Expert Systems With Applications 182 (2021) 115287

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Expert Systems with Applications
An International Journal
Editor-in-Chief
Binshan Lin

People re-identification using depth and intensity information from an overhead camera

Carlos A. Luna, Cristina Losada-Gutiérrez [*], David Fuentes-Jimenez, Manuel Mazo

Department of Electronics, University of Alcalá, Ctra. Madrid-Barcelona, km. 33600, 28805 Alcalá de Henares, Spain

ARTICLE INFO

Keywords:
People re-identification
Depth information
Intensity images
Overhead depth camera
Real-time

ABSTRACT

This work presents a new people re-identification method, using depth and intensity images, both of them captured with a single static camera, located in an overhead position. The proposed solution arises from the need that exists in many areas of application to carry out identification and re-identification processes to determine, for example, the time that people remain in a certain space, while fulfilling the requirement of preserving people's privacy. This work is a novelty compared to other previous solutions, since the use of top-view images of depth and intensity allows obtaining information to perform the functions of identification and re-identification of people, maintaining their privacy and reducing occlusions. In the procedure of people identification and re-identification, only three frames of intensity and depth are used, so that the first one is obtained when the person enters the scene (frontal view), the second when it is in the central area of the scene (overhead view) and the third one when it leaves the scene (back view). In the implemented method only information from the head and shoulders of people with these three different perspectives is used. From these views three feature vectors are obtained in a simple way, two of them related to depth information and the other one related to intensity data. This increases the robustness of the method against lighting changes. The proposal has been evaluated in two different datasets and compared to other state-of-the-art proposal. The obtained results show a 96,7% success rate in re-identification, with sensors that use different operating principles, all of them obtaining depth and intensity information. Furthermore, the implemented method can work in real time on a PC, without using a GPU.

1. Introduction

People re-identification consists of determining if two persons in different images, taken using a different camera or with a different point of view, correspond to the same individual or not. This task has attracted the attention of the scientific community in recent years, emerging an increasing number of proposals based on different approaches (Zheng, Yang, & Hauptmann, 2016; Islam, 2020), since it is a fundamental task in different applications such as video-surveillance (Vezzani, Baltieri, & Cucchiara, 2013) or customer behavior analysis (Merad, Aziz, Iguernaissi, Fertil, & Drap, 2016; Liciotti, 2017). However, it is still and open and challenging problem due to changes in the appearance of people caused by lighting changes, occlusions, unconstrained poses or camera point of view variations.

The first works in this topic were based on two main steps: feature extraction and classification using different metrics to compare people detected by different cameras (Bedagkar-Gala & Shah, 2014). In these works, the feature vectors are usually based on people appearance (D'Angelo & Dugelay, 2011; Satta, 2013; de Carvalho Prates & Schwartz, 2015; Liao, Hu, Xiangyu Zhu, & Li, 2015; Mingyong Zeng, Wu, Tian, Lei Zhang, & Lei Hu, 2015; Matsukawa, Okabe, Suzuki, & Sato, 2016; Devyatkov, Alfimtsev, & Taranyan, 2018), being widely used features based on color information.

In this context, the authors of (D'Angelo & Dugelay, 2011) propose a Probabilistic Color Histogram (PCH) descriptor that is then classified using a fuzzy K-Nearest Neighbours (KNN) for people re-identification in a video-surveillance scenario. Similarly, the proposal in (de Carvalho Prates & Schwartz, 2015) describes a Color-based Ranking Aggregation (CBRA) method for combining different color features in a descriptor. Covariance-based feature descriptors are also used to fuse color and gradient information in (Devyatkov et al., 2018; Zeng et al., 2015). Thus, the authors of (Zeng et al., 2015) also propose a new coding based Multi-

[*] Corresponding author.
E-mail addresses: carlos.luna@uah.es (C.A. Luna), cristina.losada@uah.es (C. Losada-Gutiérrez), d.fuentes@edu.uah.es (D. Fuentes-Jimenez), manuel.mazo@uah.es (M. Mazo).

https://doi.org/10.1016/j.eswa.2021.115287
Received 12 February 2021; Received in revised form 10 April 2021; Accepted 23 May 2021
Available online 30 May 2021

0957-4174/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

shot method named CRC-S (CRC in Subtraction form) to compare the obtained descriptors based for people re-identification. Furthermore, there are several approaches for dealing with lighting change. In (Liao et al., 2015), it is proposed a new feature representation for people re-identification called Local Maximal Occurrence (LOMO), and a subspace and metric learning method named Cross-view Quadratic Discriminant Analysis (XQDA). Based in the same idea than LOMO, the authors of (Matsukawa et al., 2016) describe each local path color and texture using a set of Gaussians, proposing a novel region descriptor based on hierarchical Gaussian distribution of pixel features, and test two widely used metrics: XQDA (Liao et al., 2015) and the KISS Metric learning (KISSME) (Kostinger, Hirzer, Wohlhart, Roth, & Bischof, 2012). However, these appearance descriptors can change quickly since people may variate their clothes, thus they only work for people re-identification during short periods of time.

In recent years, there have appeared the RGB-D cameras (Smisek, Jancosek, & Pajdla, 2011; Sell & O'Connor, 2014) that provides both, color and depth information (distance from each point to the camera) from the environment. Thus, RGB-D data allow obtaining a 2.5D point cloud (only the objects in front of the camera appears in the point cloud). In addition, these cameras also provide an intensity image with information about the received infrared (IR) intensity at each pixel. The emergence of these cameras has led to an important number of proposals that use these RGB-D data for people re-identification (Barbosa, Cristani, Del Bue, Bazzani, & Murino, 2012; Baltieri, Vezzani, & Cucchiara, 2015; Gharghabi, Shamshirdar, Shangari, & Maroofkhani, 2015; Pala, Satta, Fumera, & Roli, 2016; Patruno, Marani, Cicirelli, Stella, & D'Orazio, 2019). Most of these works also includes the two previously mentioned stages: first there are extracted 3D feature descriptors that are then classified to detect if they correspond to a previously detected person or to a new one. The authors of (Pala et al., 2016) propose combining clothing appearance descriptors extracted from RGB images with anthropometric measures extracted from depth data, based on the Multiple Component Dissimilarity (MCD) representation. Other works use skeleton data obtained from detected people (Baltieri et al., 2015; Gharghabi et al., 2015; Patruno et al., 2019), thus Gharghabi et al. (2015) introduce the novel VHF 3D descriptor of the body shape combined with skeleton data for people re-identification, that is invariant to color and lighting changes. Similarly, Baltieri et al. (2015) propose combining 3D skeleton data with color and gradient histograms, reducing the effects of occlusions, partial views or pose changes, whereas in (Patruno et al., 2019), there is built a person signature from its skeleton standard posture (SSP). Most of the aforementioned proposals work well under controlled conditions, but they fail in complex environments, with multiple people and occlusions. To reduce the occlusions, there are other proposals in which the camera is located in a top-view configuration (Liciotti, 2017; Liciotti, Paolanti, Frontoni, Mancini, & Zingaretti, 2017; Paolanti et al., 2018). It is noteworthy that descriptors based on shapes and dimensions of the skeleton have the drawback of the errors in 3D measurements with current depth sensors.

As in other fields, recently, numerous deep-learning alternatives for people re-identification have emerged (Wu et al., 2019; Ye et al., 2020). These works include supervised (Ahmed, Jones, & Marks, 2015; Chen, Zhu, & Gong, 2017; Chung, Tahboub, & Delp, 2017; Li, Zhao, Xiao, & Wang, 2014; Qian, Fu, Jiang, Xiang, & Xue, 2017; Schumann & Stiefelhagen, 2017), semipervised (Xin et al., 2019) and unsupervised (Chen, Zhu, & Gong, 2018; Li, Zhu, & Gong, 2018; Wang, Zhu, Gong, & Li, 2018) approaches. The use of deep learning techniques for many applications presents two fundamental drawbacks: (a) the need to have a large number of images for training, that may not be available for the re-identification task; (b) although there are proposals to reduce computation time (Satta, Fumera, & Roli, 2012; Wang, Gong, Cheng, & Hou, 2020), it is usually high or the approaches require specific hardware resources.

To reduce the problems raised above, the proposal described in this paper use a camera located in an overhead position. In addition, the use of overhead information, specifically depth and intensity images, enables the method to be used in applications where it is required to preserve the privacy of people and reduces the occlusions. Furthermore, the use of depth data increases the robustness against lighting changes. This proposal is based on a classic method with ad hoc descriptors and uses only three depth and intensity frames to carry out re-identification, reducing computational cost.

In what follows, the structure of the paper is: Section 2 describes the proposed solution, then Section 3 includes the experimental setup, the obtained results and discussion, and finally, Section 4 presents the main conclusions and some ideas for future work.

2. Proposed solution

The proposal described in this work is based on the use of a single static camera located in an overhead position, which allows to obtain both depth and intensity images. Due to the overhead location of the camera, the information in both types of images (depth and intensity) are related to the most relevant parts of a top-view of a person, i.e., the head and shoulders.

First, people detection is performed from depth images, following the procedure described in (Luna et al., 2017), after that, feature extraction is carried out. For each person who comes into the scene, there are computed the following characteristics: the person height, a depth feature vector and an intensity feature vector for their identification and re-identification. To increase the robustness of the proposal to the changes of the appearance depending on the position of the person and the camera, for each person who comes into the scene these characteristics are calculated for three different person positions: frontal, overhead and back. Therefore, three mean person height, three depth feature vectors, and three intensity feature vectors are used in the identification and re-identification processes. In the identification process, there are obtained the features for each person who comes into scene (three mean person height, three depth and three intensity feature vectors) and them are saved into a Dataset. Whereas, in the re-identification process, for each person that leaves the room, the three mean person height, three depth and three intensity feature vectors are obtained and compared with all the previously saved into the Dataset. The person leaving the room is identified as the person in the Dataset whose characteristics are most similar.

A general block diagram of the proposal is shown in Fig. 1. There are two main stages: the identification and the re-identification processes. In the identification (ID) process, there are obtained the feature vectors from the depth and intensity images of the people that enter into the room (Person in), an saved into a Dataset. Then, in the re-identification process, for each user that leaves the room (Person out), the depth and intensity feature vectors are obtained and compared with all the previously saved feature vectors, that are included in the Dataset during the identification stage. As it can be seen in Fig. 1, the three first steps are the same for the identification and re-identification processes, whereas the last one is different. Besides, in both processes, the feature vectors are determined from three different images of each user crossing the scene: the first one is acquired when the user comes into the scene (frontal view), the second when in the center (overhead view) and the third when the user leaves (back view).

An example of the depth and intensity images of the same person in the three positions and in both directions is shown in Fig. 2, where the three positions are indicated in the depth image with red, yellow and blue lines. It is important to highlight that all these images have been obtained from a single static camera in overhead position, which has captured the same person in three positions of the scene, which are named as frontal, overhead and back. As it can be seen in Fig. 2, since the viewing angles with which the images have been acquired are small, there is no information on people's faces, being not possible to recognize them. This allows performing people re-identification while preserving their privacy.

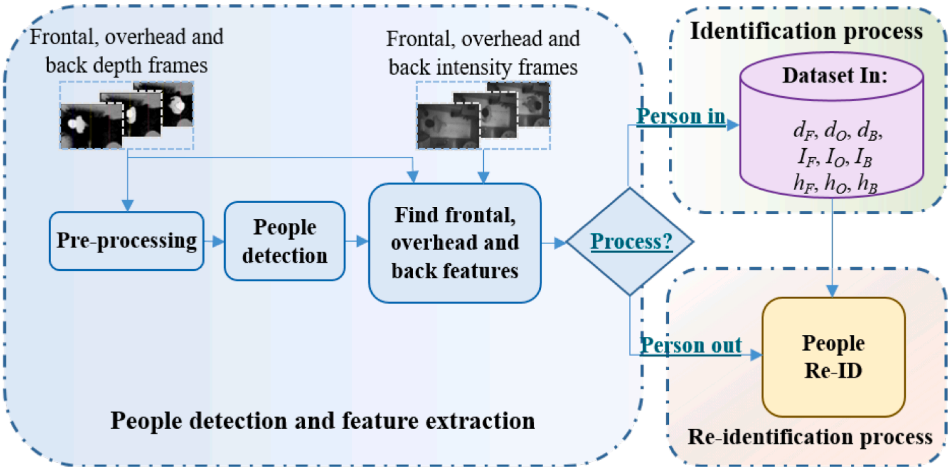

Frontal, overhead and back depth frames
Frontal, overhead and back intensity frames
Pre-processing
People detection
Find frontal, overhead and back features
Process?
Person in
Person out
Identification process
Dataset In:
$d_F, d_O, d_B,$
I_F, I_O, I_B
h_F, h_O, h_B
People Re-ID
People detection and feature extraction
Re-identification process

Fig. 1. General block diagram of the identification and re-identification procedure.

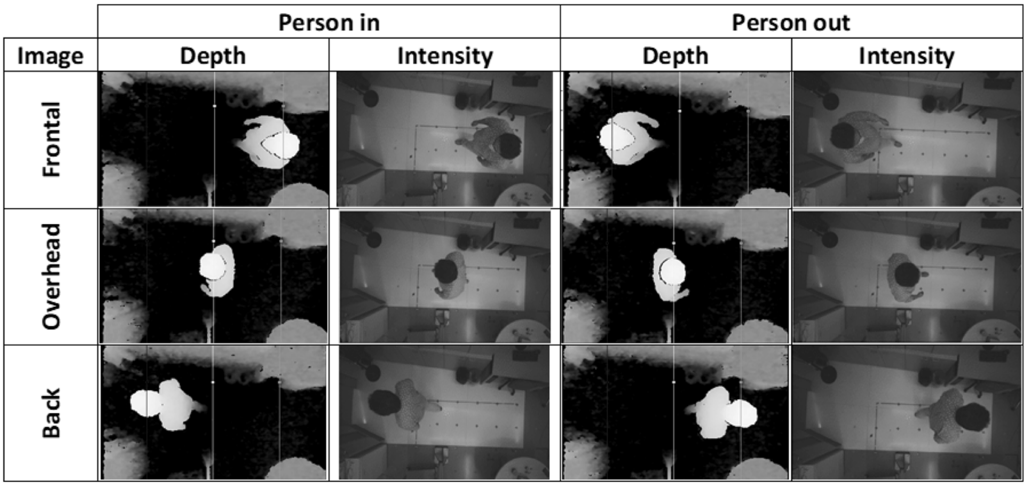


Fig. 2. Example of the depth and intensity frames of a person in the three selected positions.

Each of the stages shown in Fig. 1 are detailed in the following sections.

2.1. Pre-processing

In order to reduce the noise in depth images we have implemented a noise reduction algorithm that includes three steps:

1. The depth image is transformed to a height image, so each pixel represents the height from the floor, instead of the distance to the camera. To do that, each pixel of the height image is computed as the height of the camera with respect to the ground minus the value of the pixel in the depth image. Then, all pixels with values under $100\,cm$ or over $220\,cm$ are removed (assigning $height = \ 0\,cm$), assuming that this is the people height range to be re-identified. This range has been set taking into account anthropometric characteristics of the human body (Matzner et al., 2015), since the height of adult people must be included in this range. Thus all the adults, and even most of the children must be correctly detected.
2. The background is removed using the procedure described in (Luna, Losada-Gutiérrez, Fuentes-Jiménez, & Mazo, 2021).
3. A mean filter (of 3×3 elements) is used to smooth the object surfaces.

Invalid pixels are not taken into account neither in the background extraction nor in the mean filter. Fig. 3 shows an example of an image after each of the pre-processing steps of depth and intensity images, especially in the case of depth images.

2.2. People detection

The detection of the people present in the scene is carried out from the pre-processed depth images. Since most of the objects in the scene are removed in the pre-processing stage and only small portions of them remain, the people detection process can be considerably simplified. We have carried out the people detection using the regions of interest (ROIs) estimation procedure described in (Luna et al., 2017), choosing the object with the largest area as the person, since we assume that there is only one person in each frame. It is worth highlighting that the proposal in (Luna et al., 2017) is able to detect people, as well as the pixels that corresponds to each person, with high accuracy (over 99%) even if there are multiple people in the scene and they are close to each other, so the re-identification can be correctly performed for each person in the scene if there are multiple people.

Regardless of the sensor used and the height where it is located, in this procedure we perform the following steps:

1. Divide the height image into subregions of 20×20 pixels.
2. Find the mean value of the set of valid height measurements for each subregion

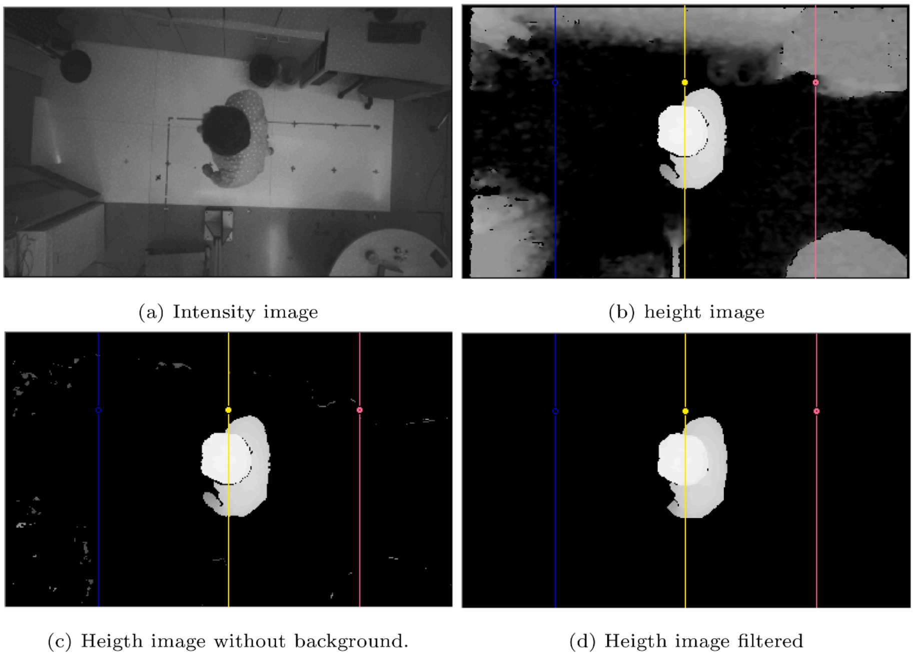

(a) Intensity image
(b) height image
(c) Heigth image without background.
(d) Heigth image filtered

Fig. 3. Sample images showing the results after each of the pre-procesing steps.

3. Determine the subregion with the highest mean value, and take this mean value as the maximum person height (h).
4. Determine the set (τ) of subregions that belong to the person.

2.3. *Feature extraction*

In this stage, there are obtained depth and intensity features from those pixels associated with the person head, neck and shoulders. For each person who comes into the scene, there is computed a feature vector that includes a mean person height ($\overline{h}$) and six vectors, three of depth (**d**) and other three of intensity (**I**). These characteristics are calculated from depth and intensity images captured from three different positions (frontal, overhead and back) of the person in the scene.

The value of $\overline{h}$ is obtained as the average of the three values of h obtained in the frontal (h_F), overhead (h_O) and back (h_B) positions, as shown in Eq. 1.

$$\overline{h} = \frac{h_F + h_O + h_B}{3} \tag{1}$$

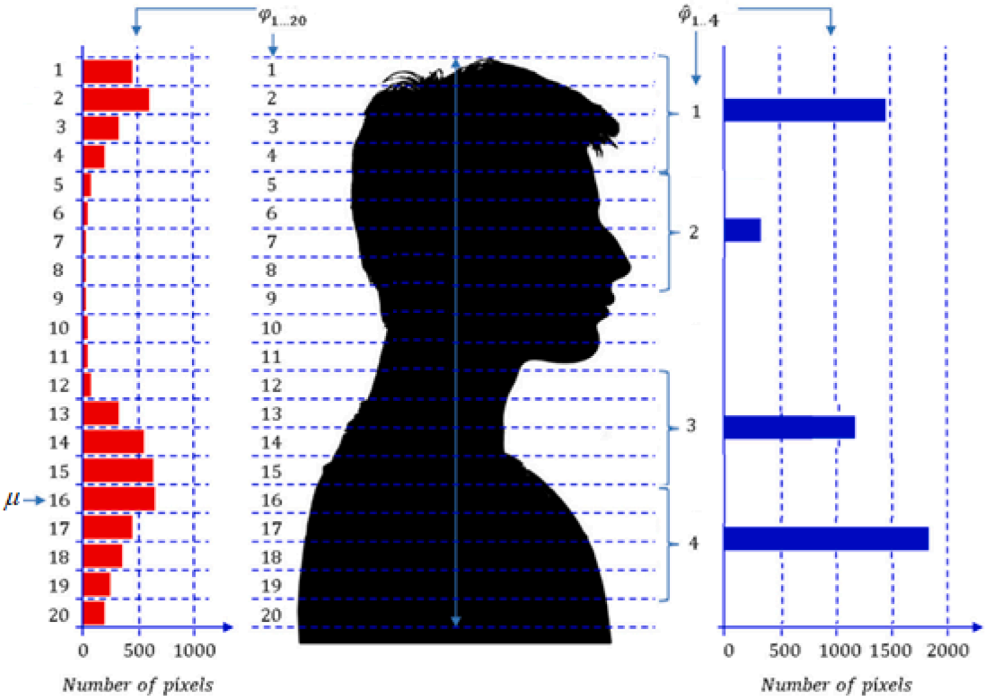


Fig. 4. Representation of slice segmentation for a person. The number of valid pixels in each slice (φ) is shown in the left. The values for the 4 components of the depth feature vector ($\widehat{\varphi}_i$) are shown in the right.

Each of the depth feature vectors $\mathbf{d}_P$ (where $P = F, O and B$ for Frontal, Overhead and Back position respectively) is composed of 4 elements:

$$d_F = [\widehat{\varphi}_{1F} \quad \widehat{\varphi}_{2F} \quad \widehat{\varphi}_{3F} \quad \widehat{\varphi}_{4F}]^T$$

$$d_O = [\widehat{\varphi}_{1O} \quad \widehat{\varphi}_{2O} \quad \widehat{\varphi}_{3O} \quad \widehat{\varphi}_{4O}]^T \tag{2}$$

$$d_B = [\widehat{\varphi}_{1B} \quad \widehat{\varphi}_{2B} \quad \widehat{\varphi}_{3B} \quad \widehat{\varphi}_{4B}]^T$$

Since the procedure for obtaining the components $\widehat{\varphi}_{iP} \quad (i = 1, \ldots, 4)$ is the same for all positions P, in what follows we will call them as $\widehat{\varphi}_i$. The steps for computing these components are detailed below:

1. From the maximum person height h, there are created 20 slices $s \quad (s = 1, \ldots, 20)$ of heights of $2\,cm$, that include the number of valid pixels at different heights. In the left part (red bars) of Fig. 4 there is shown an example. Taking the silhouette of a person as a reference, red bars on the left represent the number of valid depth measurements in each of the 20 slices. Besides, the components $\widehat{\varphi}_i$ are represented by blue bars in the right part of Fig. 4.
2. The pixels included in the subregions belonging to τ are assigned to the corresponding slice s. It is worth highlighting that the first slice $(s = 1)$ also includes pixels with a height greater than h, since it is the average value of the corresponding subregion.
3. For each s, the number of pixels is counted, being φ_S the number of pixels of slide s, that provides information about the pixel density for the identification of the head and shoulders areas of a person.
4. $\widehat{\varphi}_1$ and $\widehat{\varphi}_2$ are related to the head and they are obtained as shown in Eq. 3.

$$\widehat{\varphi}_1 = \sum_{s=1}^{s=4} \varphi_s \quad and \quad \widehat{\varphi}_2 = \sum_{s=5}^{s=8} \varphi_s \tag{3}$$

5. $\widehat{\varphi}_3$ and $\widehat{\varphi}_4$ are related to the shoulders and they are obtained as shown in Eq. 4.

$$\widehat{\varphi}_3 = \sum_{s=\mu-4}^{s=\mu-1} \varphi_s \quad and \quad \widehat{\varphi}_4 = \sum_{s=\mu}^{s=\mu+3} \varphi_s \tag{4}$$

where μ is the slice between φ_{11} and φ_{17} with the maximum number of valid pixels. This maximum value coincides with the upper part of the shoulders, which due to the biometric characteristics of an adult person must be in this range of heights.

The features related to the intensity image include 200 components. These are the normalized histograms of the intensity values corresponding to the valid pixels $\mathbf{I}_P$ (where the subscript $P = F, O, B$ is used to identify the position of each frame, as in the depth images). The vectors, obtained in frontal (I_F), overhead (I_O) and back (I_B) positions, are defined as:

$$\begin{aligned} I_F &= [\phi_{1F} \quad \phi_{2F} \quad \phi_{3F} \quad \phi_{4F}]^T \\ I_O &= [\phi_{1O} \quad \phi_{2O} \quad \phi_{3O} \quad \phi_{4O}]^T \end{aligned} \tag{5}$$

$$I_B = [\phi_{1B} \quad \phi_{2B} \quad \phi_{3B} \quad \phi_{4B}]^T$$

Each component is described in Eq. 6:

$$\phi_{iF} = \frac{H_{iF}}{\widehat{\varphi}_{iF}}, \quad \phi_{iO} = \frac{H_{iO}}{\widehat{\varphi}_{iO}}, \quad \phi_{iB} = \frac{H_{iB}}{\widehat{\varphi}_{iB}} \tag{6}$$

where $H_{iF} = [b_{i1F} b_{i2F} \ldots b_{i50F}]$, $H_{iO} = [b_{i1O} b_{i2O} \ldots b_{i50O}]$, $H_{iB} = [b_{i1B} b_{i2B} \ldots b_{i50B}]$ and the b_{ijF}, b_{ijO} and b_{ijB} $(i = 1, \ldots, 4, \quad j = 1, \ldots, 50)$ are the 50 bins of the intensity histogram, where invalid pixels are discarded as they can add uncertainty to intensity values.

2.4. People re-identification

In the re-identification process, for each person that leaves the scene, there are computed the Euclidean distances between feature vectors and $\overline{h}$ of this person and feature vectors and $\overline{h}$ of the N people previously registered in the Dataset *Person In*, as shown in Eq. 7.

$$\begin{aligned} \Delta d_{p(n)} &= |d_{Pout} - d_{Pin(n)}| \\ \Delta \mathbf{I}_{p(n)} &= |\mathbf{I}_{Pout} - \mathbf{I}_{Pin(n)}| \end{aligned} \tag{7}$$

$$\Delta \overline{h}_{(n)} = \left| \overline{h}_{out} - \overline{h}_{in(n)} \right|$$

where $d_{Pin(n)}, I_{Pin(n)}$ and $\overline{h}_{in(n)}, (n = 1, 2, \ldots N \quad and \quad P = F, O, B)$ are the vectors and $\overline{h}$ associated with the person n in the Dataset and $d_{Pout}, \mathbf{I}_{Pout}$ and $\overline{h}_{out}$ are the vectors and $\overline{h}$ obtained for the person who leaves the room. Due to the great differences between the magnitudes of the computed Euclidean distances, there are normalized dividing $\Delta d_{P(n)}$ by 1000 and $\Delta \overline{h}_{(n)}$ by 10, to do a re-scaling, in order the values of the three variables $\Delta d_{P(n)}, \Delta \mathbf{I}_{P(n)}$ and $\Delta \overline{h}_{(n)}$ have magnitudes on similar scales.

For each person, the average of the distances $\Delta d_{P(n)}$ and $\Delta \mathbf{I}_{P(n)}$for each position P are then obtained, as shown in Eqs. 8 and 9:

$$\Delta \overline{d}_{(n)} = \frac{\Delta d_{F(n)} + \Delta d_{O(n)} + \Delta d_{B(n)}}{3} \tag{8}$$

$$\Delta \overline{I}_{(n)} = \frac{\Delta \mathbf{I}_{F(n)} + \Delta \mathbf{I}_{O(n)} + \Delta \mathbf{I}_{B(n)}}{3} \tag{9}$$

The person k is identified as the one who is leaving the room, if the quadratic sum $\delta_{(k)}$ of the mean values $\Delta \overline{d}_{(k)}, \Delta \overline{I}_{(k)}$ and $\Delta \overline{h}_{(k)}$ is the smallest among the N people $\delta_{(n)}$ with $n = 1, 2, \ldots, N$ present in the room, fulfilling Eq. 10.

$$\delta_{(k)} \leqslant \delta_{(n)}, \quad \forall n \neq k \tag{10}$$

where $\delta_{(k)}$ and $\delta_{(n)}$ are defined in Eq. 11 and 12 respectively:

$$\delta_{(k)} = \left(\Delta \overline{d}_{(k)}\right)^2 + \left(\Delta \overline{I}_{(k)}\right)^2 + \left(\Delta \overline{h}_{(k)}\right)^2 \tag{11}$$

$$\delta_{(n)} = \left(\Delta \overline{d}_{(n)}\right)^2 + \left(\Delta \overline{I}_{(n)}\right)^2 + \left(\Delta \overline{h}_{(n)}\right)^2 \tag{12}$$

It is worth note that we have squared the Euclidean distances to achieve better discrimination, since the obtained Euclidean distances are usually lower than the unit for $k = n$ and greater than the unit for $k \neq n$.

3. Results and discussion

This section presents the main experimental results obtained with the proposal. First, there is described the experimental setup. Next there are exposed the performance evaluation results, and the comparison with the state-of-the-art proposals. To end this section, the computational cost is analyzed.

Regarding the comparison to other works, it is worth highlighting that, to the best of authors' knowledge, there are no other works that carry out re-identification from depth and intensity images acquired form an overhead camera. The most similar is the proposal of (Paolanti et al., 2018) that used RGB-D from an overhead camera. Therefore, no other comparisons have been made with respect to other works, as the working conditions are very different, especially in terms of camera placement and the information used, being in most of these cases RGB.

3.1. Datasets

The experimental evaluation has been carried out using two different datasets: the first one is GODPR, that has been recorded and manually labeled by the authors, and the other one is the publicly available TVPR dataset (Liciotti, 2017; Liciotti et al., 2017). Both of them are briefly described below.

- **GODPR**: the GODPR (Geintra Overhead Depth People Re-identification) (Fuentes-Jimenez, Gutierrez, Guarasa, Luna, & Pizarro, 2020) is a dataset recorded with 2 high resolution RGB-D sensors, being the first a Kinect V2 time of flight depth sensor with a resolution of 512x424 pixels and framerate of approximate 30 fps. The second is an Intel Real-Sense D435 active stereo depth sensor, with a resolution of 1280x720. The used Both sensors provide depth measurements, but their working principles are different. In the first case the principle of the sensor is the time of flight, while in the second case the principle is the active stereo. In this dataset the camera is positioned in a top view configuration with different heights from 2760 mm to 3400 mm. The captured sequences were obtained in scenarios with different backgrounds and different natural and non-natural lighting conditions. We used for the experiments different users people to evaluate with a variety of features (clothes, face cover masks, hair, age, etc.). Each sequence contains a person walking in forward or backward direction, as it can be seen in the sample images shown in Figs. 2 and 3. This dataset contains a total of 136136 sequences, divided in three sets:
  - GODPR1: 42 sequences acquired with Kinect V2 at a height of 3360 mm and framerate of approximate 30 fps. This set contains 21 different people in each direction.
  - GODPR2- 48 sequences acquired with Intel Real sense D435 at a height of 3400 mm and low framerate of approximate 3 fps. This set contains 24 different people in each direction. The people wear face cover masks.
  - GODPR3- 46 sequences with Intel Real sense D435 at a height of 2760 mm and framerate of approximate 30 fps. This set contains 23 different people in each direction. The people wear face cover masks.

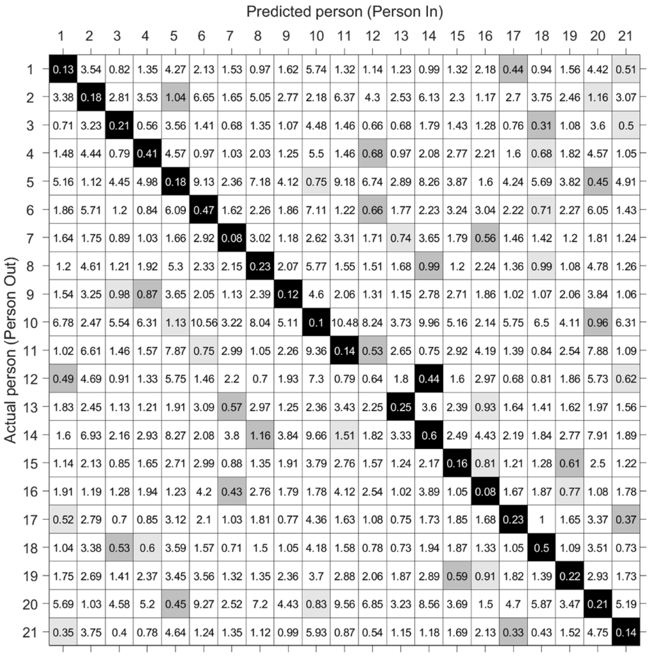
  For each person, there are recorded two sequences, walking along the scene in different directions. One of these sequences is used for identification (People In) whereas the other one is used for re-identification (Person Out). This dataset has been made available to the scientific community (Fuentes-Jimenez et al., 2020) including the used vectors with the number of pixels (φ_s) in each slice, the intensity histogram of each slice and the maximum person height (h), for the three frames used in each direction.
- **TVPR**: the TVPR (Top View Person Re-identification) (Liciotti, 2017; Liciotti et al., 2017) dataset is an RGB-D dataset that contains depth frames obtained from an Asus Xtion Pro Live RGB-D camera positioned in an overhead configuration. Depth measurements are made by means of a camera dedicated to IR detection and an infrared structured light source. All the images are captured with a resolution of 640x480 pixels and at an approximate framerate of 30 fps. TVPR recorded 100 different people, in 23 registrations sessions, that took place in 8 different days. The illumination conditions are not constant due to the natural light changes at different hours of the day and the weather changes. In the recorded sequences, RGB and depth images are not synchronized. For this reason, we firstly manually synchronize both sequences and then convert the RGB images to intensity (16 bits), however, there is always a small desynchronization between both images. We also determined that the camera is located at an approximate height of 3180 mm. As in GODPR, there are provided two sequences for each person in different directions, using one of them for identification, and the other one for re-identifications.

3.2. Performance evaluation

To evaluate the performance of the proposed classification model, for each dataset (including between 21 and 32 people), we assume that all the people have entered the room and then, we compare each of the person who leave the room with all those who have previously entered. Figs. 5–8 show the confusion matrices for each of the analyzed datasets. In these matrices, each row represents the value $\delta_{(n)}$ (defined in Eq. 12) obtained after comparing each person n with all the previously identified people. As it has been explained before, The person k is identified as the one who is leaving the room, if the quadratic sum $\delta_{(k)}$ of the average values $\Delta\overline{d}_{(k)}, \Delta\overline{I}_{(k)}$ and $\Delta\overline{h}_{(k)}$, is the smallest among the N identified people, so the smallest value in each row represents the re-identified person. For easier interpretation, the position of the three smallest values in each row are coloured: the smallest one in black color, whereas the other two in grey color. In each element of the matrix the value of $\delta_{(n)}$ is also shown.

Table 1 summarizes the percentage of correctly re-identified people for each dataset. As it can be seen, the results obtained are highly satisfactory, since for the four datasets the percentage of correct re-identifications exceeds 91%. The worst result is obtained for GODPR2. There are two causes that influence this result: the first one is that the sampling rate is so low (3 fps), thus the images are not capture at the same positions when each person enters and leaves the room, which can lead to changes in their appearance and the extracted features. The second cause is related to the height at which the camera is located (higher than in the other datasets), because, in general, the precision in the depth measurements worses as the object is further from the sensor, being this effect significantly noticeable in the Real-Sense D435 sensor. It is worth highlighting that, sometimes, although the person is correctly detected, there are other candidates whose value of $\delta_{(n)}$ is close to that of the correct person $\delta_{(cp)}$. To evaluate the robustness of the system, we have determined the number of wrong candidates whose difference between $\delta_{(n)}$ and $\delta_{(cp)}$ is less than a certain percentage X% of the value of $\delta_{(cp)}$ (Eq. 13). Table 1 also shows the percentage of wrong candidates with a difference less than 10% and less than 30%. Assuming a difference between $\delta_{(n)}$ and $\delta_{(cp)}$ less than 10%, the percentage of wrong candidates in the worst case (12.5%) is also for dataset GODPR2 and the best result is for dataset GODPR3, where there is no candidate with this difference.

$$d_{correct} = \frac{\delta_{(cp)} - \delta_{(n)}}{\delta_{(cp)}} \leqslant X\% \tag{13}$$



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0.13	3.54	0.82	1.35	4.27	2.13	1.53	0.97	1.62	5.74	1.32	1.14	1.23	0.99	1.32	2.18	0.44	0.94	1.56	4.42	0.51
2	3.38	0.18	2.81	3.53	1.04	6.65	1.65	5.05	2.77	2.18	6.37	4.3	2.53	6.13	2.3	1.17	2.7	3.75	2.46	1.16	3.07
3	0.71	3.23	0.21	0.56	3.56	1.41	0.68	1.35	1.07	4.48	1.46	0.66	0.68	1.79	1.43	1.28	0.76	0.31	1.08	3.6	0.5
4	1.48	4.44	0.79	0.41	4.57	0.97	1.03	2.03	1.25	5.5	1.46	0.68	0.97	2.08	2.77	2.21	1.6	0.68	1.82	4.57	1.05
5	5.16	1.12	4.45	4.98	0.18	9.13	2.36	7.18	4.12	0.75	9.18	6.74	2.89	8.26	3.87	1.6	4.24	5.69	3.82	0.45	4.91
6	1.86	5.71	1.2	0.84	6.09	0.47	1.62	2.26	1.86	7.11	1.22	0.66	1.77	2.23	3.24	3.04	2.22	0.71	2.27	6.05	1.43
7	1.64	1.75	0.89	1.03	1.66	2.92	0.08	3.02	1.18	2.62	3.31	1.71	0.74	3.65	1.79	0.56	1.46	1.42	1.2	1.81	1.24
8	1.2	4.61	1.21	1.92	5.3	2.33	2.15	0.23	2.07	5.77	1.55	1.51	1.68	0.99	1.2	2.24	1.36	0.99	1.08	4.78	1.26
9	1.54	3.25	0.98	0.87	3.65	2.05	1.13	2.39	0.12	4.6	2.06	1.31	1.15	2.78	2.71	1.86	1.02	1.07	2.06	3.84	1.06
10	6.78	2.47	5.54	6.31	1.13	10.56	3.22	8.04	5.11	0.1	10.48	8.24	3.73	9.96	5.16	2.14	5.75	6.5	4.11	0.96	6.31
11	1.02	6.61	1.46	1.57	7.87	0.75	2.99	1.05	2.26	9.36	0.14	0.53	2.65	0.75	2.92	4.19	1.39	0.84	2.54	7.88	1.09
12	0.49	4.69	0.91	1.33	5.75	1.46	2.2	0.7	1.93	7.3	0.79	0.64	1.8	0.44	1.6	2.97	0.68	0.81	1.86	5.73	0.62
13	1.83	2.45	1.13	1.21	1.91	3.09	0.57	2.97	1.25	2.36	3.43	2.25	0.25	3.6	2.39	0.93	1.64	1.41	1.62	1.97	1.56
14	1.6	6.93	2.16	2.93	8.27	2.08	3.8	1.16	3.84	9.66	1.51	1.82	3.33	0.6	2.49	4.43	2.19	1.84	2.77	7.91	1.89
15	1.14	2.13	0.85	1.65	2.71	2.99	0.88	1.35	1.91	3.79	2.76	1.57	1.24	2.17	0.16	0.81	1.21	1.28	0.61	2.5	1.22
16	1.91	1.19	1.28	1.94	1.23	4.2	0.43	2.76	1.79	1.78	4.12	2.54	1.02	3.89	1.05	0.08	1.67	1.87	0.77	1.08	1.78
17	0.52	2.79	0.7	0.85	3.12	2.1	1.03	1.81	0.77	4.36	1.63	1.08	0.75	1.73	1.85	1.68	0.23	1	1.65	3.37	0.37
18	1.04	3.38	0.53	0.6	3.59	1.57	0.71	1.5	1.05	4.18	1.58	0.78	0.73	1.94	1.87	1.33	1.05	0.5	1.09	3.51	0.73
19	1.75	2.69	1.41	2.37	3.45	3.56	1.32	1.35	2.36	3.7	2.88	2.06	1.87	2.89	0.59	0.91	1.82	1.39	0.22	2.93	1.73
20	5.69	1.03	4.58	5.2	0.45	9.27	2.52	7.2	4.43	0.83	9.56	6.85	3.23	8.56	3.69	1.5	4.7	5.87	3.47	0.21	5.19
21	0.35	3.75	0.4	0.78	4.64	1.24	1.35	1.12	0.99	5.93	0.87	0.54	1.15	1.18	1.69	2.13	0.33	0.43	1.52	4.75	0.14

Fig. 5. Confusion matrix obtained for GODPR1 dataset.

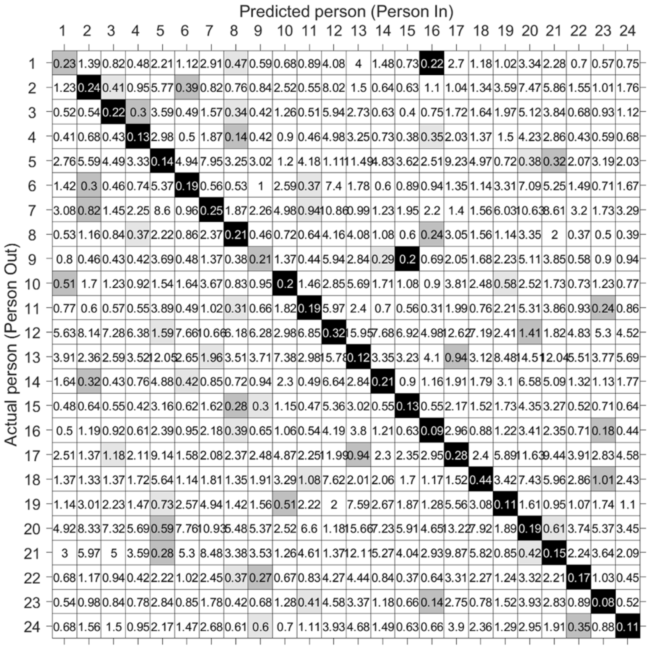


Fig. 6. Confusion matrix obtained for GODPR2 dataset.

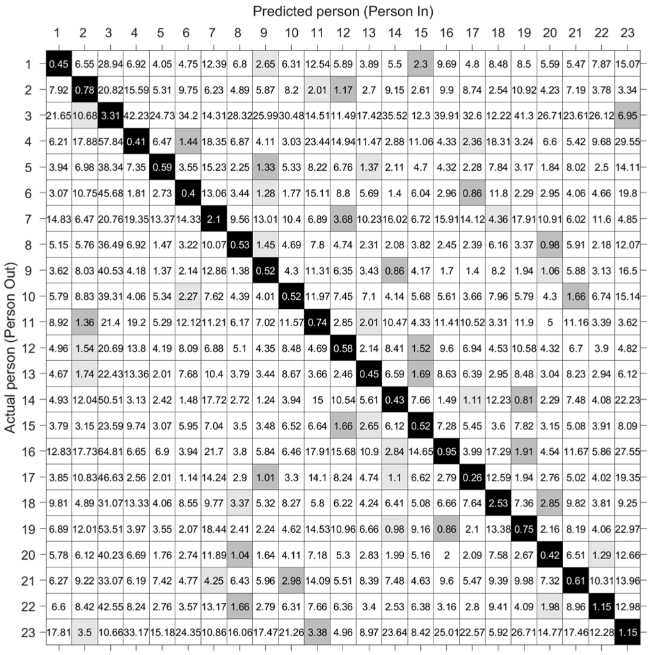


Fig. 7. Confusion matrix obtained for GODPR3 dataset.

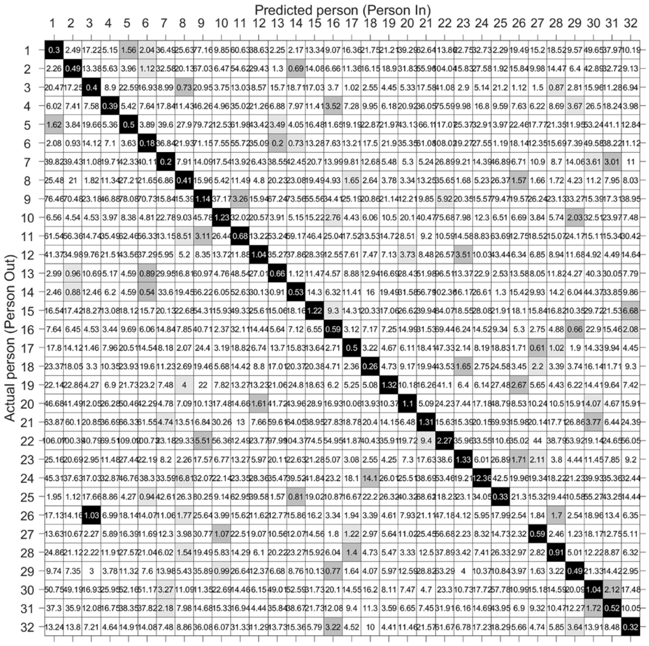


Fig. 8. Confusion matrix obtained for TVPR dataset.

Table 1
Percentage of correctly re-identified people and percentage of wrong candidates with a difference between $\delta_{(n)}$ and $\delta_{(k)}$ lower than 10% and 30%.

Dataset	Correct re-ID (%)	People with $d_{correct} \leq$ 10%		People with $d_{correct} \leq$ 30%	
		Quantity	%	Quantity	%
GODPR1	95.24	2	9.52	2	9.52
GODPR2	91.67	3	12.50	4	16.67
GODPR3	100.0	0	0.00	2	8.70
TVPR	96.88	2	6.25	6	18.75

Table 2
Precision, recall and F1-score obtained for the different tested datasets.

Method	Dataset	Precision	Recall	F1-score
Our proposal	GODPR1	0.952	0.952	0.952
	GODPR2	0.917	0.917	0.917
	GODPR3	1.0	1.0	1.0
	TVPR	0.969	0.969	0.969
(Paolanti et al., 2018)		0.86	0.85	0.83

The precision, recall and F1-score are shown in Table 2. To obtain the results in this table, if the re-identified person is the correct one, it is considered a True Positive (TP), whereas if the re-identification is not correct, it is considered as a False Negative (FN) for the actual person, and a False Positive (FP) for the incorrect one. Since all the evaluated people are re-identified as one of the previously identified people, the number of FP and FN are equal. Due to that, the precision and recall have the same value for each dataset.

To ease comparison with the state-of-the-art, Table 2 also shows the results obtained in (Paolanti et al., 2018) for the TVPR dataset, but it is worth highlighting that, although this proposal uses an overhead camera, their method is based on RGB-D data, instead of intensity and depth data. As it can be observed, our method obtains a precision, recall and F1-score over 91,7% for all the tested datasets. Besides, the precision for the TVPR datasets is over 96%, what outperforms the results in (Paolanti et al., 2018) with a precision of 86%.

3.3. Computational cost

To analyze the computational efficiency of the proposed algorithm, we have determined the average time to find and store the used features: d, $\mathbf{I}$ and $\overline{h}$ and the average time to perform the people re-identification, showing the obtained values in Table 3. These values are computed without any optimization and taking into account the time used to store and read the vectors into the SSD. The time values has been obtained in computer with a processor Intel (R) Core(TM) i5-7500 CPU, 3.40 GHz and 32 GB of RAM. The time needed to find features depends directly on the resolution of the sensor. Although it presents relatively high values for both sensors, it allows the system to operate in real time, since it is

Table 3
Average processing time of the different process of the re-identification proposal.

Process	Resolution 1280 × 720	Resolution 640 × 480	Units
Find and store features: d, I and $\overline{h}$	19.9	10.1	ms/frame
People re-identification	0.1	0.1	ms/person

much less than the time it would take people to enter or leave the room, crossing along the area of interest. In the case of the time to carry out the people re-identification, it is important that it should be as short as possible, since it is multiplied by the number of people models in the *Person in* dataset (and this value can be high). With the obtained value (0.1 ms), the person who leaves can be compared with up to 10 000 people models in one second, what allows real-time people re-identification.

4. Conclusions

In this work, it has been presented a method to identify and re-identify people entering and leaving a room, from the depth and IR intensity information provided by a depth camera located in an overhead position. The proposal is based on ad-hod feature vectors, that include information about people head and shoulders anthropometric and texture characteristics, and a classificator based on the Euclidean distance, and it is able to work in real time without specific hardware. The proposal has been evaluated in two different datasets, including information acquired with sensors that use different operating principles to perform depth and intensity measurements, with re-identification results over the 90% in all cases, outperforming other state-of-the-art approaches for top-view people re-identification. From the obtained results, the following conclusions can be reached:

- The height at which the camera is located has a great influence on the re-identification process, with the Real Sense sensor at the height of 3400 mm (GODPR2 dataset) and 2760 mm (GODPR3 dataset), re-identification values of 91.67% and 100% respectively are obtained. It is because the height of the camera can modify the depth and intensity features, as the measurement errors increases significantly with the distance.
- The method works correctly even when there is a small desynchronization between the depth and intensity frame, as it can be seen in the results obtained with the TVPR dataset (with a 96.88% of precision).
- The method correctly re-identifies people even if they wear masks on their faces (GODPR2 and GODPR3 datasets), since it is used information related to anthropocentric characteristics.
- The method can be implemented in a low-performance PC and it is able to work in real time.

Regarding future works, we have observed that depending on the type of sensor and the height at which it is located, the different features change their weight in the re-identification process. Thus, we propose to carry out a study of the weight of each feature to improve the results.

CRediT authorship contribution statement

Carlos A. Luna: Conceptualization, Methodology, Software, Writing - review & editing. **Cristina Losada-Gutiérrez:** Conceptualization, Methodology, Data curation, Writing - review & editing. **David Fuentes-Jimenez:** Methodology, Data curation, Software, Writing - review & editing. **Manuel Mazo:** Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by the Spanish Ministry of Economy and Competitiveness under projects HEIMDAL-UAH (TIN2016-75982-C2-1-R) and ARTEMISA (TIN2016-80939-R), and by the University of Alcalá under projects ACUFANO (CCG19/IA-024) and ARGOS (CCG20/IA-043).

References

Ahmed, E., Jones, M., & Marks, T.K. (2015). An improved deep learning architecture for person re-identification. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3908–3916). IEEE.https://doi.org/10.1109/CVPR.2015.7299016.

Baltieri, D., Vezzani, R., & Cucchiara, R. (2015). Mapping Appearance Descriptors on 3D Body Models for People Re-identification. *International Journal of Computer Vision, 111*, 345–364. https://doi.org/10.1007/s11263-014-0747-z

Barbosa, I.B., Cristani, M., Del Bue, A., Bazzani, L., & Murino, V. (2012). Re-identification with RGB-D Sensors. In Computer Vision – ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7–13, 2012, Proceedings, Part I (pp. 433–442).https://doi.org/10.1007/978-3-642-33863-2_43.

Bedagkar-Gala, A., & Shah, S. K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing, 32*, 270–286. https://doi.org/10.1016/j.imavis.2014.02.001

Chen, Y., Zhu, X., & Gong, S. (2017). Person Re-identification by Deep Learning Multi-scale Representations. In 2017 IEEE International Conference on Computer Vision Workshops (ICCVW) (pp. 2590–2600). IEEE.https://doi.org/10.1109/ICCVW.2017.304.

Chen, Y., Zhu, X., & Gong, S. (2018). Deep association learning for unsupervised video person re-identification. http://arxiv.org/abs/1808.07301. arXiv:1808.07301.

Chung, D., Tahboub, K., & Delp, E.J. (2017). A Two Stream Siamese Convolutional Neural Network for Person Re-identification. In 2017 IEEE International Conference on Computer Vision (ICCV) (pp. 1992–2000). IEEE.https://doi.org/10.1109/ICCV.2017.218.

D'Angelo, A., & Dugelay, J.-L. (2011). People re-identification in camera networks based on probabilistic color histograms. In A. Said, O.G. Guleryuz, & R.L. Stevenson (Eds.), Visual Information Processing and Communication II (p. 78820K). SPIE volume 7882.https://doi.org/10.1117/12.876453.

de Carvalho Prates, R.F., & Schwartz, W.R. (2015). CBRA: Color-based ranking aggregation for person re-identification. In 2015 IEEE International Conference on Image Processing (ICIP) (pp. 1975–1979). IEEE volume 2015-Decem.https://doi.org/10.1109/ICIP.2015.7351146.

Devyatkov, V. V., Alfimtsev, A. N., & Taranyan, A. R. (2018). Multicamera Human Re-Identification based on Covariance Descriptor. *Pattern Recognition and Image Analysis, 28*, 232–242. https://doi.org/10.1134/S1054661818020025

Fuentes-Jimenez, D., Gutierrez, C.L., Guarasa, J.M., Luna, C., & Pizarro, D. (2020). Depth person detection database (gfpd-uah).https://www.kaggle.com/dsv/1664233.https://doi.org/10.34740/KAGGLE/DSV/1664233.

Gharghabi, S., Shamshirdar, F., Shangari, T.A., & Maroofkhani, F. (2015). People re-identification using 3D descriptor with skeleton information. In 2015 International Conference on Informatics, Electronics & Vision (ICIEV) (pp. 1–5). IEEE.https://doi.org/10.1109/ICIEV.2015.7333986.

Islam, K. (2020). Person search: New paradigm of person re-identification: A survey and outlook of recent works. *Image and Vision Computing, 101*, Article 103970. https://doi.org/10.1016/j.imavis.2020.103970

Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P. M., & Bischof, H. (2012). Large scale metric learning from equivalence constraints. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2288–2295). https://doi.org/10.1109/CVPR.2012.6247939

Li, M., Zhu, X., & Gong, S. (2018). Unsupervised Person Re-identification by Deep Learning Tracklet Association. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (pp. 772–788). volume 11208 LNCS.https://doi.org/10.1007/978-3-030-01225-0_45. arXiv:1809.02874.

Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). DeepReID: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition (pp. 152–159). https://doi.org/10.1109/CVPR.2014.27

Liao, S., Hu, Y., Xiangyu Zhu, & Li, S.Z. (2015). Person re-identification by Local Maximal Occurrence representation and metric learning. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2197–2206). IEEE.https://doi.org/10.1109/CVPR.2015.7298832.

Liciotti, D., Frontoni, E., Mancini, A., & Zingaretti, P. (2017a). Pervasive system for consumer behaviour analysis in retail environments. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (pp. 12–23). Springer Verlag volume 10165 LNCS.https://doi.org/10.1007/978-3-319-56687-0_2.

Liciotti, D., Paolanti, M., Frontoni, E., Mancini, A., & Zingaretti, P. (2017b). Person Re-identification Dataset with RGB-D Camera in a Top-View Configuration. In Video Analytics. Face and Facial Expression Recognition and Audience Measurement: Third International Workshop, VAAM 2016, and Second International Workshop, FFER 2016, 2016, Revised Selected Papers (pp. 1–11). Springer, Cham.https://doi.org/10.1007/978-3-319-56687-0_1.

Luna, C. A., Losada-Gutierrez, C., Fuentes-Jimenez, D., Fernandez-Rincon, A., Mazo, M., & Macias-Guarasa, J. (2017). Robust people detection using depth information from an overhead Time-of-Flight camera. *Expert Systems with Applications, 71*, 240–256. https://doi.org/10.1016/j.eswa.2016.11.019

Luna, C. A., Losada-Gutiérrez, C., Fuentes-Jiménez, D., & Mazo, M. (2021). Fast heuristic method to detect people in frontal depth images. *Expert Systems with Applications, 168*, Article 114483. https://doi.org/10.1016/j.eswa.2020.114483

Matsukawa, T., Okabe, T., Suzuki, E., & Sato, Y. (2016). Hierarchical Gaussian Descriptor for Person Re-identification. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1363–1372). IEEE.https://doi.org/10.1109/CVPR.2016.152.

Matzner, S., Heredia-Langner, A., Amidan, B., Boettcher, E. J., Lochtefeld, D., & Webb, T. (2015). Standoff human identification using body shape. In *2015 IEEE International Symposium on Technologies for Homeland Security (HST)* (pp. 1–6). https://doi.org/10.1109/THS.2015.7225300

Merad, D., Aziz, K. E., Iguernaissi, R., Fertil, B., & Drap, P. (2016). Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis. *Pattern Recognition Letters, 81*, 11–20. https://doi.org/10.1016/j.patrec.2016.04.011

Mingyong Zeng, Wu, Z., Tian, C., Lei Zhang, & Lei Hu (2015). Efficient person re-identification by hybrid spatiogram and covariance descriptor. In 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 48–56). IEEE.https://doi.org/10.1109/CVPRW.2015.7301296.

Pala, F., Satta, R., Fumera, G., & Roli, F. (2016). Multimodal Person Reidentification Using RGB-D Cameras. *IEEE Transactions on Circuits and Systems for Video Technology, 26*, 788–799. https://doi.org/10.1109/TCSVT.2015.2424056

Paolanti, M., Romeo, L., Liciotti, D., Pietrin, R., Cenci, A., Frontoni, E., & Zingaretti, P. (2018). Person re-identification with RGB-D camera in top-view configuration through multiple nearest neighbor classifiers and neighborhood component features selection. *Sensors (Switzerland), 18*, 1–18. https://doi.org/10.3390/s18103471

Patruno, C., Marani, R., Cicirelli, G., Stella, E., & D'Orazio, T. (2019). People re-identification using skeleton standard posture and color descriptors from RGB-D data. *Pattern Recognition, 89*, 77–90. https://doi.org/10.1016/j.patcog.2019.01.003

Qian, X., Fu, Y., Jiang, Y.-G., Xiang, T., & Xue, X. (2017). Multi-scale Deep Learning Architectures for Person Re-identification. In 2017 IEEE International Conference on Computer Vision (ICCV) (pp. 5409–5418). IEEE.https://doi.org/10.1109/ICCV.2017.577.

Satta, R. (2013). Appearance Descriptors for Person Re-identification: a Comprehensive Review, http://arxiv.org/abs/1307.5748. arXiv:1307.5748.

Satta, R., Fumera, G., & Roli, F. (2012). Fast person re-identification based on dissimilarity representations. *Pattern Recognition Letters, 33*, 1838–1848. https://doi.org/10.1016/j.patrec.2012.03.026

Schumann, A., & Stiefelhagen, R. (2017). Person Re-identification by Deep Learning Attribute-Complementary Information. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 1435–1443). IEEE. https://doi.org/10.1109/CVPRW.2017.186.

Sell, J., & O'Connor, P. (2014). The Xbox One System on a Chip and Kinect Sensor. *IEEE Micro, 34*, 44–53. https://doi.org/10.1109/MM.2014.9

Smisek, J., Jancosek, M., & Pajdla, T. (2011). 3D with Kinect. In Computer Vision Workshops (ICCV Workshops). In *2011 IEEE International Conference on* (pp. 1154–1160). https://doi.org/10.1109/ICCVW.2011.6130380

Vezzani, R., Baltieri, D., & Cucchiara, R. (2013). *People reidentification in surveillance and forensics: A survey*. https://doi.org/10.1145/2543581.2543596

Wang, G., Gong, S., Cheng, J., & Hou, Z. (2020). Faster Person Re-identification. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (pp. 275–292). Springer Science and Business Media Deutschland GmbH volume 12353 LNCS.https://doi.org/10.1007/978-3-030-58598-3_17. arXiv:2008.06826.

Wang, J., Zhu, X., Gong, S., & Li, W. (2018). Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-identification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 2275–2284). IEEE.https://doi.org/10.1109/CVPR.2018.00242. arXiv:1803.09786.

Wu, D., Zheng, S. J., Zhang, X. P., Yuan, C. A., Cheng, F., Zhao, Y., Lin, Y. J., Zhao, Z. Q., Jiang, Y. L., & Huang, D. S. (2019). Deep learning-based methods for person re-identification: A comprehensive review. *Neurocomputing, 337*, 354–371. https://doi.org/10.1016/j.neucom.2019.01.079

Xin, X., Wang, J., Xie, R., Zhou, S., Huang, W., & Zheng, N. (2019). Semi-supervised person Re-Identification using multi-view clustering. *Pattern Recognition, 88*, 285–297. https://doi.org/10.1016/j.patcog.2018.11.025

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S.C.H. (2020). Deep Learning for Person Re-identification: A Survey and Outlook, http://arxiv.org/abs/2001.04193. arXiv:2001.04193.

Zheng, L., Yang, Y., & Hauptmann, A.G. (2016). Person Re-identification: Past, Present and Future, http://arxiv.org/abs/1610.02984. arXiv:1610.02984.