A low power colour-based skin detector for smart environments

Michela Lecca, Massimo Gottardi[1], Bojan Milosevic[2] and Elisabetta Farella[2]

Fondazione Bruno Kessler, Center for Information and Communication Technology, Technologies of Vision, Trento, Italy
[1]Fondazione Bruno Kessler, Center for Materials and Microsystems, Integrated radiators and Image Sensors, Trento, Italy
[2]Fondazione Bruno Kessler, Center for Information and Communication Technology, Energy Efficient Embedded Digital Architectures, Trento, Italy
Emails: lecca | gottardi | milosevic | efarella@fbk.eu

We describe an embedded optical system detecting human skin under a wide range of illuminant conditions. Our attention to such a system is justified by the many applications for which skin detection is needed, e.g. automatic people monitoring and tracking for security reasons or hand gesture recognition for fast and natural human-machine interaction. The presented system consists of a low power RGB sensor connected to an energy efficient microcontroller. The RGB sensor acquires the RGB signal from a region in front of it over a wide dynamic range, converts it in the rg chromaticity space directly on chip and delivers the processed data to the microcontroller. This latter classifies the input signal as skin or non-skin, testing its membership to a skin locus, i.e. to a compact set representing the chromaticities of the human skin tones acquired under several illuminant conditions. The system architecture distributes the computational load of skin detection on both hardware and software, providing a reliable skin detection with a limited energy consumption. This makes the system suitable to be used in smart environments, where energy efficiency is highly desired in order to keep the sensors always ready to receive, process and transmit data without affecting the performance.

Received 08 April 2016; revised 15 June 2016; accepted 16 June 2016

Published online: 3 August 2016

Introduction

Smart sensors, cyber-physical systems, pervasive computing, multi-modal human-machine interactions are only some of the keywords that characterise an emerging, new world where advanced technologies work in a continuous and collaborative way to make human life more comfortable [1-3]. In these smart environments, the sensing technologies, such as cameras, wearables, tablets, phones, are usually wirelessly connected to each other, and are always ready to acquire, transmit and/or process incoming data. The continuous data processing and exchanging require an efficient management of the energy and of the computational resources. In fact, in wireless sensor network, the sensing functionality is constrained by the battery life and the data transmission is one of the most power hungry tasks. Embedded systems are the current solution to the need of technologies guaranteeing high performance with low power consumption. They are hardware and software dedicated technologies, where the processing of the acquired data is performed as early as possible, integrated on hardware or performed by a microcontroller [4-5], usually with low power consumption characteristics.

In this work, we describe an embedded optical system performing human skin detection, extending a work recently presented in [6]. Here we propose a slightly different, more efficient version of it, and we discuss some possible applications in smart environments. Our attention to such a smart *skin detector* is motivated by the many applications requiring the visual detection of the human skin in a certain scene, like e.g. face detection/recognition [7-10] or people tracking for security reasons [11], hand gesture recognition for human-machine interaction [12-13], automatic image annotation for multimedial browsing [14].

The system on which this work is based is composed by a low power RGB vision sensor, connected to an energy efficient microcontroller (MCU). The RGB sensor embeds signal processing on board: it acquires the RGB signal in auto-exposure mode and converts it into the rg space directly on chip with low power consumption [15]. Such a signal is acquired and digitalised by the MCU using the integrated Analog to Digital Converter (ADC), and it is labeled as skin or non-skin based on a skin locus [7, 16]. The skin locus is a compact set in the rg space formed by the human skin tones captured under different illuminants, e.g. direct sunlight, daylight, candles, and fluorescent lamps. The classification of any colour signal as skin or non-skin is performed by checking its membership to the skin locus: rg points with chromaticities falling out of (falling in, resp.) the skin locus are classified as non-skin (skin, resp.). According to this principle, the skin locus acts as illuminant invariant filter to detect the signals possibly coming from skin regions.

In this system, the optimisation of the computational resources is achieved by using a smart vision sensor that embeds the conversion from RGB to rg in hardware. The on-chip implementation of the conversion loads a part of the computational burden on the hardware. Moreover, the classification task is implemented on the MCU directly connected to the RGB sensor, without requiring the transmission of the data to external processors, such as a PC.

A prototype of this system, consisting of a RGB photodiode and built up with off-the-shelf components, was used as a smart switch [6], i.e. for touchless activation of an appliance connected to it, such as the monitor of a computer, a sliding door, a smart-phone. In this application, the system worked in tandem with a proximity sensor. Precisely, the vision system captured continuously the visual signal of a region in front of it, then, when the acquired signal was classified as skin, it woke up the proximity sensor. This latter measured the distance of the incoming signal from the sensor: if this distance was in a desired range, the system sent a pre-defined command to wake up an appliance. The experiments showed that the usage of the colour information remarkably improve the performance of a switch based on a proximity sensor only.

The system version we present here differs from the original one for the way it represents the skin locus and checks the membership of a rg signal to it. In the original version of the system, the skin locus is analytically modeled by specifying its boundaries as two quadratic polynomials, and the membership of a rg signal to the skin locus is established by evaluating the position of the signal with respect to these polynomials. As specified next, this task requires floating point multiplications and additions. In our new version, we model the skin locus as a binary map, i.e. as an image where black pixels represent the rg coordinates of the skin while the white pixels represent the rg coordinates of non-skin objects. The membership of a rg signal to the skin locus is thus performed by checking if the corresponding entries in the binary map is black or white. This requires only a memory access and a comparison, thus it is more efficient. With this approach, we avoid the use of floating point operations, with the advantage of a higher efficiency of the implementation on a low-power MCU. The dimension of such binary map sets the trade-off between memory requirements and accuracy and will be evaluated and compared with the floating-point analytical implementation.

In this paper we describe the general architecture of the system and we report the details of the modified classification algorithm. Moreover, we discuss two possible applications of the skin detector: its usage as a smart switch as previously proposed [6], and as a people counter for statistical tasks, e.g. counting the number of persons visiting a shop or a certain room of a museum.

Colour-based skin detection in rg chromaticity space

Detecting skin in a digital image consists in classifying each image pixel as skin or non-skin. The colour information is widely used for this task [3, 17-20]. In fact, in many real-world scenarios, skin colour has been proved to be very distinctive from the rest [7, 16, 21]. Moreover, colour is robust against several geometric transforms, such as translation, rescaling, in-plane rotation, some perspective changes, and against Gaussian noise addition [22], therefore it provides a robust description of the visual appearance of any object under many different conditions.

The starting point of any colour-based skin detection algorithm is the choice of a colour space, where to represent the human skin tones. The work in [23] pointed out that there is not an *absolute optimal* colour space for skin detection, but the choice of a colour space among the others usually depends on the application at hand. Therefore, many colour spaces have been proposed, e.g. RGB [24], YCbCr [21], rg [16], HSV [25], CIELab [26], CIELUV [27], TSL [28]. A comprehensive review about the colour spaces for skin detection is available in [20, 29-31].

In the following subsection, we introduce the rg chromaticity space, explaining why we choose this space for modeling the skin colour, and we describe the skin detection algorithm.

The rg chromaticity space

The proposed system represents the human skin tones in the rg chromaticity space. The rg chromaticities of a pixel x are the components of a 2D vector $(r, g) \in [0, 1]^2$ obtained from the (R, G, B) coordinates of x by the following equation:

$$(r, g) = \left(\frac{R}{R+G+B}, \frac{G}{R+G+B}\right) \quad (1)$$

where $R + G + B \neq 0$, and $(r, g) = (0, 0)$ otherwise.

The choice of the rg space for modeling the skin tones is motivated by two main reasons:

1. the rg space has been proved to be adequate for modeling and detecting the skin tones [9] and it has been used in several previous works [e.g. 7, 16];
2. the conversion from RGB to rg colour space is suitable to be implemented on hardware with low energy consumption [15, 32].

In principle, the rg coordinates are invariant to changes of light intensity [33], so that they allow a good description of the skin tones also in presence of shadows or in scenes with a non uniform light intensity. Nevertheless, in practice, this invariance is guaranteed only upon a certain extent because of the physical constraints of the acquisition device. In fact, in case of very brilliant lights or in the darkness, a standard camera cannot acquire the RGB colour components in a reliable way and this negatively affects the computation of the rg coordinates. Enabling an auto-exposure time control overcomes this problem.

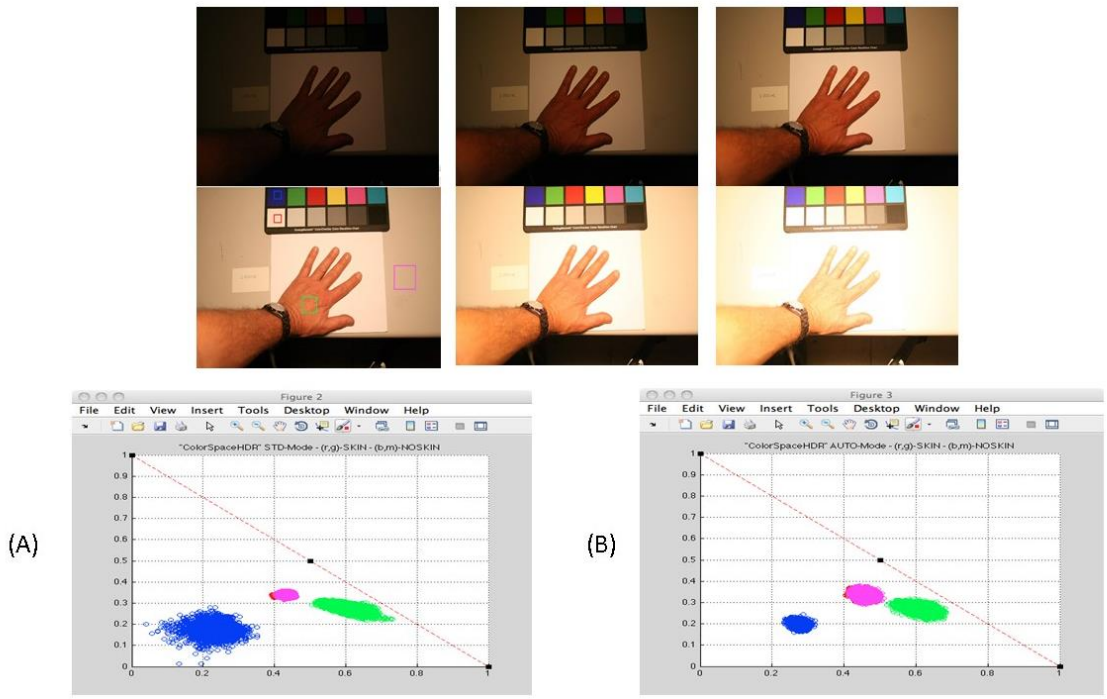

Figure 1: While in principle, the rg coordinates are insensitive to changes of light intensity, in practice, in a standard camera, this invariance is guaranteed only upon a certain extent. This figure shows on top an image acquired under the same light with six increasing intensity values, and on bottom the rg point clouds – across all the images - of the rectangular regions indicated on the 4th picture. According to their colours, the clouds identify the points within the corresponding rectangles. In (A) a fixed exposure time is used, while in (B) an auto-exposure time is employed.

Figure 1 shows the same scene captured under a light with fixed chromaticity, but increasing intensity. The first images are quite dark, while the latter ones are very bright. Figure 1A shows the rg coordinates of four regions, indicated by the rectangles in the fourth picture, across the different images, as obtained by capturing the colour signal by a standard camera with a fixed exposure time. Figure 1B shows the same points captured by tuning the exposure time. The point clouds in Figure 1A are much more sparse than those in Figure 1B, meaning that the data acquired with an adaptive exposure time is more robust against changes of the light intensity. In order to provide accurate data, the RGB sensor used in our prototype [15] enables an auto-exposure that allows a reliable signal acquisition over a dynamic range of about 105 dB.

The skin locus and the classification algorithm

In the rg chromaticity space, the rg coordinates of the human skin acquired under different light conditions, regardless of the user's ethnicities, form a compact set that is called *skin locus* [32]. This name derives from its shape, which follows the curve (termed Planckian locus) that the rg chromaticities of an incandescent black body track in the rg space when the temperature of the black body changes [32].

The skin classification based on the skin locus is implemented by testing the membership of any incoming rg signal to the skin locus. If the rg point falls into the skin locus, then it is labeled as skin, otherwise it is classified as non-skin.

This classification approach has two main advantages:

1. since it uses colour, the classification based on the skin locus is robust against geometric distortions, as rescaling, rotations, translations, skew; this makes the approach appropriate for detecting deformable objects like hands, thus for hand gesture recognition, and for detecting faces that can have different spatial orientations;
2. since the skin locus collects the skin tones under many illuminants, it acts as an *illuminant invariant filter*, making possible the skin recognition under varying light conditions.

We observe that the signal captured by a camera depends on the spectral characteristics of the photodiodes of the camera, thus the skin locus generally varies from camera to camera [9]. This means that the skin locus based classification is not invariant against changes of devices.

For a given standard RGB camera, the skin locus is computed by acquiring the RGB signals coming from human skin regions under different lights and converting them to the rg coordinates. The shape and the size of the skin locus depend on the number of people involved in the acquisition process and on the correlated colour temperature of the lights under which the skin has been captured. Generally, both the number of people and the illuminants are chosen according to the application at hand. For example, if we are interested to detect skin in outdoor environments only, we can skip the acquisition of the human skin under neon or tungsten lamps, which are typical of indoor scenarios. Similarly, there are applications for which the acquisition of the skin locus of a single person is sufficient. This is the case of a cell phone, used by its owner only, that wants to switch on his/her phone from the stand-by state simply by putting the hand palm in front of the phone for some seconds. In this framework, the skin detection can be performed simply by testing the membership of the RGB signal to the skin locus of the owner. However, we highlight that the skin colour is not discriminative enough to distinguish two persons, i.e. different persons may have similar skin tones.

In many other applications, as the activation of a device in public places (e.g. the sliding doors of a shop or a ticket machine at a train station), the skin locus must represent different human skin tones under several illuminants.

The main drawback of this skin detection approach is that the classification is performed point by point, i.e. it analyses a single signal located in a very small region Ω in front of the sensor, regardless of the signals coming from other regions, adjacent to Ω. The pointwise approach may generate many false positives which can be avoided by considering additional information, regarding for instance the rg statistical distribution in a neighborhood of the acquired signal [7-8, 16]. However, the experiments reported next showed that in many applicative scenarios, the classification provided by the skin locus ensures good performance.

Modeling the skin locus

There exist several models for representing the skin locus. For instance, as a regular, smooth region in the rg plane [7, 16] and approximated by a bounding box [34].

In previous work [6], we modeled the skin locus S by specifying its boundary by two quadratic polynomials $d(r)$ and $u(r)$ [7, 11], so that:

$$S = \{(r, g) \in [0,1] \times [0,1] : d(r) \leq g \leq u(r)\}. \quad (2)$$

The classification of a rg point as skin or non-skin was performed by computing the values $d(r)$ and $u(r)$ and by checking the conditions $d(r) \leq g$ and $g \leq u(r)$ [6]. In this analytical model, the (r, g) values and the polynomial coefficients are represented using floating point (FP) values and the evaluation of each new point requires 12 FP operations (6 multiplications, 4 additions and 2 comparisons). Such an implementation is not optimised for low-power MCUs: they are usually not

provided with a dedicated hardware for the execution of FP operations, which are computed using successive integer approximations and thus require extensive processing time.

In this work, we propose to use a discrete model making the skin classification computationally more efficient. Precisely, we discretise the rg space and represent the skin locus as a binary map S_D, of size $N \times N$ where the black pixels (= 0) represent the skin chromaticities, while the white ones (= 1) represent the chromaticities of non-skin materials. In this framework, any input signal (r_i, g_i) is mapped to $\left([Nr_i], [Ng_i]\right)$, where $[\cdot]$ denotes the cast of the argument to the closest integer number. The membership of the signal to the discretised skin locus is tested by checking if the entry $S_D\,[Nr_i][Ng_i]$ is zero (i.e. black pixel, thus skin) or one (white pixel, thus non-skin). This approach avoids the use of FP operations and reduces the signal classification to the computation of the integer indexes of the signal $\left([Nr_i], [Ng_i]\right)$ and the evaluation of the corresponding bit in the binary map. The drawback here is that the binary map has to be stored in the MCU's memory. Its dimension is N^2 $[bits]$ and it determines the trade-off between memory requirements and accuracy. A low value of N provides a coarse representation of the skin locus, and may adversely affect the classification performance; a high value of N approximates better the skin locus, reproducing the shape of the set S but requires more memory.

The system: general characteristics

The presented system is composed by a RGB sensor and a MCU. The RGB sensor's electronics and its operating principle are sketched in Figure 2. The sensor acquires the visual signal through a three channels photodiode, where each channel discharges from a fixed voltage V_{res} with a slope proportional to the light intensity (see Figure 2B). The sensor can acquire the signal in two ways:

1. standard mode: with a fixed exposure time, i.e. the signal RGB is sampled at a fixed time T, then converted to rg space;
2. auto exposure mode: during the exposure time, the RGB sensor computes the intensity $S(t) = R(t) + G(t) + B(t)$ and samples the R, G, B values when $S(t)$ reaches a pre-fixed, maximum voltage value T. The signal is sampled at the time $t = T_s$ such that $S(t) = \mathrm{T}$ and the rg coordinates are computed as:

$$r = \frac{R(T_s)}{\mathrm{T}}, g = \frac{G(T_s)}{\mathrm{T}} \tag{3}$$

The data acquisition in auto-exposure mode provides a more accurate signal than that obtained by exploiting the standard mode (see Figure 1). The rg chromaticities are in this case captured with a high dynamic range of up to 105 dB [15], that is twice greater than the dynamic range of a standard images (about 55dB).

Figure 2A sketches the electronics of the sensor: the photodiodes provide the RGB voltage signals to an analog ADDER, which computes $S(t)$ on-the-fly, during the exposure time. A voltage comparator matches $S(t)$ with the threshold T and triggers the SAMPLE&HOLD block, which in turn samples the signals R, G. The output (r, g) is delivered to the connected MCU for the AD conversion and processing.

The estimated power consumption is about 170pW per sample, while that of a standard photodiode is 15nW per sample, making the proposed sensor suitable for smart environments. The main drawback in using this sensor is the larger pixel pitch (10μm-15μm), due to the electronics needed to embed the local processing. Alternate solutions for touchless interaction use proximity sensors, which estimate the distance between the sensor and the object in front of it. A state of the art commercial proximity sensor,

such as the VL6180X from ST Microelectronics [35], has an average active power consumption of 5.1 mW, making it considerably less efficient than the proposed solution.

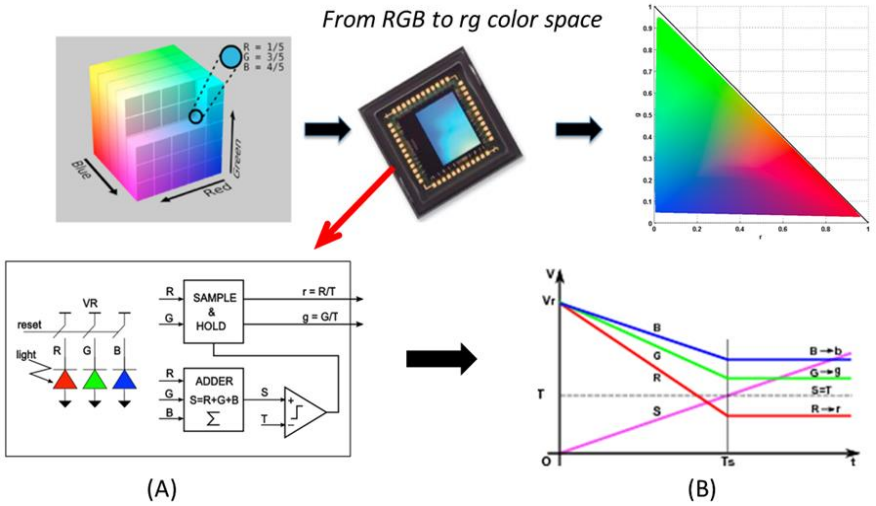

Figure 2: Work-flow of the RGB sensor. The sensor acquires over a high dynamic range a RGB signal and converts it into the rg space. Subfigure A shows the operating principle of the pixel, while subfigure B shows the sampling principle.

The prototype

In the current skin detector prototype, we equipped the RGB system with a 3-channel photodiode (RGB) from Hamamatsu [36], sensitive to the red, green, blue regions of the visible spectrum, with maximum sensitivities to the wavelengths of 630 nm, 540 nm, and 460 nm respectively. Moreover, we adopted a 6 mm optics that provides a field of view of 20°.

As MCU, we employed an 8 bit Atmel ATMega328 [37], which is an 8bit AVR RISC-based solution running up to 20 MHz and achieving 1 MIPS per MHz, balancing power consumption and processing speed. Such MCU is available on the Arduino boards, which provide a miniaturized and ready to use board with integrated serial to USB interface and pin connectors [38]. Moreover, the use of an Arduino-enabled board allows to take advantage of its high-level programming interface and to easily implement the desired control and processing functionalities.

In this implementation, the interface between the RGB sensor and the MCU is composed by two control signals (RESET and SH) and the two analogue r, g signals to acquire. As indicated by its name, the RESET signal, set by the MCU, resets the sensor and triggers the pre-charge of the photodiodes necessary for each acquisition. The SH signal, set by the sensor in auto-exposure mode, notifies the end of the sensor sampling and the availability of the r and g signals, that are ready to be delivered to the microcontroller. This acquires the processed signal using two integrated Analogue-to-Digital converters. As already mentioned before, in addition to this auto-exposure functionality, the system can use a fixed exposure time by ignoring the SH signal and sampling the (r, g) signals after a fixed amount of time.

In the current version of the skin detector, once the (r, g) point is acquired, the MCU tests if it represents skin by checking its membership to the discretised skin locus S_D, which is computed offline using a sample skin dataset and stored in the MCU's FLASH memory as a binary array. During data acquisition and testing, the prototype is interfaced to a PC via a serial port and a *Processing* application [39] is used to visualise and log the data on the PC. Figure 3 shows the proposed prototype mounted on

a compact tripod, while Figure 4 shows the proposed setup used for the acquisition of skin to nes under varying illuminant conditions.

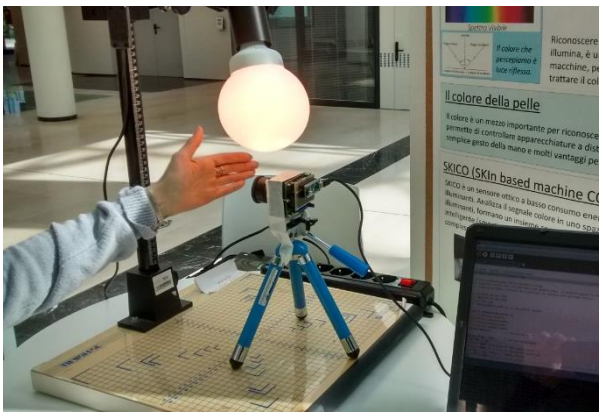

Figure 3 (left): The prototype used in the experiments.
Figure 4 (right): Experimental set up used for acquiring the human skin tones under various illuminants.

Experiments and applications

Here we report the experiments carried out to evaluate the performance of our skin detector.

The section is organised in two parts. In the first one, we describe the skin locus for the used prototype. In the second one, we describe two possible applications of the system in smart environments. Precisely, we discuss its usage as a smart switch, and as people counter for statistical analysis.

In all these experiments, we characterise the light of an environment by a RGB triplet, that we refer here as the *RGB light colour*, measured by means of a colour checker. Precisely, for each light we acquire with our system the rg signals coming from the 24 coloured squares of the colo ur-checker using the auto-exposure mode. Then we compute the mean values (r_m, g_m, b_m) of the (r, g, b) coordinates, where $b = 1 - r - g$. The RGB colour of the light has been obtained as:

$$(R, G, B)_{light} = 255 \left(\frac{r_m}{M}, \frac{g_m}{M}, \frac{b_m}{M}\right) \quad (4)$$

where $M = \max(r_m, g_m, b_m)$.

The first two squares of the colour-checker represent the dark and light colours of the human skin, and thus they have been used, along with skin samples collected from a variety of subjects, to measure the skin detection performance of the system. The other 22 squares of the colour-checker represent the colours of non-skin surfaces, like foliage, flowers, sky, fruits. The rg chromaticities of these non-skin surfaces have been used to measure the number of false positives, i.e. to evaluate how many times the system classifies non-skin regions as skin.

The skin locus

The skin locus for the used prototype as computed in the previous work [6] is shown in 5 on the left. It has been build up by considering 15 illuminants with correlated colo ur temperatures ranging over [2700K, 7000K], including natural and artificial lights. For each illuminant, we collected in auto-

exposure mode the skin tones of a hand palm of 10 volunteers with different ethnicities. We remark that the volunteers whose hand has been taken under an illuminant might differ from those whose hand has been taken under another illuminant. The total number of volunteers participating to these experiments was 75. Figure 5 shows some examples of skin colours captured by our sensor from the hand palm of some volunteers under a natural light. The RGB colours shown in Figure 5 have been obtained by mapping the (r, g, b) triplet detected by the sensor on the RGB space, as done for the light colour (see Equation (4)). Despite the skin tones of the palm are less dispersed over ethnicities than those of other body parts, we observe a certain variability that must be modeled by the skin locus. Our choice to consider skin tones from the hand palm only is justified by the applications considered here: the hand palm it is commonly used in many gesture for the touchless control of machines, e.g. activating/stopping a device or scrolling a menu on a screen.

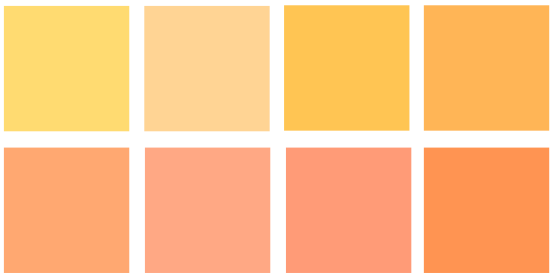

Figure 5: RGB skin tones from the human hand palm of some volunteers captured by the RGB sensor under a natural light. The RGB values have been obtained from the $(\boldsymbol{r}, \boldsymbol{g})$ coordinates acquired by the sensor as done for defining the light colour (see text for more details).

We filtered the collected rg data by removing rg points rarely occurring and some other data that were erroneously captured due to a not proper usage of the sensor (e.g the volunteers moved the hand during the acquisition or they stayed too far or too close to the sensor). The retained rg points have been processed offline to compute the functions d and u delimiting the skin locus by a least square approach applied to the rg points close to the minimum bounding rectangle containing the cloud of the acquired data. We obtained the following functions:

$$u(r) = -2.668\, r^2 + 2.336\, r - 0.152 \quad (5)$$

$$d(r) = 0.027\, r^2 - 0.242\, r + 0.372 \quad (6)$$

In the new version of the system, we describe the regions bounded by the curves u and d in a discrete way, i.e. as a binary matrix with size $N \times N$. Figure 6 shows on the right the discretised skin locus obtained for $N = 8, 16, 32, 64, 128, 256$. The entry ij ($i, j = 1, \ldots, N$) of the binary matrix has been set to 0 (i.e. black) if $d(i) \leq j \leq u(i)$ and to 1 (i.e. white) otherwise.

The shape of the discrete skin locus S_D depends on the value of N. For $N = 8$, the skin locus is approximated by a rectangle, while for $N = 256$ the discrete skin locus is very close to the skin locus described analytically by the functions d and u. As shown in the experiments described next, the shape of the skin locus influences the classification performance. In general, a too coarse quantisation of the rg space diminishes the discriminative power of the skin locus, providing many false positives. As noted, the memory occupation of S_D is equal to N^2 bits, which for $N = 256$ translates in 65536 bits (i.e. 8 kByte). The MCU employed by our prototype is equipped with 32 kByte of Flash memory, which can thus accommodate to store the bitmap even in the most demanding case, i.e. with $N = 256$.

Using the analytical approach, the classification operation requires the storage and evaluation of the two functions reported in Equations 5 and 6. With the employed MCU, this leads to a 180 us computation time per sample and the use of 48 Bytes of memory. With the discretised skin locus, regardless of it size, the classification of each new sample takes 52 us. Hence, both approaches guarantee real-time performance and meet the memory requirements of the platform, but the discretised approach leads to a 71% reduction in the computation time.

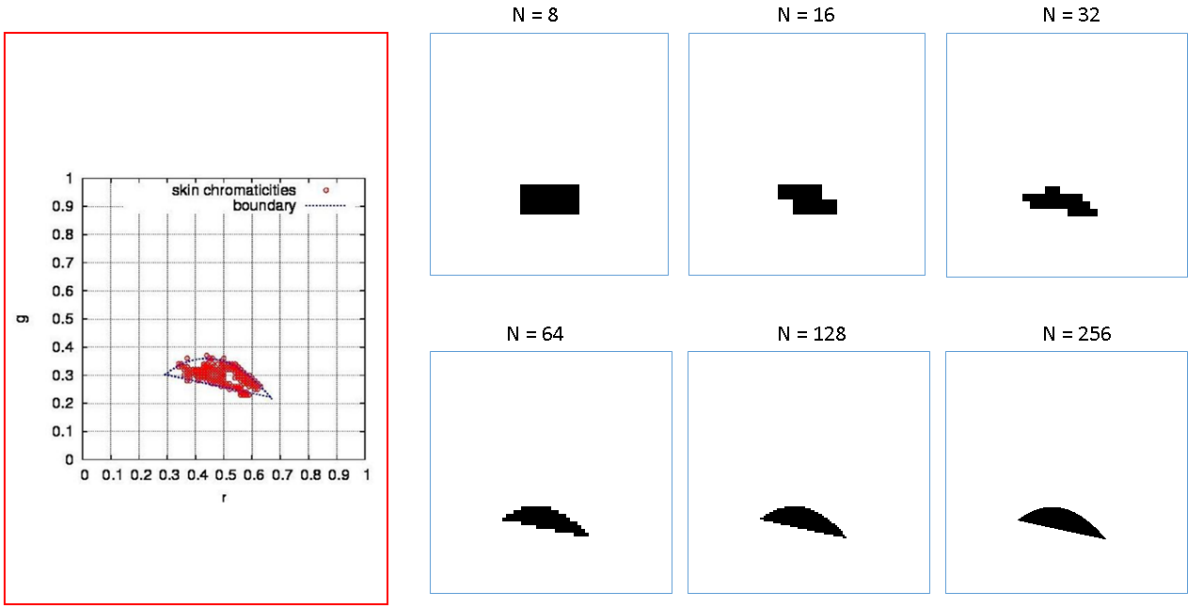

Figure 6: The plot in the red rectangle on left shows the rg data acquired for the computation of the skin locus (red points) and the analytic representation of the skin locus by its boundary (the blue line, see Equations 5 and 6). On right, the discrete representations of the skin locus as binary matrices, for different values of the space quantisation. The binary maps have been rescaled for a better visualisation.

Using the skin detector for contactless machine control and people counting

As already mentioned in Section 1, skin detection is relevant to many applications. Here we discuss two possible uses of our system in smart environments: (1) as a smart switch to control other connected devices, as previously proposed by Lecca *et al.* [6]; and (2) as people counter for statistical tasks.

In the first application, we employ our skin detector for contactless machine control which is required in many smart environments. For example, in smart cities, touchless energy efficient switches may be installed on information kiosks releasing touristic information like e.g. the city maps, the locations of monuments, museums, drugstores, the current status of the main parking, the public transportation timesheets. Specifically, in a museum, smart switches can be used to activate audio guides explaining the history and the meaning of a certain artwork. Currently, the activation of such kind of devices is performed by means of proximity sensors that switch on/off the connected appliances each time something (also not a person) passes in front of it. For instance, in many shops, sliding doors open also when someone approaches to the gates to have a look inside, but is not intentioned to enter. This might causes a waste of energy not only employed to open/close the doors but it could also determine a dysregulation of the shop climate system. Circumscribing the device activation to the presence of human skin helps to reduce this energy waste.

In order to solve this problem, we proposed to use our skin detector in tandem with a proximity sensor as follows. The skin detector is always on, captures continuously the colour signal coming from

a region in front of it and classifies it as skin or non-skin. Each time the incoming signal is labeled as skin, the skin detector wakes up the proximity sensor that measures the distance of the object in front of it. If this distance is in the range 5-10 cm from the optical sensor, the MCU launches a pre-defined command to another appliance connected to it, e.g. it opens a sliding door. The experiments showed that the colour information remarkably improved the performance and the efficiency of a switch based on proximity information only. The results, summarised in Figure 7 , have been obtained from two sets of experiments. In the first set, a colour-checker and the palm of a hand have been illuminated by 8 different lights, similar to those used for the skin locus acquisition. The two squares of the colour-checker representing human skin and the skin tone of the volunteer were always recognised as skin; the 85.23% of the other 22 colours of the colour-checker was correctly classified as non-skin, while a 14.77% was wrongly classified as skin. In the second set of experiments, we positioned the smart switch in 15 different places, characterised by different illuminant conditions, including indoor and outdoor scenarios. In each place, we asked each user to put in front of the switch his/her hand and another non-skin objects randomly chosen in that place. The 93.33% of the human skin chromaticities and the 73.34% of non-skin objects have been correctly classified as skin and non-skin regions respectively, so that we obtained a 26.67% of false positives.

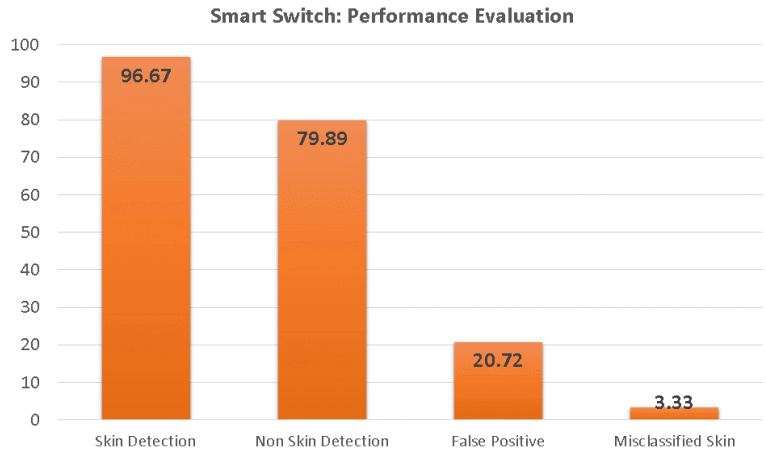

Figure 7: Performance of the colour based smart switch proposed by Lecca et al. *[6].*

Usually such statistical information is collected by counting the number of invoices of the shops or museums, but these data cannot take into account the people visiting the shops without buying something or people inside a museum that visit or not a certain room. A solution is provided by proximity sensors, which can be used to count the times something activates the doors of a shop or of a museum. However, as already highlighted above, proximity sensors provide many false positives and as a consequence noisy data. Our low power system offers a smart solution for accurate, non-intrusive and anonymous people counting.

In this framework, the skin detector is located close to a door or to an object of interest and the RGB sensor is oriented against a dark panel, whole chromaticities fall out of the skin locus under the set of lights used for modeling the skin locus. Each one interested to activate the connected appliance, puts his/her hand between the RGB sensor and the dark panel for some seconds. Each time the skin detector reveals a skin region, the MCU increases an integer counter that is initially set to zero. The usage of the dark panel is a trick to avoid the usage of the proximity sensor in order to further reduce the energy consumption.

We observe that in this second application, people counting is not passive: as before, the system is used to control an appliance (i.e. the door) by a gesture that expresses the intentionality of the users to activate the appliance (i.e. open the door). There are two main differences between the first and the second applications: (1) in the second application, the system has been customised so that to avoid the use of the proximity sensor (i.e. we introduced the dark panel); (2) the MCU firmware has been modified by adding an integer counter.

Set	Time (hour)	Light [RGB] colour	No. of people
1	1 – afternoon	Light 1 [247 255 133]; Natural light + artificial lamp (4000K)	29
2	1.5 – afternoon	Light 2 [236 255 236]; Natural daylight	42
3	1 – morning	Light 3 [221 255 228]; Natural daylight	21
4	1 – morning	Light 4 [255 205 95]; Natural light + artificial lamp (2700 K)	21

Table 1: Specifics of the four sets of rg tones acquired in our institute in two different corridors and under different lights.

To measure the performance of our counter, we selected two corridors of our institute and installed our skin detector there, so that when at rest it captured the RGB colour of a black panel in front of it. We asked each colleague passing there to put the palm of his/her hand in front of the sensor for some seconds, and we captured 5 samples of his/her skin tone. People counting was implemented on the MCU by increasing an integer variable, initially set to zero, each time an acquisition was completed. We collected four sets of rg data, corresponding to four different illuminants, whose RGB colours are reported in Table 1 along with the acquisition time and the number of people sampled. The rg points sampled from human skin in these experiments are plotted in Figure 8. Figure 9-12 show in a coloured square the RGB colour of the used light, the percentage of skin tones correctly classified as skin, and that of the false positives measured on the set of the 22 non-skin surfaces of the colour-checker. The percentage of detected skin and that of the false positives have been measured by using different representations of the skin locus. i.e. the analytic model based on the quadratic functions d and u, and the discrete models obtained for different values of N.

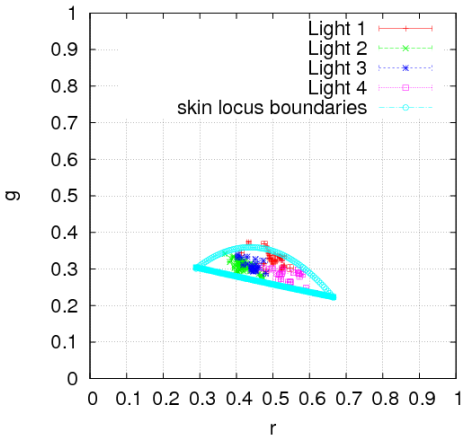

Figure 8: rg data of skin tones acquired under four different illuminants. For each cloud of points, the standard deviation of each chromaticity component C = r, g is about the 5% of the mean value of C.

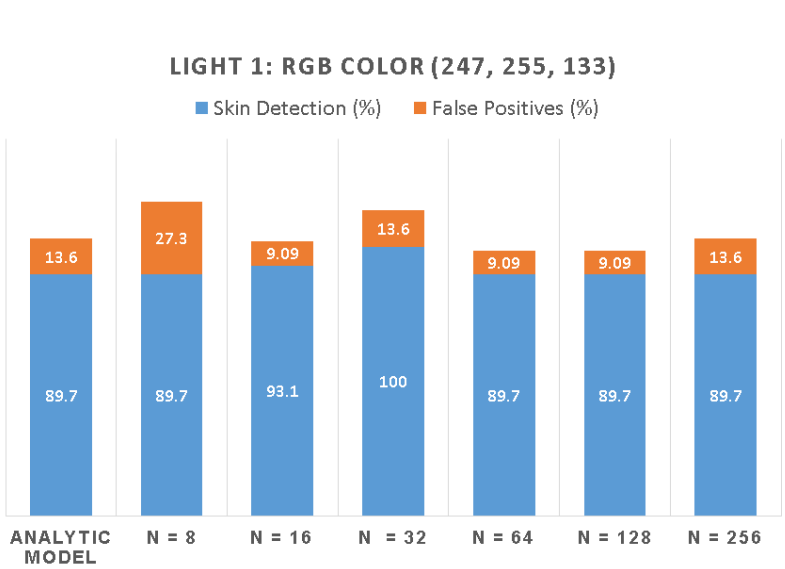

Figure 9: Percentage of skin detection and false positives for Set 1 (Light 1).

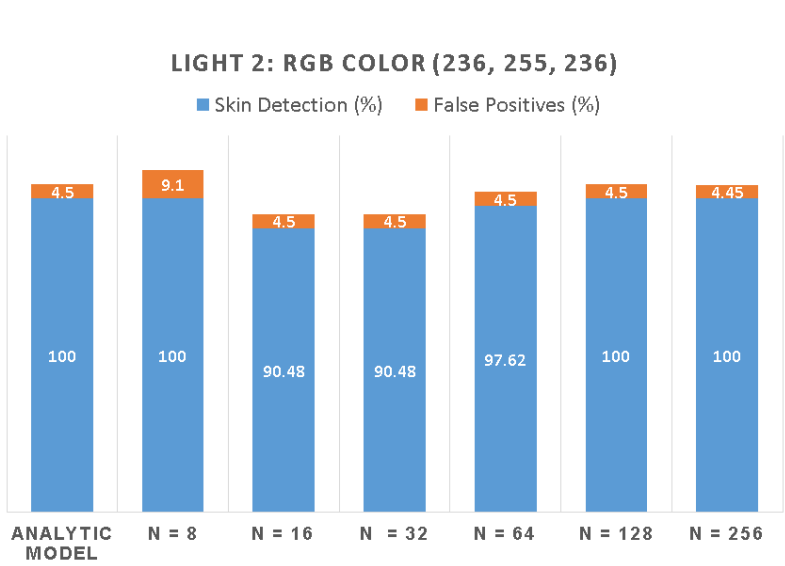

Figure 10: Percentage of skin detection and false positives for Set 2 (Light 2).

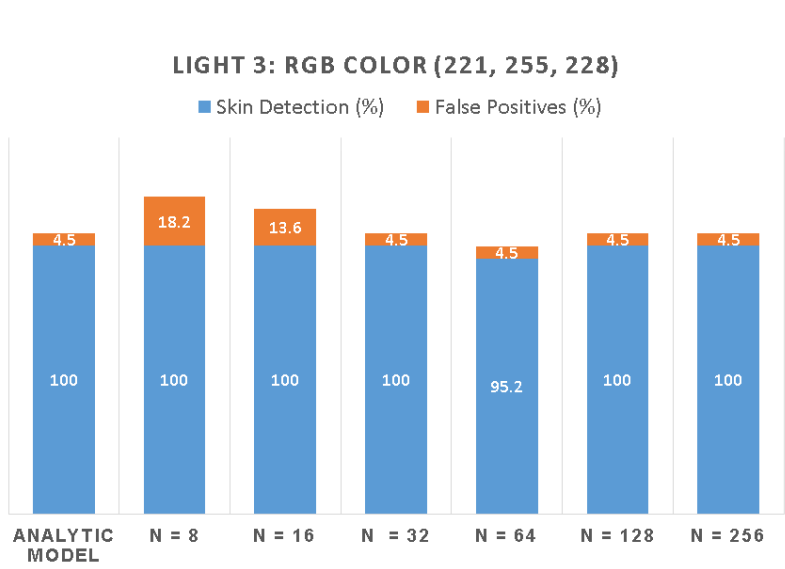

Figure 11: Percentage of skin detection and false positives for Set 3 (Light 3).

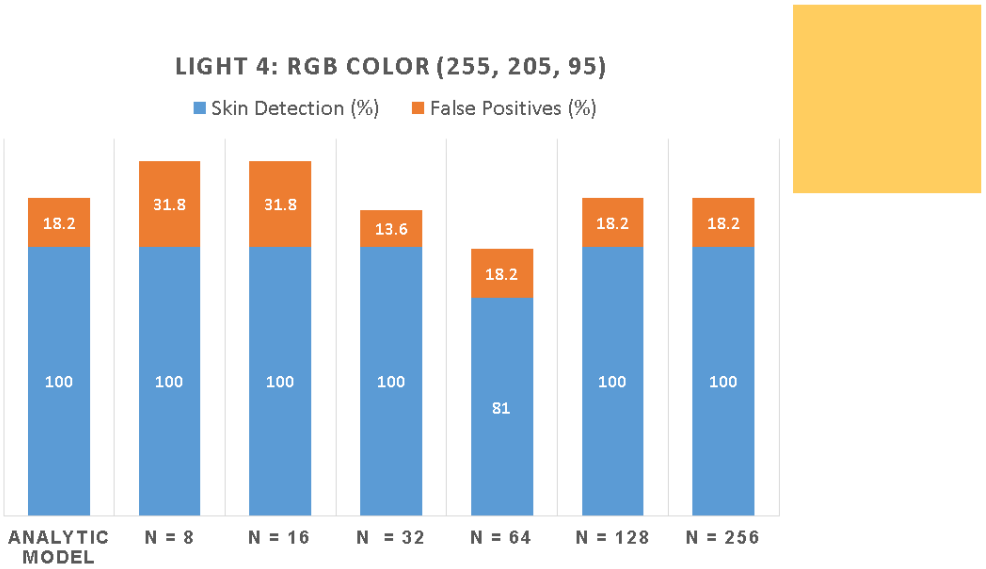

Figure 12: Percentage of skin detection and false positives for Set 4 (Light 4).

Generally, the coarsest discretisation ($N = 8$) poorly represents the skin locus: it allows a high skin recognition rate, but also produces an undesired high percentage of false positives. For $N = 16, 32, 64$ the skin locus borders are poorly approximated, and this adversely affects the classification of the rg skin data close to the skin locus boundary. For $N = 128, 256$, the discrete skin locus is close to the analytic model and the classification performance is the same. Figure 13 shows the rg data collected in Set 3, under the illuminant with RGB colour (221, 255, 228). The rg skin points are drawn against the discrete skin loci for $N = 8, 16, 32, 64, 128, 256$.

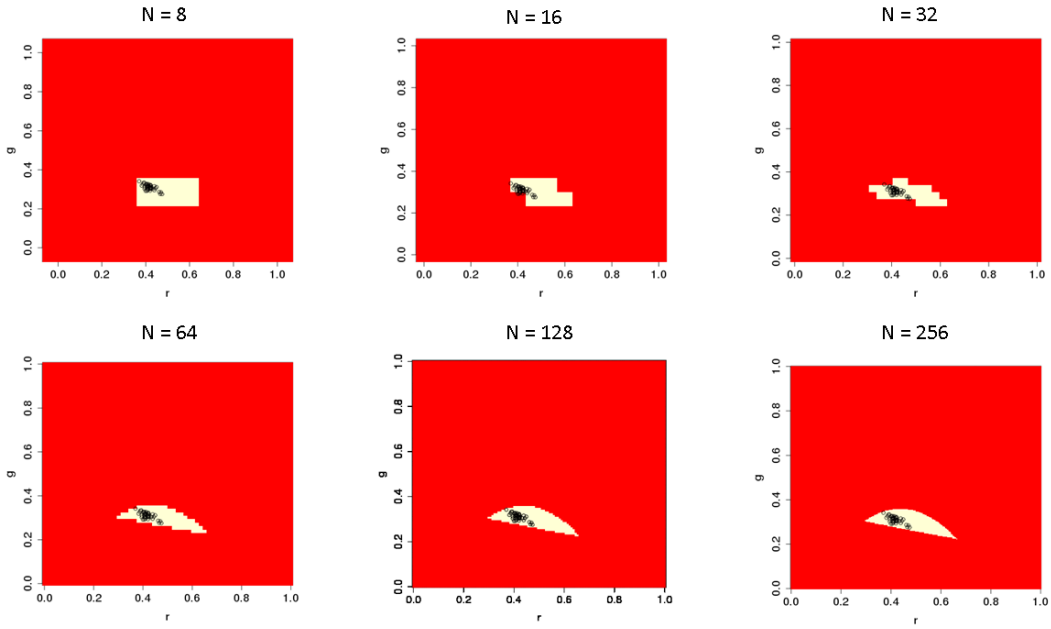

*Figure 13: For **N** = 16, 32, 64, the rg points of skin tones fall out of the skin locus. For **N** = 8, the points fall in the skin locus, but the region also contain many non-skin chromaticities.*

For the highest space quantisation ($N = 256$), the mean skin detection percentage is 97.4%, while the percentage of mean false positive is 6.7%. These results are encouraging: the total number of people sampled in the corridors we monitored is 113, and only three of them have been missed. We observed

that, beside one person whose skin tone was on the skin locus boundary, the other two not recognised cases were due to a not correct location of the hand in front of the sensor, i.e. the users puts fingers of an open hand instead of the palm, so that part of the scene background was in the RGB sensor's field of view as well.

We used the collected data also to measure the performance of the system as a smart switch with the additional functionality of people counting. In this framework, we imagined to remove the black panel and to equip the system with a proximity sensor. As reported above, the mean skin detection performance is the 97.4%, while the 2.6% of the skin has been missed. The mean value of the false positive percentage measured on the 22 non-skin surfaces of the colour-checker under the four lights is the 10.2%. As already found in the experiments described in [6] and reported above, the improvement on the machine control provided by our system with respect to a proximity sensor is remarkable: joining colour and distance avoided the activation of the connected appliance in the 98.9% of the cases in which a non-skin region has been positioned in front of the system.

Finally, Figure 14 reports an example of statistical analysis: in this case, we computed the histogram of the people detected by the skin detector during the acquisition of the second set of data (Set 2 in Table 1).

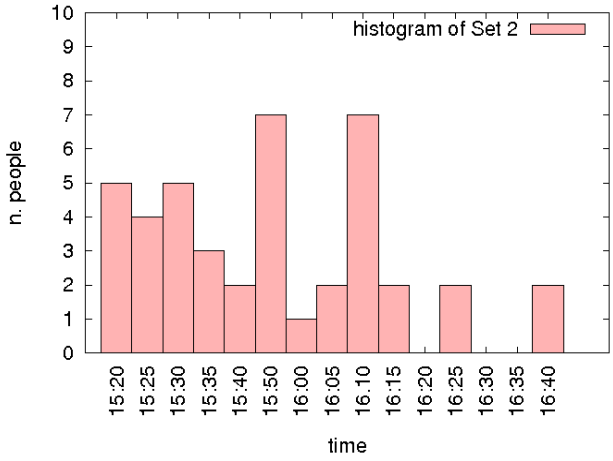

Figure 14: Histogram of the number of people detected in Set 2 at different times. The labels on the x-axis report the time (in PM format).

Conclusions

Smart devices, capable to work continuously with high performance and with reduced energy consumption, are the main components of the future incoming world, where hardware and software co-designed technologies are expected to be always on and ready to receive, process and transmit information to improve the human life. In this framework, visual devices, like smart cameras and embedded vision systems play a crucial role in many applications [40]. In this work, we focused on the human skin detection, that is often the first step to localise, recognise and/or track a person in a certain ambient or to recognise his/her gesture, behavior, or emotions. These information are highly relevant for several different tasks, such as: controlling and monitoring the presence of people in a certain place for security reasons, understanding the intention of someone from its behavior to help him/her in life assisted living, controlling appliances by simple, natural gestures of the hands in order to simplify the human-machine. Here we described an optimised version of an embedded visual system firstly presented by Lecca *et al.* [6]. The system implements skin detection in an energy efficient way through

a smart architecture that distributes the computational load of the skin detection algorithm on both hardware and software. The optimisation with respect to its previous version regards the skin classification algorithm that has been re-written in a discrete space to lower its computational complexity. We discussed two different applications of this system in smart environments, i.e. as a smart switch for touchless control of appliances, as already proposed by Lecca *et al.* [6], and as an automatic people counter for statistical analysis. The experiments showed good performance of the system in terms both of energy consumption and accuracy of the skin detection.

Acknowledgement

The authors would like to thank all the people participating to our experiments.

References

1. Charlton S (2009), *Ubiquitous Computing: Smart Devices, Environments and Interactions*, New Jersey: John Wiley.
2. Das SK and Cook DJ (2006), Designing and modeling smart environments, *Proceedings of the 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks*, 490-494, Washington, DC, USA.
3. Vermesan O and Friess P (2013), *Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems*, River Publishers.
4. Yen T-Y and Wolf W (2013), *Hardware-software Co-synthesis of Distributed Embedded Systems*, Springer Science & Business Media.
5. Marwedel P (2010), *Embedded System Design: Embedded Systems Foundations of Cyber-physical Systems*, Springer Science & Business Media.
6. Lecca M, Gottardi M, Farella E, Milosevic B and Bilal M (2015), A low power color sensor for illuminant invariant skin detection, Proceedings of the XI Colour Conference - Colour and Colorimetry, **XI B**, 43-54, Milano, Italy.
7. Soriano M, Martinkauppi B, Huovinen S and Laaksonen M (2000), Using the skin locus to cope with changing illumination conditions in color-based face tracking, *Proceedings of the IEEE Nordic Signal Processing Symposium*, **38**, 383-386. Kolmarden, Sweden.
8. Ban Y, Kim SK, Kim S, Toh KA and Lee S (2014), Face detection based on skin color likelihood, *Pattern Recognition*, **47** (4), 1573-1585.
9. Li SZ and Jain AK (2005), *Handbook of Face Recognition*, Springer-Verlag New York Inc., Secaucus, NJ, USA.
10. Sigal L, Sclaroff S and Athitsos V (2004), Skin color-based video segmentation under time-varying illumination, *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26** (7), 862-877.
11. Sultana M and Gavrilova ML (2014), Face recognition using multiple content–based image features for biometric security applications, *International Journal of Biometrics*, **6** (4), 414-434.
12. Malima A, Ozgur E and Çetin M (2006), A fast algorithm for vision-based hand gesture recognition for robot control, *Proceedings of the IEEE on Signal Processing and Communications Applications*, 1-4.
13. Karishma SN and Lathasree V (2014), Fusion of skin color detection and background subtraction for hand gesture segmentation, *International Journal of Engineering Research and Technology*, **3** (2), 1835-1839.
14. Wei G and Sethi IK (1999), Face detection for image annotation, *Pattern Recognition Letters - Special issue on pattern recognition in practice VI*, **20** (11-13), 1313-1321.
15. Lecca M, Gasparini L and Gottardi M (2014), Ultra-low power high-dynamic range color pixel embedding RGB to r-g chromaticity transformation, *Proceedings of the SPIE Photonics Europe*, **9141**, 914107-914107-6, Brussels, Belgium.
16. Soriano M, Martinkauppi B, Huovinen S and Laaksonen M (2000), Skin detection in video under changing illumination conditions, *Proceedings of the 15th International Conference on Pattern Recognition*, **1**, 838-842, Barcelona, Spain.
17. Khan R, Hanbury A, Stöttinger J and Bais A (2012), Color based skin classification, *Pattern Recognition Letters*, **33** (2), 157-163.
18. Huang L, Ji W, Wei W, Chen B-W, Yan CC, Nie J, Yin J and Jiang B (2015), Robust skin detection in real-world images, *Journal of Visual Communication and Image Representation*, **29**, 147-152.
19. Xu Z-W and Zhu M-I (2007), Color-based skin detection: a survey, *Journal of Image and Graphics*, **12** (3), 377-388.

20. Kawulok M, Nalepa J and Kawulok J (2014), Skin detection and segmentation in color images, in *Advances in Low-Level Color Image Processing*, Celebi ME and Smolka B (eds.), 329-366, Springer: Netherlands.
21. Hsu R-L, Abdel-Mottaleb M and Jain A (2002), Face detection in color images, *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24** (5), 696-706.
22. Gevers T, Gijsenij A, Weijer Jvd and Geusebroek J-M (2012), *Color in Computer Vision: Fundamentals and Applications*, Wiley Publishing.
23. Albiol A, Torres L and Delp E (2011), Optimum color spaces for skin detection, *Proceedings of the International Conference on Image Processing*, 122-124, Brussels, Belgium.
24. Aznaveh M, Mirzaei H, Roshan E and Saraee M (2008), A new color based method for skin detection using RGB vector space, *Proceedings of the 2008 Conference on Human System Interactions*, 932-935, Krakow, Poland.
25. Hashem H (2009), Adaptive technique for human face detection using HSV color space and neural networks, *Proceedings of the National Radio Science Conference*, 1-7, New Cairo, Egypt.
26. Zeng HLG (2012), A new method for skin color enhancement, *Proceedings of the SPIE Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications*, 82920K-82920K-9, Burlingame, CA, USA.
27. Yang M-H and Ahuja N (1999), Gaussian mixture model for human skin color and its applications in image and video databases, *Proceedings SPIE Storage and Retrieval for Image and Video Databases VII*, 458-466, San José, CA, USA.
28. Terrillon J-C, Shirazi M, Fukamachi H and Akamatsu S (2000), Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 54-61, Grenoble, France.
29. Jones M and Rehg J (1999), Statistical color models with application to skin detection, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 280, Fort Collins, CO, USA.
30. Bianco S, Gasparini F and Schettini R (2013), Computational strategies for skin detection, *Proceedings of the Computational Color Imaging Workshop (CCIW'13)*, 199-211, Chiba, Japan.
31. Kakumanu P, Makrogiannis S and Bourbakis N (2007), A survey of skin-color modeling and detection methods, *Pattern Recognition*, **40** (3), 1106-1122.
32. Storring M, Andersen H and Granum E (2001), Physics-based modelling of human skin colour under mixed illuminants, *Robotics and Autonomous Systems*, **35** (3), 131-142.
33. Lecca M (2014), On the von Kries model: estimation, dependence on light and device, and application, *Advances in Low-Level Color Image Processing*, Celebi ME and Smolka B (eds.), 95-135, Springer: Netherlands.
34. Etienne-Cummings R, Pouliquen P and Lewis MA (2003), A vision chip for color segmentation and pattern matching, *EURASIP Journal on Advances in Signal Processing*, **7**, 703-712.
35. http://www.st.com/content/st_com/en/products/imaging-and-photonics-solutions/proximity-sensors/vl6180x.html [last accessed 12 July 2016]
36. http://www.hamamatsu.com/ [last accessed 12 July 2016]
37. http://www.atmel.com/devices/atmega328p.aspx [last accessed 12 July 2016]
38. https://www.arduino.cc/en/Main/ArduinoBoardUno [last accessed 12 July 2016]
39. https://processing.org/ [last accessed 12 July 2016]
40. Belbachir AN (2010, ed.), *Smart Cameras*, Springer.