camera, which makes results for the thermal face detector have a very low detection rate, because of their low resolution. (Recall that the thermal camera has a resolution of only 320 × 256 pixels). This does not happen with larger faces, and in the 'Who is Who?' benchmark (Section 5.4) the thermal frontal face detector works well.

It can be observed in Table 7 that using the *Frontal Face Detector* (either thermal or visible) alone is not enough for detecting humans. The *Frontal Face Detector* for visible images can detect frontal faces robustly (83.78%) with a very low false positive rate (only 25 false positives in 101 images), but it can detect only 51.91% of the total number of faces (frontal and non-frontal) in a home environment.

The *Human Body Detector* can detect most of the humans, but it has a large number of false positives. It cannot detect faces. On the other hand, the *Thermal Skin Blob Detector*, which is not able to detect humans, can find all faces, but has a very large number of false positives.

The *Person Detector* can reduce the number of false positives considerably (to less than one-third) when detecting faces, and it can reduce the number of false positives considerably (to one-fifth) when detecting frontal faces, without diminishing the detection rate.

These results show the robustness of the proposed system for detecting people, faces and frontal faces. The proposed *Person Detector* can detect all humans, ~50% of face candidates, and ~83% of the frontal faces with a relatively low number of false positives, particularly in the case of frontal faces. The large number of false positives when detecting humans can be reduced by discarding false positives based on the blob size (see, for example, the small red blobs in Fig. 2d). These results demonstrate that the proposed system appropriately solves the robot detection of humans problem in domestic environments.

5.4 'Who is Who?' Benchmark

'Who is Who?' is one of the standard benchmarks used in the RoboCup@Home competitions. The main goal of the benchmark is to test the ability of domestic robots "to autonomously detect and recognize people in an unknown environment" [36]. In order to accomplish this task it is expected that "without manual calibration, a robot will have to introduce itself to a group of people, ask for their names, memorize them and recognize the people when meeting them again". The test focuses "on human detection/recognition, face detection/recognition, safe navigation and human-robot interaction with unknown people" [36]. Basically the test is as follows: The robot enters the arena through the door and stops next to it. Two people enter through the door and introduce themselves to the robot, one by one. The robot asks for their names and memorizes them. When told to do so by an operator, the robot goes to the room and starts looking for guests. In the room, there are two other people who are unknown to the robot. One of them is sitting and the other one is standing. There is also one person standing in the room who is known to the robot. When the robot finds a person, it has to approach it and say that it has found a person. Then it has to recognize the person by saying its name or state that the person is unknown. The distance from the robot to the person must not exceed 1 m.

Table 7 Results of human detection, face detection and frontal face detection in the *arena* database

Method	Human detection		Face detection		Frontal face detection	
	DR %	FP	DR %	FP	DR %	FP
Visual frontal face detector	35.67	25	51.92	25	83.78	25
Thermal frontal face detector	3.51	5	5.77	5	16.22	5
Human body detector	99.42	241	–	–	–	–
Thermal skin blob detector	–	–	100.00	499	–	–
Person detector using the visual face detector	99.42	241	51.92	7	83.78	5
Person detector using the thermal face detector	99.42	241	7.69	1	16.22	1

There are 101 images in total containing 171 Humans, 104 faces and 37 frontal faces
DR detection rate; *FP* false positives (for the complete DB)

The following human detection and identification systems were tested in this benchmark:

- *Full Visible:* Only the visible face detector is used to find people in the arena. A face recognition system working on visible images is used to identify them. This is the standard approach used by most teams participating in the RoboCup@Home world competitions.
- *Full Thermal:* The *Person Detector* using thermal face detection, as tested and presented in the previous sections, is used to find people in the arena. A face recognition system working on thermal images is used to identify them.
- *Hybrid Thermal and Visible: The Person Detector* using thermal face detection, as tested and presented in the previous sections, is used to find people in the arena. A face recognition system working on visible images is used to identify them.

Table 8 presents the results of the evaluation of the 'Who is Who?' benchmark. In each scene there were five people, of whom three were known to the robot. Two people were standing with their faces looking towards the outside of the arena. Thus the frontal face detector was not able to detect those two faces. The experiment was run three times using each method mentioned above, and in each run the people known to the robot were the same, and their position in the room was not changed. The detectors and recognition modules used correspond to the ones described in Section 3 and characterized in Section 4.

The obtained results show that the *Full Visible* system works well when detecting faces (it detected 88.9% of the frontal faces), and the visible

Table 8 'Who is Who?' evaluation

Method		Face detection		Person detection		Face recognition		
		DR % | TP	FP	DR % | TP	FP	Correct	Incorrect	Missed
Full visible	Run 1	66.7 | 2	0	40 | 2	0	2	0	3
	Run 2	100 | 3	0	60 | 3	0	2	1	2
	Run 3	100 | 3	0	60 | 3	0	3	0	2
	Avg.	88.9 | 2.7	0	53.3 | 2.7	0	2.7	0.3	2.7
Full thermal	Run 1	100 | 3	0	80 | 4	0	4	0	1
	Run 2	100 | 3	0	100 | 5	0	3	2	0
	Run 3	66.67 | 2	0	80 | 4	0	3	1	1
	Avg.	88.9 | 2.7	0	86.6 | 4.3	0	3.3	1	0.7
Hybrid thermal and visible	Run 1	66.7 | 2	0	80 | 4	0	3	1	1
	Run 2	100 | 3	0	80 | 4	0	4	0	1
	Run 3	100 | 3	0	100 | 5	0	4	1	0
	Avg.	88.9 | 2.7	0	86.6 | 4.3	0	3.7	0.7	0.7

Of the five humans presented in the image, two of them were standing in positions where a frontal face detector would not work (see Fig. 7). All methods were run three times each. Best average results shown in bold letters. See text for details
DR detection rate; *FP* false positives (for each run)

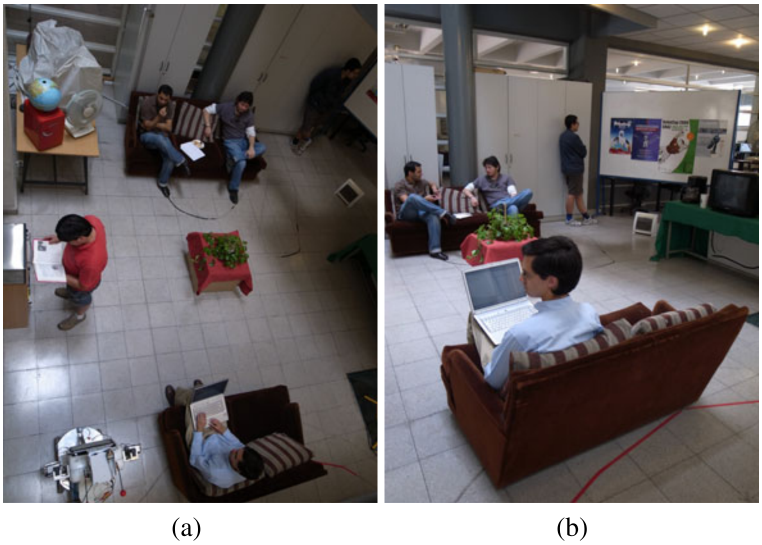
(a) (b)

Fig. 7 Examples of the 'Who is Who?' experiment run. **a** View from the top of the home environment. **b** Robot's view

face recognition method has a high recognition rate (on average it correctly recognized 2.7 of the 3 detected people), while mistaking them in only one of nine cases. Nevertheless it cannot detect people whose faces are not looking in a direction where the robot can detect them.

The obtained results for the *Full Thermal* system show that it is very good for detecting people (it detected 86.6% of the people with 0 false positives per run), it gave good results when detecting and verifying the frontal faces in thermal images (it detected 88.9% of the frontal faces presented), and has a good recognition rate (on average it recognizes correctly 3.3 of 4 detected people), but it recognizes wrongly in three of 12 cases.

The *Hybrid Thermal and Visible* has a similar performance than the *Full Thermal* system for people and face detection, but it improves the recognition rate (on average it recognizes correctly 3.7 of 4 detected people), and it recognizes wrongly in two of 12 cases.

The obtained results show that the joint use of thermal and visible information allows achieving high human detection and high human recognition rates at the same time.

One of the main problems observed with the thermal camera is that its response is dependent on the length of time the camera has been on, considerably affecting the recognition rate, because images become saturated and the thresholds need to be changed accordingly. Another problem is that people's body temperatures could vary widely (e.g. perspiration). These problems will be addressed by us in future work.

6 Conclusions

The development of robots for domestic environments is a challenging task. One of the most basic problems is how to enable them to detect and identify humans robustly. In the present manuscript, a system to solve this problem is proposed. Thermal and visual information sources are used for detecting humans, locating their faces, and recognizing them robustly. An integrated analysis is performed to detect human-candidate objects, and to process them further in order to verify the presence of humans and their identity. A face detector is used to verify the presence of humans, and a face recognition system is used to identify them. In case direct identification is not possible, an active vision search mechanism is employed to improve the relative pose of a candidate object/person. The response of the different proposed modules is characterized, and the proposed system is validated using image databases of real domestic environments, and a human detection and identification benchmark of the RoboCup@Home research community.

The reported results demonstrate that the proposed system solves the robot human detection and identification problem in domestic environments appropriately. Thermal skin detection and thermal human detection are robust under variable illumination and view angles, and allow detecting human bodies and human body parts at appropriate distances for domestic applications (~6 m). The experiments also confirmed our hypothesis that a Thermal Face Detector based on the use of a cascade of boosted classifiers, and implemented in this work for the first time, is able to detect human faces robustly.

The use of a thermal camera allows robots to work under difficult illumination conditions (low illumination, uneven illumination, illumination from different sources), and to detect humans that are far from the camera with higher accuracy, while the use of a visual camera allows work with un-calibrated images, in environments with many warm objects, with objects that have textures or textured appearance, and a wide range of objects because of the availability of a larger number of databases for training detectors or classifiers.

The background of the images can be quite bothersome when trying to detect faces or objects using normal cameras; this problem can be solved easily by using a thermal camera. Besides, since the information given by the thermal system is complementary to the information provided by the visible system, the false detections generated by the thermal system can be removed by the visible system and vice versa. The problem of skin detection can be solved much more easily in the thermal spectrum than in the visible spectrum.

It is important to mention that the proposed human detection and identification system is able to work with night light (almost no illumination), thanks to the use of thermal images. This is a very important capability for general-purpose domestic robots, which should be able to take care of home tasks (e.g. surveillance, elderly care) during both the day and the night.

As previously mentioned, one of the main problems observed in the use of thermal cameras is that their response is dependent on the length of time the camera is on, affecting the recognition rate, because images become saturated and the thresholds need to be adjusted accordingly. This issue will be addressed as part of our future work. In addition, we will investigate the use of thermal images for hand gesture detection and recognition, and we will continue our work on blood-vein based face recognition using thermal images. We will also further develop our multisensory approach (time of flight cameras, thermal cameras, visible spectrum cameras, and lasers) for human and object detection and recognition by robots in domestic environments.

Acknowledgements This research was partially funded by FONDECYT-Chile, grant 1090250.

References

1. Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: 19 results all computer vision researchers should know about. Proc. of the IEEE. **94**(11), 1948–1962 (2006)
2. Hermosilla, G., Loncomilla, P., Ruiz-del-Solar, J.: Thermal face recognition using local interest points and descriptors for HRI applications. Lect. Notes Comput. Sci. (RoboCup Symposium 2010) (2010, in press)
3. Hermosilla, G., Ruiz-del-Solar, J., Verschae, R., Correa, M.: Face recognition using thermal infrared images for human-robot interaction applications: a comparative study. In: 6th IEEE Latin American Robotics Symposium – LARS 2009, Valparaíso, Chile (CD Proceedings), 29–30 Oct. 2009
4. Ruiz-del-Solar, J., Verschae, R., Correa, M.: 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, p. 19 (2009)
5. Kong, S., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition - a review. J. Comput. Vis. Image Understanding **97**(1), 103–135 (2005)
6. Meis, U., Oberlander, M., Ritter, W.: Reinforcing the reliability of pedestrian detection in far-infrared sensing. 2004 IEEE Intelligent Vehicles Symposium, pp. 779–783, 14–17 June 2004
7. Binelli, E., Broggi, A., Fascioli, A., Ghidoni, S., Grisleri, P., Graf, T., Meinecke, M.: A modular tracking system for far infrared pedestrian recognition. 2005 IEEE Intelligent Vehicles Symposium, pp. 759–764, 6–8 June 2005
8. Mudaly, S.S.: Novel computer-based infrared pedestrian data-acquisition system. Electron. Lett. **15**(13), 371–372 (1979)
9. Nanda, H., Davis, L.: Probabilistic template based pedestrian detection in infrared videos. IEEE Intell. Veh. Symposium **1**, 15–20 (2002)

10. Bertozzi, M., Broggi, A., Fascioli, A., Graf, T., Meinecke, M.-M.: Pedestrian detection for driver assistance using multiresolution infrared vision. IEEE Trans. Veh. Technol. **53**(6), 1666–1678 (2004)
11. Wu, S.-Q, Song, W., Jiang, L.-J., Xie, S.-L., Pan, F., Yau, W.-Y., Ranganath, S.: Infrared face recognition by using blood perfusion data. Lect. Notes Comput. Sci. **3546**, 527–531 (2005)
12. Wilder, J., Phillips, P.J., Jiang, C., Wiener, S.: Comparison of visible and infra-red imagery for face recognition. In: Proc. of the 2nd Int. Conf. on Automatic Face and Gesture Recognition, pp.182–187, 14–16 Oct 1996
13. Li, S., Chu, R., Liao, Sh., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 627–639 (2007)
14. Li, S., Chu, R., Ao, M., Zhang, L., He, R.: Highly accurate and fast face recognition using near infrared images. Lect. Notes Comput. Sci. **3832**, 151–158 (2005)
15. Verschae, R., Ruiz-del-Solar, J., Correa, M.: A unified learning framework for object detection and classification using nested cascades of boosted classifiers. Mach. Vis. Appl. **19**(2), 85–103 (2008)
16. Correa, M., Ruiz-del-Solar, J., Bernuy, F.: Face recognition for human-robot interaction applications: a comparative study. Lect. Notes Comput. Sci.,(RoboCup Symposium 2008) **5399**, 473–484 (2009)
17. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)
18. Tan, X., Chen, S., Zhou, Z.-H., Zhang, F.: Face recognition from a single image per person: a survey. Pattern Recogn. **39**, 1725–1745 (2006)
19. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces: a survey. Proc. IEEE **83**(5), 705–740 (1995)
20. Face Recognition Homepage. Available in January 2008. http://www.face-rec.org/
21. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cognitive Neurosicence **3**(1), 71–86 (1991)
22. Mendez, H., San Martin, C., Kittler, J., Plasencia, Y., Garcia, E.: Face recognition with LWIR imagery using local binary patterns. In: Proceedings ICB2009 (2009)
23. Ruiz-del-Solar, J., Verschae, R.: Robust skin segmentation using neighborhood information. In: The Eleventh International Conference on Image Processing (ICIP 2004), 24–27 October 2004, pp. 207–210. IEEE Press, Singapore (2004)
24. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. Proc. Conf. Comput. Vis. Patt. Recogn. **1**, 511–518 (2001)
25. Jones, M., Rehg, J.M.: Statistical color models with application to skin detection. Int. J. Comput. Vis. **46**(1), 81–96 (2002)
26. Hjelmas, E., Kee Low, B.: Face detection: a survey. Comput. Vis. Image Understanding, **83**(3), 236–274 (2001)
27. Satake, J., Miura, J.: Robust Stereo-based Person Detecting and Tracking for a Person Following Robot. In: Proc. ICRA 2009 workshop on person detection and tracking. Kobe, Japan (2009)
28. Wilhelm, T., Böhme, H.-J., Gross, H.-M.: A multimodal system for tracking and analyzing faces on a mobile robot. Robot. Auton. Syst. **48**(1), 31–40, (2004), European Conference on Mobile Robots (ECMR ’03)
29. Medionia, G., R.J. Françoisa A., Siddiquia, M., Kima, K., Yoonb, H.: Robust real-time vision for a personal service robot. Comput. Vis. Image Understanding **108**(1–2), 196–203, Special Issue on Vision for Human-Computer Interaction, October–November 2007
30. Li, L., Koh, Y.T., Ge, S.S., Huang, W.: Stereo-based human detection for mobile service robots. Control, Automation, Robotics and Vision Conference, 2004, ICARCV, vol. 1, pp. 74–79 (2004)
31. Bellotto, N., Hu, H.: Multisensor-based human detection and tracking for mobile service robots systems, man, and cybernetics, Part B: cybernetics. IEEE Trans. **39**(1), 167–181 (2009)
32. Böhme, H.J., Wilhelma, T., Keya, J., Schauera, C., Schrötera, C., Großa, H-M., Hempelb, T.: An approach to multi-modal human–machine interaction for intelligent service robots. Robot. Auton. Syst. **44**(1), 83–96 (2003)
33. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
34. Wisspeintner, T., van der Zant, T., Iocchi, L., Schiffer, S.: RoboCupHome: scientific competition and benchmarking for domestic service robots. Interaction Studies **10**(3), 392–426(35) (2009)
35. iRobot Official Website. Available on Dec. 2010. http://store.irobot.com/home/index.jsp
36. RoboCupHome Official Website. Available on Dec. 2010. http://www.ai.rug.nl/robocupathome/
37. FLIR TAU 320 thermal camera. Information available on Dec. 2010. http://www.flir.com/cvs/cores/uncooled/products/tau/
38. PMD Technologies website: http://www.pmdtec.com/. Accessed Dec 2010
39. Carmen Robot Navigation Toolkit website: http://carmen.sourceforge.net/. Accessed Dec 2010
40. Ruiz-del-Solar, J., Correa, M., Lee-Ferng, J., Hevia-Koch, P., Parra, I., Mascaró, M.: UChile HomeBreakers 2010 Team Description Paper. RoboCup Symposium 2010, 19-25 June 2010. Singapore (CD Proceedings)
41. Munder, S., Gavrila, D.M.: An experimental study on pedestrian classification. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1863–1868 (2006)
42. Gavrila, D.M., Munder, S.: Multi-cue pedestrian detection and tracking from a moving vehicle. Int. J. Comput. Vis. **73**, 41–59 (2007)
43. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: survey and experiments. Pattern Anal. Mach. Intell., IEEE Trans. **31**(12), 2179–2195 (2009)
44. Felzenszwalb, P.F., Girshick, R.B., Mcallester, D.: Cascade object detection with deformable part models. In: Proc. of IEEE Int’l Conference on Computer Vision and Pattern Recognition (2010)

45. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. 1, pp. 886–893 (2005)
46. Viola, P.A., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. IJCV, **63**(2), 153–161 (2005)
47. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on riemannian manifolds. IEEE T-PAMI, **30**(10), 1713–1727 (2008)
48. Paisitkriangkrai, S., Shen, C., Zhang, J.: Fast pedestrian detection using a cascade of boosted covariance features. Circuits Syst. Video Technol., IEEE Trans. **18**(8), 1140–1151 (2008)
49. Zhu, Q., Yeh, M.-C., Cheng, K.-T., Avidan, S.: Fast human detection using a cascade of histograms of oriented gradients. In: Computer Vision and Pattern Recognition, IEEE Computer Society Conference, vol. 2, pp. 1491–1498 (2006)
50. Wojek, C., Schiele, B.: A performance evaluation of single and multi-feature people detection. In: DAGM Symp. on Patt Rec, pp. 82–91 (2008)
51. Tao, J., Odobez, J.-M.: Fast human detection from videos using covariance features. In: Workshop on VS at ECCV (2008)
52. Dalal, N., Triggs, B., Schmid C.: Human detection using oriented histograms of flow and appearance. In: ECCV (2), pp. 428–441 (2006)
53. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: a benchmark. In: CVPR, pp. 304–311 (2009)