Draft version. To appear in Advances in Intelligent and Soft Computing 2011

Improving Human Face Detection through TOF Cameras for Ambient Intelligence Applications

J.R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez

Abstract One of the cornerstones of ambient intelligence technology is the need of sensory systems to reliably notice the presence of people. Several approaches for detecting humans within a non-controlled scenario can be found in the literature but they exhibit, not enough effectiveness, i.e. a high rate of false positive or true negative detections. This becomes a drawback for the development of a variety of ambient intelligence applications which depend on such sensory capability.

In this paper we propose the use of a TOF camera for noticing human presence by detecting their faces. Apart from a typical intensity image, this camera also provides a range image of the scene. The proposed methodology first detects faces from the intensity image (by using the Viola-Jones algorithm) and then analyzes those detections in the range image to discard false positives. Experimental evaluations of the proposed process have yielded excellent results in non-controlled scenarios, eliminating most of false positive detections.

1 Introduction

Ambient Intelligence (AmI) has recently come out to increase the comfort of people within their daily life. AmI refers to the enhancement of the environment through a set of devices that intelligently controls certain functionalities, pursuing a variety of aims. For instance, an intelligent fridge can take account on the aliments that it contains, automatically ordering those which have been already consumed. Such an intelligent fridge, equipped with gas sensors, could also alert the user when a particular meal is rotten [7]. Thus, the inhabitants of the home can be easygoing with their food supply, simplifying this chore. Other practical cases where AmI may help is in the adjustment of resources consumption at the same time that all the comforts of

J.R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez
System Engineering and Automation Dpt., University of Malaga, Spain, e-mail: jotaraul@isa.uma.es,{cipriano,jgonzalez}@ctima.uma.es

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

the environment are kept. A sensory system can detect the presence/absence of people in a given area to control, for example a heating device. Even more interesting is the possibility to estimate the number of people and to be able to recognize them in order to accordingly adjust the temperature of a room following pre-specified preferences. These are some examples that reveal the human-centric characteristic of Ambient Intelligence technology. Given that humans are the main actors, it becomes clear the interest of Ami in reliably noticing the presence of people in order to act adequately.

Computer vision contributes to this issue by providing algorithms to detect human faces in images [3, 10, 11], through their reliability are still not high enough for many real AmI applications. Very recently, a new type of sensor, called *TOF camera* (Time-Of-Flight camera), has appeared in the market providing both intensity and range data. These cameras permit us to improve the detection of the truly presence of human faces not only through visual information, e.g. skin-like colour, 2D shape, etc., but also by exploiting 3D characteristic information of human faces, for example, the area of the nose is prominent with respect to the area of the cheeks.

In the recent literature some works can be found exploring the capabilities of TOF cameras. Many of them address the physical characterization of TOF cameras [2, 12], and others focus on their application in a variety of fields [13, 14]. Some related works to ours are [4, 6], which propose simple and direct approaches for face detection using TOF cameras. In this paper we improve their achievements through a battery of 3D-shape tests that leads to a reliable and robust face detector. Our experiments have been successfully conducted on real and non-modified scenarios, yielding results with a high rate of true positives and reducing very significantly the number of false positives, which are a serious problem for known intensity-based face detectors. Some of our experiments have been conducted with the TOF camera mounted on a mobile robot, to prove the suitability of our approach for AmI applications that may incorporate a service robot.

Next we describe the proposed method and the evaluation results of the experiences conducted on real, non-controlled scenarios.

2 The Proposed Method

Our aim is to develop a robust face detection process using a TOF camera that produces the minimum number of false positives as possible, while maintaining high detection rates. Our approach is divided into two steps (see figure 1): the first one applies the well-known Viola-Jones classifier [10] over the intensity image to determine a set of candidates regions that presumably contain a human face. In the second step, the resulting candidates are checked against a battery of 3D-shape tests that operates over the range image. This second phase aims at discarding false positives of the first stage.

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

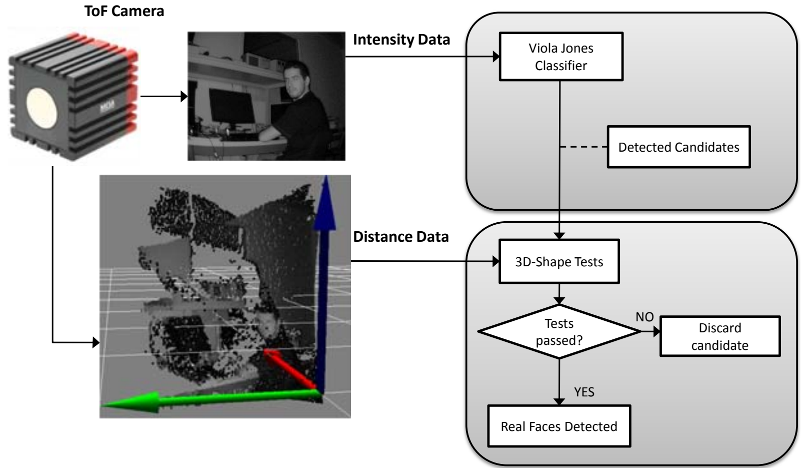

Fig. 1 Scheme of the proposed two-step process for face detection.

2.1 TOF cameras

Broadly speaking, a time-of-flight camera is a device that provides both intensity and range data, i.e. distance information to the sensed objects. In particular, the TOF camera considered in this work, manufactured by *Mesa Imaging* [8], performs by emitting a continuous wave modulation through an infrared led array (see figure 2). It calculates the phase difference between the sent and received signal, obtaining range information with respect to the camera reference system up to a distance of 5 m. with an accuracy of ± 10 *mm*.

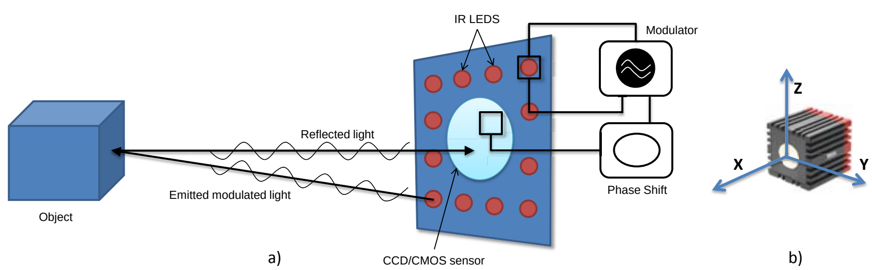

Fig. 2 a) The leds of the camera emit a continuous wave modulated signal. When the signal is reflected to the camera, their CCD/CMOS cells are excited with a signal with a certain phase shift, which permits the device to calculate the distance and reflectivity of the target through cross correlation. b) Reference system established by the used TOF camera.

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

In spite of the interesting possibilities provided by TOF cameras, they have still some limitations (see [12] for more details):

- Low resolution. The used camera has a resolution of 176 *x* 144 pixels.
- Unsuitable for highly dynamic scenes. When a scene presents fast moving objects, measured distances are prone to large errors.
- Low accuracy of the measured distance. Distance measurements are influenced by the colour and the type of material of objects.

2.2 *Viola-Jones classifier*

The Viola-Jones classifier applies a cascade of tests of increasing complexity, over the intensity image (see figure 3). First stages of the cascade are simple and quickly discard regions that do not match general human faces' features, e.g. the eyes' area is darker than the nose one. The subsequent stages are increasingly more selective and complex, minimizing the false positive rate, at the expense of a higher computational burden.

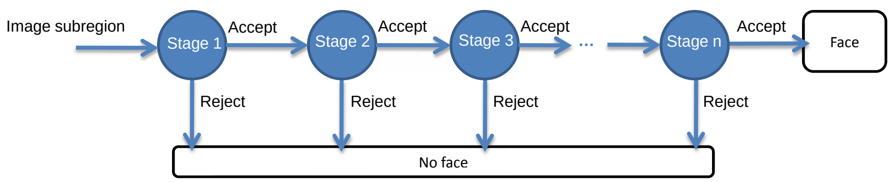

Fig. 3 Structure of a face detector cascade. Computationally less demanding stages are applied first.

Tests considered in each stage are trained using AdaBoost[1]. AdaBoost is a learning algorithm that chooses weak classifiers between a family of simple (*Haar*) features. These classifiers are combined into strong classifiers, which become constituents of the stages of the Viola-Jones cascade (refer to [10] for further detail).

Once trained, the classifier is applied to a video sequence. Within each frame, a 20*x*20 pixels window is slid, checking if it passes all the stages of the cascade. This window is moved and scaled to cover the whole frame in order to detect a face at different locations and scales. The considered window size for the Viola-Jones classifier and the low resolution of our TOF camera limit the distance for detecting faces to 2.5 meters.

Although this face detector provides good results, the number of false positives it produces is a serious drawback for many AmI applications.

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

2.3 3D-shape tests

To overcome the problem of such a high number of false positives, we propose a second step where three 3D-shape tests are applied over the range image to confirm or not every candidate. Each test checks out if a particular 3D-shape feature is present in the candidate region or not, identifying and eliminating different types of false positives. Figure 4 shows some examples of false positives detected by the two of the proposed tests which are described in the following sections.

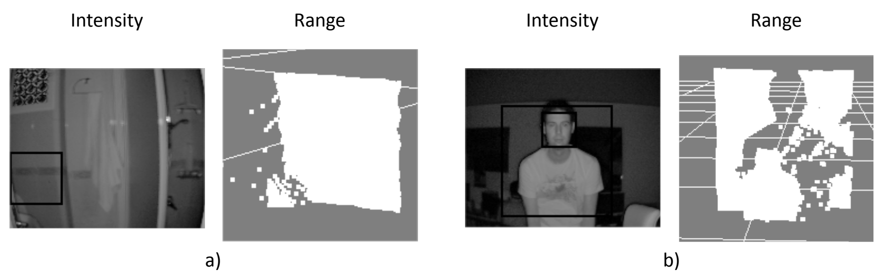

Fig. 4 a) A false positive detected because it is a flat region. b) A false positive because of an abnormal face size for that distance.

2.3.1 Test #1: Flat region

The first test is based on the fact that human faces are not flat, so they present certain relief on the range image. An example of this type of false positives is shown in figure 4a. This test is implemented by computing the covariance matrix, C, of the spatial position (x, y, z) of the pixels that conform the candidate region. Eigenvalues of the C matrix give information about the spatial distribution of the pixels that form the region. Concretely, the lower the smallest eigenvalue, the flatter the region. Notice that a similar reasoning could be stated for the standard deviation of the z coordinates (distances to the camera) which is cheaper to compute, however, we have verified that it presents some problems for slightly oblique faces.

2.3.2 Test #2: Size-distance ratio

This test discards candidates that do not match the expected size of a human face at a given distance. Figure 4b is an example of a false positive detected by this test. In our implementation we have considered that the normal size of human faces, on average, is 290 cm^2 with a typical deviation of 60 cm^2 approximately. These figures have been obtained by analyzing around 10000 face images.

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

2.3.3 Test #3: Facial structures

This test exploits the fact that human faces have a general common morphology. It segments the candidates using a growing regions method over the range data to extract the candidate face from the background, and then it divides it into nine subregions as shown in figure 5a. The depths of these subregions to the camera must verify certain constrains characteristic of human faces. For instance, the region that presumably contains the nose (numbered as 5 in figure 5a) should stand out with respect to the lateral subregions, that correspond to the cheeks, i.e. 4 and 6. An example of a false positive declared by this test is shown in figure 5b.

In the implementation of this test, the position of each region is set as the centroid of the pixels in it. Five comparisons are performed to check whether the central regions of the three rows and of the two diagonals are closer to the camera than the side regions. More concretely, we make the comparison on the x-y plane in the following manner:

Let $x = a \cdot y + b$ be the straight line that links the centroids of the side regions, and let $P_c = (y_c, x_c)$ be the centroid of the central region we check if the central region is closer to the camera than the side regions, that is, the distance $d = y_c \cdot a + b - x_c$ must be positive. This checking is done for the five cases (3 rows + 2 diagonals).

3 Evaluation

In order to prove the effectiveness of our approach we have conducted 16 experiments in two scenarios: *fixed-camera*, where the TOF camera was fixed while people were moving around, and *moving-camera*, where the camera was mounted on a mobile robot, and thus, it was affected by important changes in illumination (a video can be watched at `http://www.youtube.com/watch?v=GJE4A7R6LNs`).

We apply the Viola-Jones cascade classifiers from the OpenCV library, concretely the one labeled as *haarcascade_frontalface_alt2*, for detecting candidate regions. Although it provides good results [9], a considerable rate of false positives are produced. In our experiments from a total of 11.184 candidates, $685(5.77\%)$ was the number of false positives. These false positives as well as others reported here were identified by visual inspection.

The proposed 3D-shape tests have been implemented in a multi-thread C++ program, using a Intel®Core$^{\mathrm{TM}}$2 Quad CPU Q6600 2.4GHz, with 4Gb RAM, which permits us to process up to 14 fps. We take advantage of the multi-core capability to run simultaneously all the tests, albeit a sequential execution solution could be also adopted.

When the 3D-shape tests are considered the number of false positives is drastically reduced, as shown in table 1. Notice that each test by itself has a modest ratio of false positive declaration, but when combined they yield excellent results, eliminating all the false positives in the fixed-camera scenario and around the 97% of cases with the camera onboard the robot. Regarding the false negative rate, i.e. faces which

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

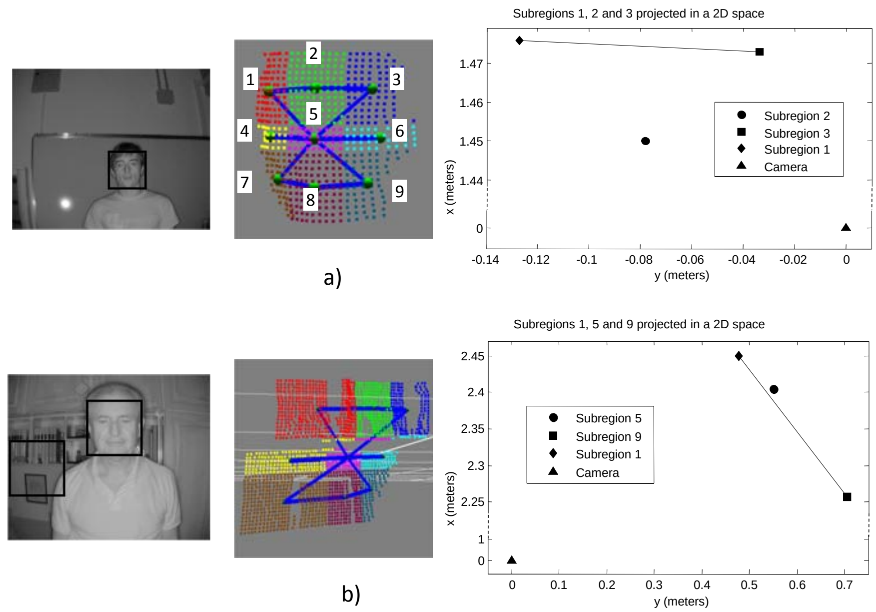

Fig. 5 a) Left: A candidate region. Middle: The centroids are represented with spheres while lines represent the constrain to be checked. Right: projection into the y-x plane. Notice that subregion 2 is closer to the camera (less x coordinate) than the two side ones on the row (1 and 3), and thus the constrain is met. b) Left: A false positive (behind the person). Middle: Subdivision of the region and the restrictions to be fulfilled. Right: projection of three subregions into the y-x plane. Note that the test is not passed in this case.

are erroneously neglected, the battery of the three tests discards 3.84% and 2.83% of actual faces in fixed- and moving-camera scenarios respectively, which does not represent a serious inconvenience since we are dealing with a video sequence.

Scenarios	Frames	Viola Jones false positives	Test #1	Test #2	Test #3	All
Fixed-camera	11230	122	60,66%	68,85%	53,27%	100%
Moving-camera	27649	563	22,74%	53,46%	93,25%	96,98%

Table 1 False positives detected by each test separately, and by the combination of all of them with respect to the total of false positives within static and dynamic configurations.

Draft version. To appear in Advances in Intelligent and Soft Computing 2011

4 Conclusions

In this work we have presented a robust human face detector for TOF cameras that highly reduces the number of false detections, while keeping a very low level of false negative cases. The obtained results demonstrate the interest in using not just intensity images, but also range data to achieve the robustness level demanded by AmI applications. Though nowadays these TOF cameras are still very expensive, we believe that the emergence of these sensors to interface to next generation of interactive games (as Kinect) will put them in the market at a very cheap price. In the future we plan to investigate the use of the proposed method in a service robot aimed to provide AmI capabilities.

References

1. Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt' 95,pages 2337. Springer-Verlag, 1995.
2. Andreas Kolb, Erhardt Barth, and Reinhard Koch. ToF Sensors: New Dimensions for Realism and Interactivity. In CVPR 2008 Workshop on Time-of-Flight-based Computer Vision (TOF-CV), 2008.
3. Cha Zhang and Zhengyou Zhang. A Survey of Recent Advances in Face Detection. Technical Report, MSR-TR-2010-66. June 2010.
4. Dan Witzner Hansen, Rasmus Larsen, and Francois Lauze. Improving Face Detection with TOF Cameras. In Proc. of the IEEE Int. Symposium on Signals, Circuits & Systems (ISSCS), vol. 1, pages 225-228, Iasi, Romania, 2007.
5. J. Gonzalez, C. Galindo, J.L. Blanco, J.A. Fernandez-Madrigal, V. Arevalo, F.A. Moreno. SANCHO, a Fair Host Robot. A Description. IEEE Int. Conf. on Mechatronics (ICM'09), 2009.
6. Martin Bhme, Martin Haker, Kolja Riemer, Thomas Martinetz, and Erhardt Barth. Face Detection Using a Time-of-Flight Camera. In Lecture Notes in Computer Science, vol. 5742, pages 167-176, 2009.
7. Mathias Broxvall, Silvia Coradeschi, Amy Loutfi and Alessandro Saffiotti. An Ecological Approach to Odour Recognition in Intelligent Environments. IEEE Int Conf on Robotics and Automation (ICRA'06, Orlando, Florida, May 2006).
8. Mesa Imagin website: http://www.mesa-imaging.ch/.
9. M. Castrillon-Santana, O. Deniz-Suarez, L. Anton-Canalis and J. Lorenzo-Navarro. Face and Facial Feature Detection Evaluation.III Int. Conf. on Computer Vision Theory and Applications, 2008.
10. Paul Viola and Michael Jones. Robust real-time face detection. In Porc. Int. Conf. on Computer Vision, 57(2):137-150, 2004.
11. Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In IEEE ICIP 2002, vol. 1, pp. 900-903, 2002.
12. Robert Lange and Peter Seitz. Solid-State Time-of-Flight Range Camera. In IEEE Journal of Quantum Electronics, Vol 37, No. 3, March 2001.
13. David Droeschel, Stefan May, Dirk Holz, Paul Ploeger, and Sven Behnke. Robust Ego-Motion Estimation with ToF Cameras. In Procc of the European Conf. on Mobile Robots (ECMR'09), Croatia 2009.
14. Sergi Foix, Guillem Alenya, Juan Andrade-Cetto and Carme Torras. Object Modeling using a ToF Camera under an Uncertainty Reduction Approach. IEEE Int. Conf on Robotics An Automation (ICRA'10), Anchorage, Alaska, 2010.