People Re-Identification using Skeleton Standard Posture and Color Descriptors from RGB-D Data

Cosimo Patruno, Roberto Marani, Grazia Cicirelli, Ettore Stella, Tiziana D'Orazio*

Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, CNR
Via Amendola 122 D/O, Bari, Italy
** Corresponding author: cosimo.patruno@stiima.cnr.it*

Abstract

This paper tackles the problem of people re-identification by using soft biometrics features. The method works on RGB-D data (color point clouds) to determine the best matching among a database of possible users. For each subject under testing, skeletal information in three-dimensions is used to regularize the pose and to create a skeleton standard posture (SSP). A partition grid, whose sizes depend on the SSP, groups the samples of the point cloud accordingly to their position. Every group is then studied to build the person signature. The same grid is then used for the other subjects of the database to preserve information about possible shape differences among users. The effectiveness of this novel method has been tested on three public datasets. Numerical experiments demonstrate an improvement of results with reference to the current state-of-the-art, with recognition rates of 97.84% (on a partition of BIWI RGBD-ID), 61.97% (KinectREID) and 89.71% (RGBD-ID), respectively.

Keywords: People re-identification; Color-based descriptor; Skeleton standard posture; Partition grid; RGB-D sensor; Color point cloud.

1. Introduction

In the last decade, people re-identification has gained an increasing interest by the scientific community. The recent terrorist attacks and the continuous demanding of monitoring public and

crowded places have been favouring the developing of new methodologies and sensors to accomplish this task. Re-identification is the ability of an intelligent system to recognize a person across disjoint camera views. In this regard, several practical applications, including video surveillance, service robots, human-robot interaction, access control, people tracking, analysis of behaviors and long-term activities of people in care centers, and sport analytics, can be identified [1-8].

Generally, in order to re-identify a person, several features can be considered. According to the cues or characteristics selected it is possible to discern two categories: soft biometric features and hard biometric ones. The soft biometrics [9] involves all the traits of a person that can be physical (e.g. face features, eye color, skin color, hair color, height, weight, distinctive marks, etc.), behavioral (e.g. gait, keystroke, voice, handwriting, etc.) or adhered human characteristics (e.g. cloth color, accessories, tattoos). On the contrary, the hard biometrics [10] includes fingerprints, DNA sequence, retina features, ear features and so on. Therefore the hard biometrics can require invasive techniques to gather the information or at least the collaboration of people.

Currently, increasing attention in the field of people re-identification has been focused on the use of novel sensors, namely RGB-D cameras [11-14], able to produce both color and spatial information of the environment. Specifically, RGB-D sensors can sample the environment, and thus people within the scene, producing dense point clouds. These are arrangements of points which model the scene in three-dimensions, also giving information about its color appearance. In addition, when the point cloud models a human being, proper available algorithms, based on machine learning [15-17], can provide a set of joints with 3D coordinates which create a simplified skeleton of the user.

In this paper, we propose a new method belonging to the soft biometrical approaches. It uses the combination of color and depth information to build a very informative signature of the person under investigation in order to increase the re-identification performance. The underlying idea follows the human ability of re-identifying people by taking into account both anthropometric and color information.

The detailed description of the proposed approach is given in Section III after a review of the related works presented in Section II. Then experimental results are reported in Section IV, where the performance of the methodology is analyzed in terms of predictive accuracies; conclusions and remarks for future investigations are given in Section V.

2. Related works

In general, people re-identification is performed by considering two distinct phases: first a signature, composed of a set of significant features, is learned for the subject under testing. Then, in the second phase, the learned signature is compared with those of a database of known subjects to sort the elements according to the similarity.

In literature, several methods have been proposed to solve this problem [18-20]. Most of them rely on the person appearance that can be described using features, such as the texture or the color of clothes. Other features as the body shape, the skeletal information or the face appearance can be employed to enhance the recognition as well.

In [21] the authors propose simple geometric features for person re-identification, which are extracted by using a top-view setup of a consumer RGB-D sensor. Only depth images are considered for the computations. The main depth image blobs, such as head and torso, are analyzed in order to extract their height, area, occupancy volume and speed useful for the trajectory descriptor. A similar approach is proposed in [22] where a new dataset, made of top-view sequences, is presented. Anthropometric and color-based features are computed by investigating the top side blobs of 100 people. Combining color and depth features a recognition rate of about 70% is obtained.

Other approaches use standard camera configurations where the individuals are framed in frontal or rear-wise side such as in [23-25]. The authors of [23] propose a re-identification algorithm able to identify particular signatures from the range data of people. The signature is composed of ten soft biometric cues. Specifically, seven skeleton-based features are derived by the skeleton information and the other three surface-based features are defined using the geodesic distances computed among

a predefined set of joints. A study on how these different features can be weighted in order to maximize the re-identification performance is presented. Then the signature matching phase is applied on several datasets acquired across intervals of days and investigating collaborative and non-collaborative settings. The average recognition rate is below the 20%. In [24], a people re-identification approach that combines 3D descriptors of body shape and skeleton data is presented. The Viewpoint Feature Histogram (VFH) is used to extract the first 3D descriptor that focuses on the spatial arrangement of the samples of the point cloud, thus getting information about the body shape of the person. Skeleton link lengths are used for composing the second descriptor. The computed features are then combined in order to obtain the final person signature. A voting system is finally proposed for finding the best matches among signatures. Recognition scores of 25% and 14.2% are obtained for datasets of people performing two actions (still and walking, respectively).

3D data without color information is also employed in [25]. The authors exploit the soft biometric cues and in particular the body skeleton information, providing two different approaches for person re-identification. One method builds a subject descriptor using the body skeleton information, whereas the other builds a standard point cloud pose which is compared with the ones of the gallery set in terms of a fitness score resulting from the application of the Iterative Closest Point (ICP) algorithm. A dataset of 50 people, performing different actions, is presented and used to evaluate the proposed algorithm for long-term re-identification. The first method achieves an average recognition rate of 23.9%, whereas the second one slightly increases it to 27.5%.

The exclusive use of color information or 3D data alone is not always sufficient to solve people re-identification [26-28]. Therefore, many literature works propose the integrated use of both the color and the 3D data for improving re-identification accuracy. In [29] the authors fuse clothing appearance descriptors with anthropometric measures extracted from depth data. Then a dissimilarity-based framework for building and fusing multi-modal descriptors of pedestrian images is also proposed. The simultaneous use of anthropometric measures and clothing appearance descriptors increases the first-rank recognition rate of about 20% in comparison with the only use of

anthropometric information. A 3D cylindrical descriptor grid that stores color features with angles and heights is proposed in [30]. A normalization process aimed at removing brightness and contrast variations is applied before matching. A new dataset involving three scenarios is presented and used for performance evaluation. Recognition scores over the 70% are found in all the investigated datasets.

Another re-identification framework [31] creates a non-articulated 3D body model, whose vertices are filled by appearance features (color and gradient histograms). The body model has a fixed shape, whereas its size is defined by a scale factor which depends on the person shape. By using the multi-shot approaches for creating the training and the query models, an average rank-1 recognition rate of about 70% is achieved on the dataset 3DPeS [32], which contains 200 people.

Finally, the recent spread of deep learning has favored the development of some person re-identification applications based on this approach [33,34]. In this regard, a multi-modal uniform deep learning method for extracting the color and anthropometric features is proposed in [35]. The method uses two Convolutional Neural Networks (CNNs), properly trained to separately analyze the depth and RGB images. Afterwards, a multi-modal fusion layer combines the computed features from the input data producing the latent variable related to a person under analysis. This method requires a preliminary training phase that imposes a preliminary labelling phase of the set of people to be re-identified and the generation of features. Furthermore, the learned models depend on the specific setup of acquisition, the lighting conditions and the camera intrinsic (resolution, exposition, focal length, etc.) and extrinsic (pose in space) parameters.

In this paper we propose a new method which uses both color features and 3D data acquired by RGB-D sensors, but it does not need a training phase to learn features and does not depend on specific camera setups. The proposed algorithm can be used for accomplishing short-term re-identification of people in large indoor environments where multiple RGBD cameras can be employed for people tracking purposes or for monitoring restricted areas. We assume that people do not change their clothes and move in the environment with typical walking gaits. In these

conditions, re-identification approaches based on color features can take advantage of depth information to compare more robust data that consider both posture and body size of different people.

The underlying idea behind our work is that 3D information is used to properly weight the color information. Color alone can be effective and enough for people re-identification when images are acquired in the same conditions, with people standing always in front of camera with the same posture. But when these constraints cannot be imposed the knowledge of posture can guide the comparison of color features and improve the recognition. In other words, 3D information is managed in order to drive the comparison, based on color information content, of consistent body parts. This is achieved by applying rigid roto-translations to the 3D point cloud of the subject under testing in order to opportunely align it in a global coordinate system. A Skeleton Standard Posture is then computed to create a new representation of the person skeleton. This representation is then used to produce an unevenly-spaced partition grid, which is properly set to non-uniformly resample the point cloud in accordance with the anthropometric properties of the subject under analysis. Each cell of the partition grid is assessed in order to compute the color statistics which defines the signature of the person to be recognized. At the same time, the point clouds of the reference database are similarly treated to extract the signatures, but using the same partition grid of the subject under analysis. In this way, the signatures have short distances only when people have similar 3D appearance.

The main contributions of paper can be mainly highlighted in the following points: **a)** introduction of Skeleton Standard Posture (SSP), **b)** computation of the Partition Grid (PG) through the SSP for generating independent and robust color-based descriptors and **c)** re-projection of the user signatures in the DataBase (DB) by means of the SSP related to the person under investigation. Different body shapes lead to very different person signatures on the basis of the same SSP as input, thus enhancing the re-identification capabilities of the proposed method.

This approach has been tested on different datasets such as the BIWI RGBD-ID [36], the KinectREID [37] and the RGBD-ID [38], in order to prove its accuracy and robustness. Experimental evidence is provided to support its outperformance by comparing it with those state-of-the-art methods specifically developed to operate in typical video surveillance contexts. In particular, the proposed approach will be compared with the work presented in [29], which uses SDALF [39], eBiCov [40] and MCMimpl [41] methods.

In the following sections the proposed algorithm will be described with more details highlighting its peculiarities and its distinctive advantages.

3. Methodology

With reference to the flow chart in Fig. 1, the method accepts as input the dense point cloud of a person and his/her 3D skeleton data. After a preliminary processing step aimed at cleaning the input data from noise, the point cloud is aligned according to the camera viewpoint by exploiting the 3D skeleton joints (see block Point Cloud Alignment in Fig. 1). The obtained new data is processed by the Skeleton Standard Posture & Partition Grid module, which computes the Skeleton Standard Posture of the person. This module outputs the person signature which consists of color descriptors associated to a partition grid related to the source user under analysis. The same partition grid is then used to extract the descriptors of other users (in DataBase re-projection module). The reference DataBase (DB) is accordingly modified as follows: 1) all the signatures in the DB are projected by using the partition grid of the source person instance; 2) the obtained signatures corresponding to the same person are averaged to obtain the reference signature which has to be used in the following matching phase. Finally, the Matching module compares the source person signature (green containers in Fig.1) against to all the re-projected reference ones (cyan containers in Fig. 1). It returns the ID number of the person who is the most similar to the considered one.

In this work we assume that the method operates for achieving short-term person re-identification into indoor scenarios. Therefore, it is supposed that people do not change suddenly their clothes

and/or accessories. Furthermore, the input point cloud of the person under analysis must be completed by a skeletal representation. Each joint of the skeleton has to be known in 3D space. The following subsections will detail the entire processing pipeline which includes data pre-processing, construction of SSP and partition grid, and signature computation.

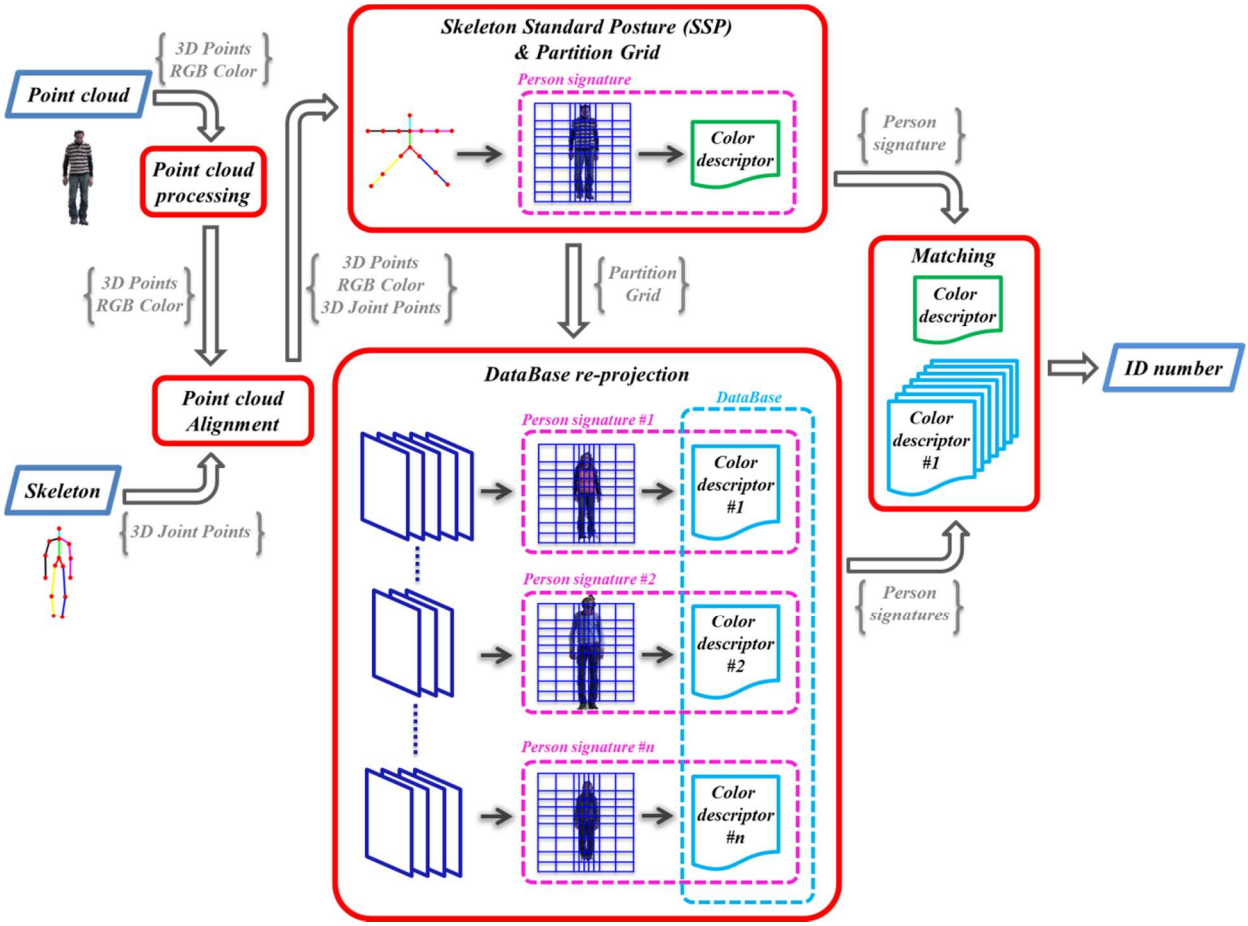


Fig. 1. Flow chart of the processing pipeline: input point clouds and skeletons are preprocessed and aligned; SSP are then computed to determine a color descriptor; feature comparison assigns the ID with respect to the re-projected DB.

3.1 Point cloud pre-processing

Point clouds, acquired by using range cameras, often suffer from noise and outliers. In particular, sparse outliers, known as shadow points, can occur at the boundaries of the acquired subjects during point cloud segmentation. Moreover, also secondary reflections or high-absorbing targets can increase the noise of point clouds. Therefore a pre-processing stage is needed in order to clean up outliers and/or irrelevant points.

Furthermore, each person of the dataset can have different pose depending on his/her relative position and orientation with respect to the camera. Different poses in space can generate not comparable point clouds because of self-occlusions. As a consequence, point clouds need to be aligned in a global coordinate framework in order to be coherently compared.

As a first step in point cloud pre-processing, a statistical filter is applied in order to remove outliers as proposed in [42]. The Fig. 2 shows the resulting point cloud obtained after applying this filter on a noisy point cloud of the person. In this case, the outliers within the red ellipse are correctly removed, thus producing the set of points shown in Fig. 2(b).

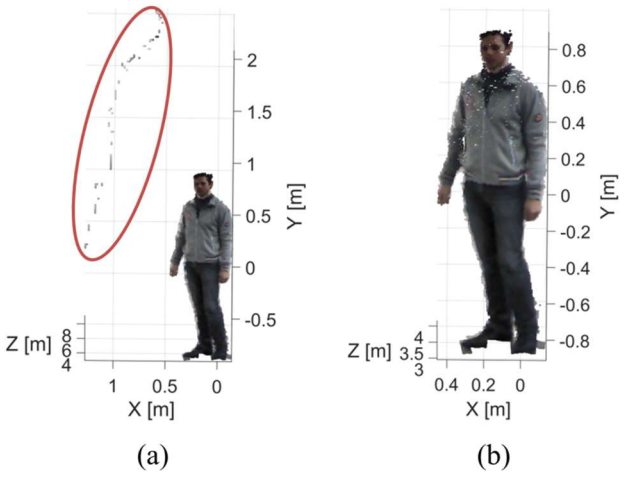


Fig. 2: (a) Input point cloud with some outliers enclosed by the red ellipse and (b) filtered point cloud.

The second step in the pre-processing phase handles the point clouds in order to align them to a common reference system. Specifically, point clouds are managed to have the front or the back side of the person always facing the camera. In this way, also the point clouds of people that are framed side-wise by the RGB-D cameras will have comparable poses.

The alignment phase is performed by taking advantage of the body skeleton information provided along with the input point clouds expressed in the global reference system (X, Y, Z), having origin in the optical center of the camera and the Z-axis along its optical axis. The pose of each person is determined by computing the plane that best fits the skeleton joints in the weighted least squares sense. As the body pose of person under analysis could affect negatively the alignment stage, the

estimation of the fitting plane is performed in a weighted manner, where each skeleton joint has a proper weight value. Specifically, the joints as the torso, the hips, the shoulders, the neck and the head, which are less affected by the person movements, involve larger influence during the plane estimation and they have consequently high weight values. Conversely, the joints of the arms and the legs have much lower weight values because they are related to body parts that move much more frequently.

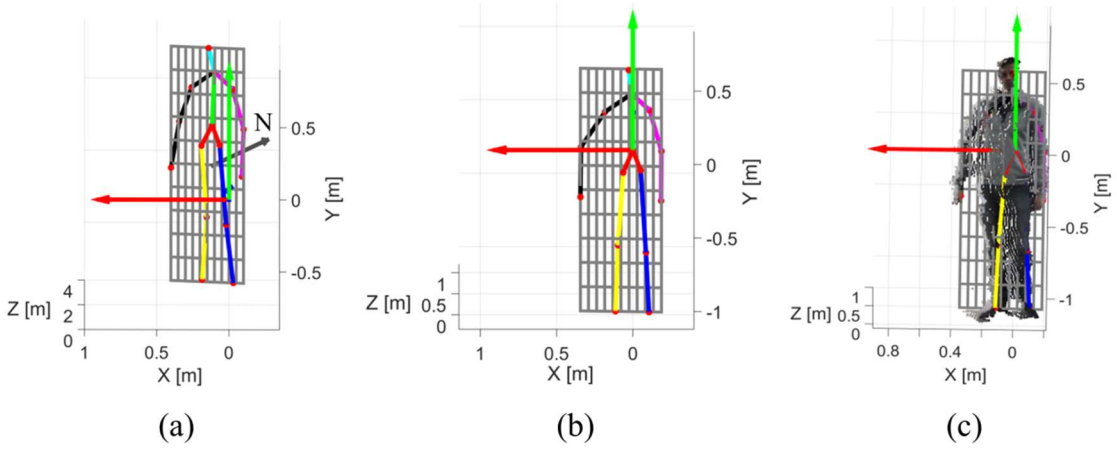


Fig. 3: (a) Planar fitting of the skeleton joints (red dots) where N is the normal unit vector to the fitting plane. (b) Alignment of joint points and (c) alignment of the person point cloud.

With reference to Fig 3(a), the unit vector $\mathbf{N}=\begin{bmatrix} N_x & N_y & N_z \end{bmatrix}^T$ represents the normal of this approximating plane, which has to be aligned with the Z-axis to obtain the facing view of the person. The point cloud is first translated in order to center the torso joint, assumed to be the center of mass of the skeleton, in the origin of the reference system (X, Y, Z). Then, by using the notation reported in [43], the point cloud is rotated by using the matrix in Eq. (1):

$$\mathbf{R}=\begin{pmatrix} N_z+hv_x^2 & hv_xv_y-v_z & hv_xv_z+v_y \\ hv_xv_y+v_z & N_z+hv_y^2 & hv_yv_z-v_x \\ hv_xv_z-v_y & hv_yv_z+v_x & N_z+hv_z^2 \end{pmatrix} \quad (1)$$

where $\mathbf{v} = \begin{bmatrix} v_x & v_y & v_z \end{bmatrix}^T = \mathbf{N} \times \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ and $h = \frac{1 - N_z}{1 - N_z^2}$. This matrix rotates the coordinates $[x_i, y_i, z_i]^T$ of a generic i-th point of the input data, as follows:

$$\begin{bmatrix} x_i' & y_i' & z_i' \end{bmatrix}^T = \mathbf{R} \times \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T \quad (2)$$

where $\begin{bmatrix} x_i' & y_i' & z_i' \end{bmatrix}^T$ are the resulting coordinates. Notice that also the skeleton joints are transformed in the new reference system. Fig. 3(b) and Fig. 3(c) show the skeleton and the point cloud after the alignment stage, respectively. It is important to observe that the point cloud alignment also improves the discrimination in space of each segment of the subject. As an example, the comparison of Fig. 2(b) and Fig. 3(c) reveals that the right arm of the framed person is further detached from the rest of body. This aspect is important for the proposed approach, as it will be discussed in the next subsections.

3.2 Skeleton standard posture and partition grid

The second step in data processing consists in the definition of the Skeleton Standard Posture (SSP) of the person by using the aligned skeleton joints. As previously stated, the SSP includes anthropometric information about the body shape of the subject and it will drive the process of signature extraction based on color collection. In fact, the SSP is used to produce a partition grid which divides the point cloud in different areas, where discriminative features can be defined.

The SSP is a new representation of the skeleton at a fixed posture. Although the proposed approach can work with any number of skeleton joints, the following discussions will consider skeletons made of 15 joints, labeled with the integer s = 1, …, 15 (s = 1 refers to the torso joint, s = 2 indicates the neck joint, and so on). Fig. 4(a) shows the fixed posture of the skeleton in the *XY*-plane of the global reference frame.

Starting from the origin of reference system (the torso joint s = 1), the SSP is built by placing the consecutive joints along specific directions, lying on the *XY*-plane, and keeping the same distances between them. As an example, the neck joint (s = 2) is located along the *Y*-axis at a distance from

the torso joint (s = 1) equal to the Euclidean distance between the neck and torso joints of the detected skeleton. Following this procedure, all the joints of the arms are properly located parallel to the X-axis. The leg joints, instead, are radially projected onto two lines making an angle of +45° and -45° with the Y-axis, respectively.

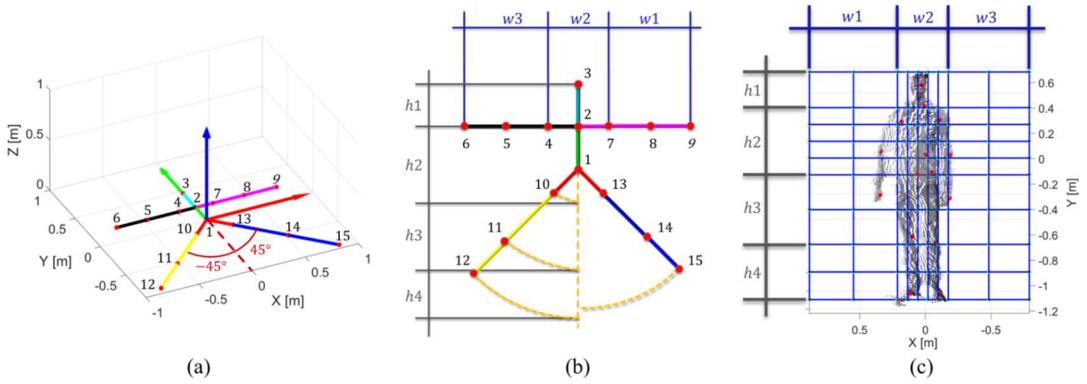


Fig. 4: (a) SSP computed using the 3D aligned joint. (b) Vertical and horizontal bands for the partition grid. (c) Example of partition grid made of 81 unevenly-spaced cells. The red dots represent the 3D joints.

The SSP is then used to create the partition grid shown in Fig. 4(c). It is of great importance to notice that the partition grid is distinctive of each person, since it depends on his specific shape. The partition grid consists of three vertical bands (w_1, w_2, w_3) and four horizontal bands (h_1, h_2, h_3, h_4) which are further divided in more cells, opportunely. In particular, as shown in Fig. 4(b), each vertical band $(w_p, p = 1, \dots 3)$ is defined by considering the position of arm joints. The horizontal bands $(h_q, q = 1, \dots 4)$, instead, are defined by considering the joints of head, neck and the radial projections of the leg joints onto the Y-axis. Differently from the vertical bands, the horizontal bands need an additional tuning due to a possible difference in the perceived legs length. For instance, in the case shown in Fig. 4(b), the horizontal bands h_2, h_3 and h_4 are defined by considering the joints of the left leg which is detected longer than the right one.

The obtained 3×4 partition grid is unevenly scaled to a final size P×Q in order to capture more information from regions which are likely expected to deliver more contents. P and Q values are defined taking into account the properties of the experimental setup, such as image resolution,

camera field-of-view, and mean distance of the subjects from the camera. In the case shown in Fig. 4(c), $P = Q = 9$: the central vertical band w_2 is equally sub-divided in five cells, whereas the other two bands w_1 and w_3 in two cells. On the other hand, the horizontal band h_2 is divided in four cells, whereas the h_3 and h_4 in two cells.

The partition grid, which depends on the 3D shape of each person, will be then used to collect color information at specific spatial positions, determined in agreement with the SSP of the person under testing.

3.3 Appearance statistics and signature computation

As previously stated, the partition grid divides the projection of the person point cloud onto the *XY*-plane in $P \times Q$ rectangular cells. The elements of the point clouds are collected in separate areas, whose position depends on the SSP of the person to be identified. Each bin is then investigated to obtain color-based descriptors.

Signatures are computed by transforming the RGB color space into another one with improved properties. In fact, in many computer vision applications, color plays an important role when some relevant features or information have to be extracted. As a matter of fact, some color spaces are formulated to aid the human understanding, whereas others are devised to help machines in data processing. Among the latter, the *CIE L*a*b** color space is highly suitable for image processing algorithms because of his significant properties [44]. Specifically, this color representation approximates the human vision and his perception of lightness, providing uniform distribution of colors [45]. The Euclidean distance between two colors in the *CIE L*a*b** model approximates the color difference that human eyes perceive. In this way, color variations are better discerned by using this mathematical model unlike other traditional ones (e.g. RGB, HSV). Consequently, more discriminative signatures can be extracted involving better re-identifications.

Hence, the RGB color information of a 3D point is converted in the corresponding *CIE Lab* one. As known, the L^* represents the luminance channel, whereas a^* and b^* channels stand for the color opponents green-red and blue-yellow of all perceivable colors. The L^* channel has a definition

range of [0, 100] of unsigned real values. Conversely, a^* and b^* channels are defined in the range [-128,127] of signed real values.

In the proposed method, the luminance channel is not considered in the computation of the signature, since it does not contain information useful for person re-identification. Most of the information carried by the luminance channel is actually related to the lighting conditions of the environment, which can oscillate unavoidably whenever the scene is enlightened by artificial light sources [46]. In this case, images can suffer from alterations of light levels, as effect of artificial light wavering, which produces global changes of the luminance channel. On the contrary, a^* and b^* channels contain the chromatic information, which is discriminative for the purpose of people re-identification. Once the colors of 3D point cloud are transformed in the CIE L*a*b* space, and the points of the input dataset are separated by the application of the partition grid, each cell is investigated to compute a 2D color histogram as a function of the a* and b* channels. The position of the bin centers of the histogram range by unit steps within the minimum and the maximum values of the two channels. Consequently, the final number of bins of each histogram depends on the color dynamics of the cell under investigation.

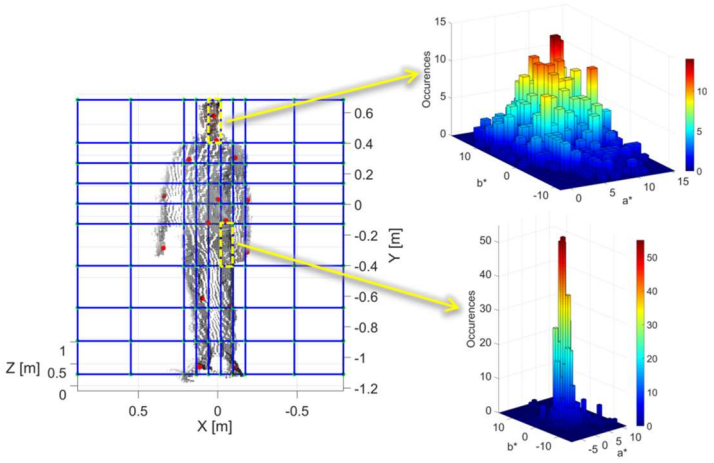


Fig. 5: 2D histograms related to the analysis of two cells of the partition grid. The histogram bins of color channels a^* and b^* are defined by the color dynamic of the cell under investigation.

For instance, cells with higher informative contents, i.e. chromatic contributions, produce histograms of more bins with respect to cells which are almost monochromatic. Accordingly, empty

cells lead to empty histograms. As an example, Fig. 5 shows two histograms collecting a^* and b^* channels. It has been empirically observed that the first five peaks of the 2D histograms carry the most contribution of the color dynamics within the cells. For this reason, each 2D histogram obtained from the analysis of the cells is shortened in terms of five tuples (n_m,a_m,b_m), $m = 1,\ldots,5$, which describe the most recurrent values of a^* and b^*. These tuples are made of the positions (a_m,b_m) of the first five peaks of the 2D histogram and n_m denotes the corresponding normalized number of occurrence of a_m and b_m for the single cell under investigation. Fig. 6 shows how the tuples of a generic i-th cell are arranged (red vector of 15 entries) to create the whole signature (blue vector of $P\times Q$ elements) of the person under investigation. It is important to notice that tuples, whose corresponding cells are empty, are all set to zero.

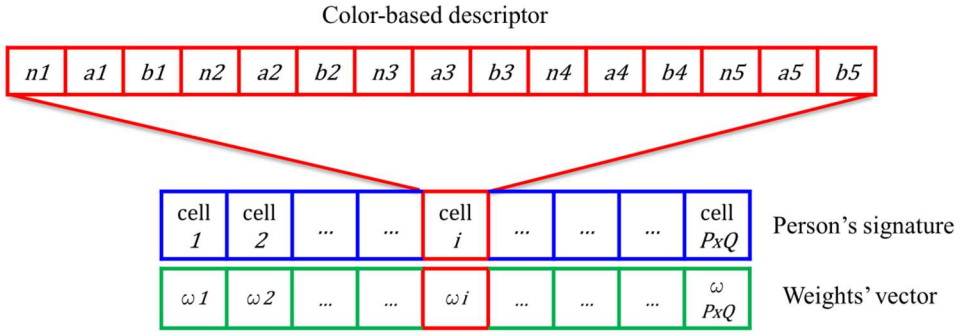


Fig. 6: The person descriptor is obtained by arranging the color statistics of each histogram into an array of $15\times P\times Q$ elements. Each cell is weighted differently during the signature comparisons. The weights of each row of partition grid are defined by opportunely sampling a Normal distribution function $F\sim N(0,s)$.

Finally, it is worth highlighting that the person signature is compared with the others by using a weighed Euclidean distance operator. Particularly, the color statistics of each cell of partition grid are weighted differently during the comparisons (see Fig. 6 for more details). The cells which mainly contain the body parts that move the more (arms and legs), are weighted less than the ones representing the more "stationary" body parts (i.e. torso, hips, shoulders and so on). In this way, the method is able to manage the body changes, thus still involving low distance values among signatures related to the same person, regardless its pose.

3.4 Reference DataBase re-projection

As previously stated, the person signature has to be compared with all the other signatures from a known database of users. These signatures are determined by exploiting the knowledge of the SSP of the subject under investigation.

In general, re-identification assigns the person under analysis to a class of a labelled database, which includes N_{users} possible classes. For each c-th class ($c = 1,\ldots,N_{users}$) the database contains $N_{c,pc}$ point clouds of the same known user. Once the partition grid of the subject under analysis has been constructed, it is used to re-project the signatures of all the point clouds of the N_{users} classes of the database. The resulting signatures belonging to the same class are then averaged in order to determine N_{users} mean signatures, one for each class. Finally, the signature of the person under testing and the mean signatures of the N_{users} classes are compared in terms of weighed Euclidean distance. These values provide the affinity of the user under analysis with respect to the N_{users} classes of the dataset. The minimum distance refers to the best matching of signatures, i.e. it provides the best assignment of the input person to the most likely class.

The novelty of the method is that the signatures of all the users of the reference database are projected on the basis of the same partition grid, i.e. the one already built for the extraction of the signature of the person to be identified. Notice that this partition grid depends on the SSP of the person under testing and, thus, it is strictly correlated to its body shape and size. In this way, when the 2D histogram analysis is applied, the color information is collected from corresponding areas in space, thus increasing discrimination between subjects of different shape. Fig. 7 better explains this concept: the partition grid of the subject to be identified is overlapped on the aligned point cloud of another person having different shape. In the case of Fig. 7, it is straightforward to observe that corresponding cells of the grid collect different portions of the two point clouds. As an example, the cells highlighted in Fig. 7 can include some portions of the point clouds or can be completely empty. These peculiarities lead to different color-based descriptors. Consequently, it is highly expected that people wearing clothes of similar colors, but having different shape (as in the example of Fig. 7), will be better differentiated during the matching step.

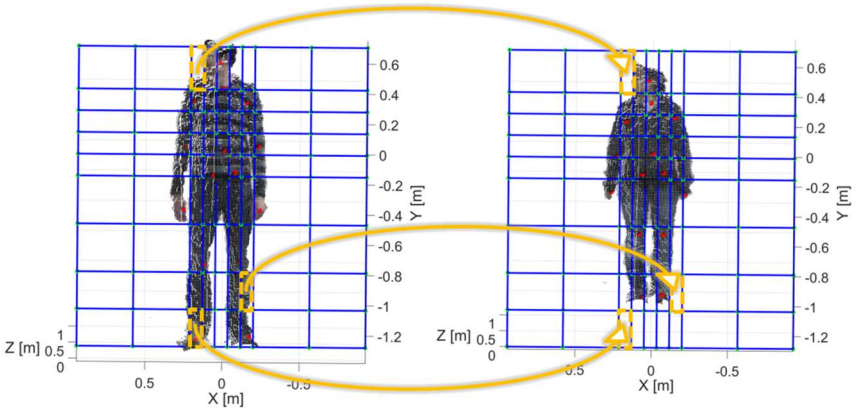


Fig. 7. Example of application of the same partition grid on people having different SSP.

4. Results and Discussions

This section reports the experiments and the corresponding outcomes of the proposed approach. Specifically, its effectiveness has been proven by evaluating the recognition performance on three different datasets: the BIWI RGBD-ID, the KinectREID and the RGBD-ID. These datasets meet the requirements needed for applying the proposed approach, since they provide both color point clouds and subject skeletons. All data are acquired simultaneously from the same typology of RGB-D sensors, placed at specific positions of the testing environments.

The BIWI RGBD-ID dataset consists of 50 subjects, captured by the Kinect v2 sensor. Video sequences are separated in training and testing videos. Since this dataset is targeted to long-term re-identification, testing sequences contain 28 subjects of the training one, but wearing different clothes. Since the algorithm works with color-based descriptors, testing sequences are not suitable. For this reason only the training data has been used for people re-identification.

On the contrary, the KinectREID and the RGBD-ID were acquired by using Kinect v1 sensors. The former contains video sequences of 71 subjects framed by considering three different viewpoints and variable light conditions, whereas the latter consists of 79 individuals acquired in two different indoor environments. Also in this case, very few sequences of the RGBD-ID dataset, where the people change their clothes, have been discarded from the computation. A total number of 769 video sequences of the RGBD-ID dataset have been considered.

The proposed method has been developed in Mathworks Matlab (R2015a 64-bit) [47]. All numerical tests have been run on a 8-GB-RAM system equipped by an Intel(R) Core^{TM} i5-3470 CPU having clock frequency of 3.20 GHz, running a Microsoft Windows 7 Professional 64-bit operating system.

4.1 Preliminary processing of input data

As stated in the previous section, preliminary processing is mandatory before the application of the proposed method, in order to clean the RGB-D datasets. For this reason, all the datasets considered within this paper have been processed using standard methods suited for the specific inputs.

In particular, two different sources of error can alter the input images:

- Random image noise on color data, such as Gaussian noise, salt-and-pepper noise and shot noise;
- Missed alignment between the color image and the corresponding depth map, which can also differ in resolution. This source of error only affects the BIWI RGBD-ID dataset.

In order to get an improvement of the image quality, undesirable noise effects on the RGB image are treated by means of standard median filters [48], having a structuring element of size 3×3. Afterwards, the color input data is further improved by means of a contrast-limited adaptive histogram equalization [49].

On the other hand, image registration has been applied in order to obtain the exact superposition of color and depth data in the case they do not match, as for the BIWI RGBD-ID dataset. This target is achieved by stretching the color image on the depth map using an affine transformation matrix, whose entries are determined during a preliminary calibration.

It is worth noticing that both pre-processing are empirically tuned on the specification of each dataset which will be considered for the next evaluations. As an example, Fig. 8 reports the application of this preprocessing phase on the input RGB image (Fig. 8(a)), which is filtered and equalized (Fig. 8(b)). In addition, a result of the registration of color and depth images are shown in Figs. 8(c-d), which report the application of a user mask, resulting from depth data, on the RGB

image, before and after the application of the affine transformation. The images in Fig. 8 are taken from the BIWI RGBD-ID dataset.

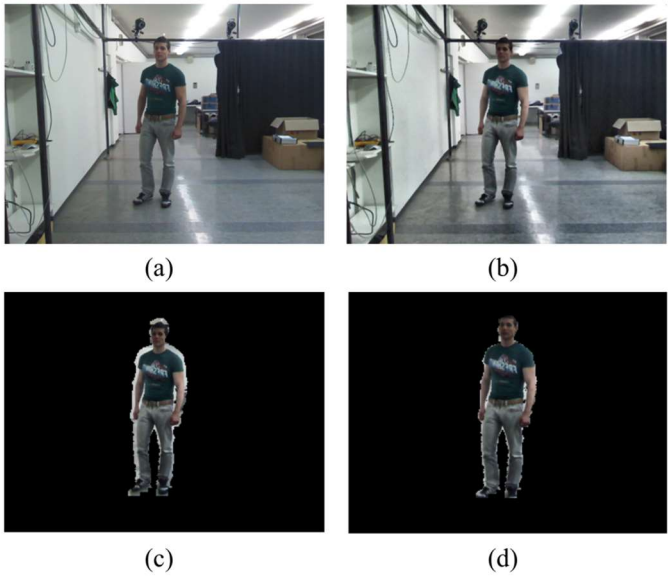


Fig. 8: First row: (a) Input image and (b) the corresponding one after the application of the median filter and the equalization step. Second row: Segmented color image before (c) and after (d) the affine transformation.

Finally, depth samples belonging to each subject are used to produce the 3D point cloud of the person under testing. Depth samples contain information about the distances between the surfaces of the targets which constitute the scene while the sensor performs the 3D model. Point clouds are obtained by treating this data via a pin-hole camera model, able to convert distances in 3D points in the world reference system (X, Y, Z) [50]. The resulting point cloud, corresponding to the data of Fig. 8, is shown in Fig. 9 with the line segments that represent the detected skeleton.

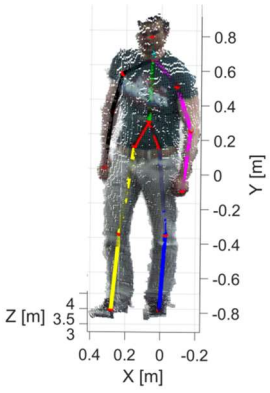


Fig. 9. Colored point cloud obtained by the application of the pin-hole model on the depth map.

4.2 Experimental analysis of public datasets

The next two sub-sections will discuss the results of the application of this method on the BIWI RGBD-ID, the KinectREID and RGBD-ID datasets. In particular, the BIWI RGBD-ID dataset will be used for testing the method on high-resolution data. On the other hand, the remaining datasets, KinectREID and the RGBD-ID will be used for providing comparisons with the current state of the art about people re-identification.

4.2.1 Tests on the BIWI RGBD-ID dataset

In the BIWI RGBD-ID, each person performs collaborative movements: two frontal walks towards the camera, one walk away from the acquisition system and several head movements and rotation actions. Following the initial hypothesis below the method, which states that re-identification is performed between point clouds acquired in short time, people are expected to be framed showing similar appearance or, equivalently, the same side of the body. For the sake of simplicity, BIWI RGBD-ID dataset has been pruned to discard people framed from their back. This result has been obtained by taking advantage of a face recognition algorithm [51].

Within the BIWI RGBD-ID, 11780 instances of $N_{users} = 50$ people have been selected. The dataset has been further divided in two subsets: the former (*source* set) is made of the users to be re-identified; the latter (*reference* set) encloses the point clouds of the labelled users. These subsets are defined by using a k-fold cross-validation partitioning [52], where k has been set to 10. For each iteration, about 10% of the instances populates the reference set, whereas all the remaining ones define the source set. Notice that the entries of the source and reference sets change at any iteration of the k-fold cross-validation partitioning. Given the SSP and the partition grid of the specific user under analysis, extracted from the source set, the method builds all the signatures of the reference set. Thus, these signatures are averaged to create a set of 50 mean signatures referred to each class of users. Once again, these mean signatures are peculiar of the specific user considered from the source set.

Tab. 1 reports the overall recognition scores referred to all the iterations. As reported in the table 1, the total average score is equal to 97.84%.

Tab. 1: Recognition rates obtained applying the proposed method on BIWI RGBD-ID database for each iteration of *k*-fold cross-validation partitioning. The forth iteration presents the lowest recognition rate in comparison with the others.

# k-fold iteration	Recognition rate [%]
1	98.31
2	98.10
3	97.63
4	**97.44**
5	97.50
6	97.99
7	98.04
8	97.77
9	97.68
10	97.92
Total average score	**97.84**

As reported in Tab. 1, the 4th iteration of the *k*-fold cross-validation partitioning gives the worst results in terms of recognition rate. Thus, the performance of the proposed re-identification algorithm, assessed by means of appropriate metrics, such as ROCs (Receiver Operating Characteristics), CMCs (Cumulative Matching Characteristics), confusion matrices and other statistical measures, will refer to this iteration.

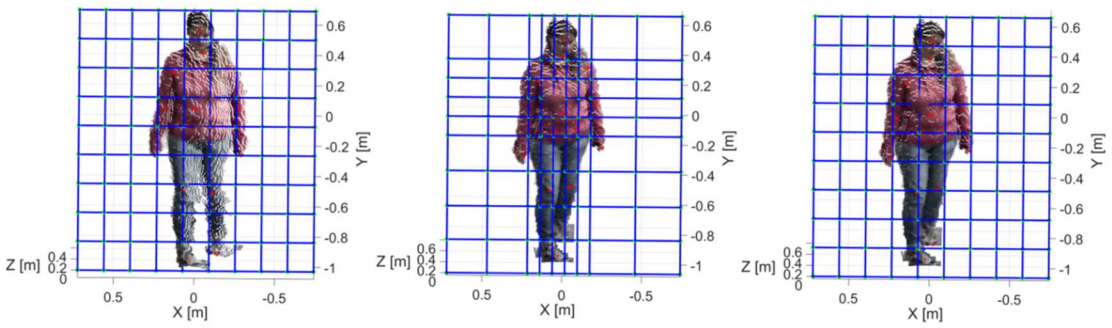


Fig. 10: (a) Aligned point cloud and evenly-spaced grid; (b) Not-aligned point cloud and unevenly-spaced grid; (c) Not-aligned point cloud and evenly-spaced grid.

The proposed method aligns the people pose and uses the unevenly-spaced partition grid. In the next experiments, three additional cases will be considered and the relative results will be shown. In

particular, they refer to the combined use of *aligned* and *not aligned point clouds* with *evenly* and *unevenly spaced* partition *grid*. Fig. 10 shows these cases: Fig. 10(a) reports an evenly-spaced partition grid over the aligned point cloud of a subject; Figs. 10(b-c) show the point cloud, without preliminary alignment, sampled using both the unevenly-spaced grid and the evenly-spaced grid, respectively.

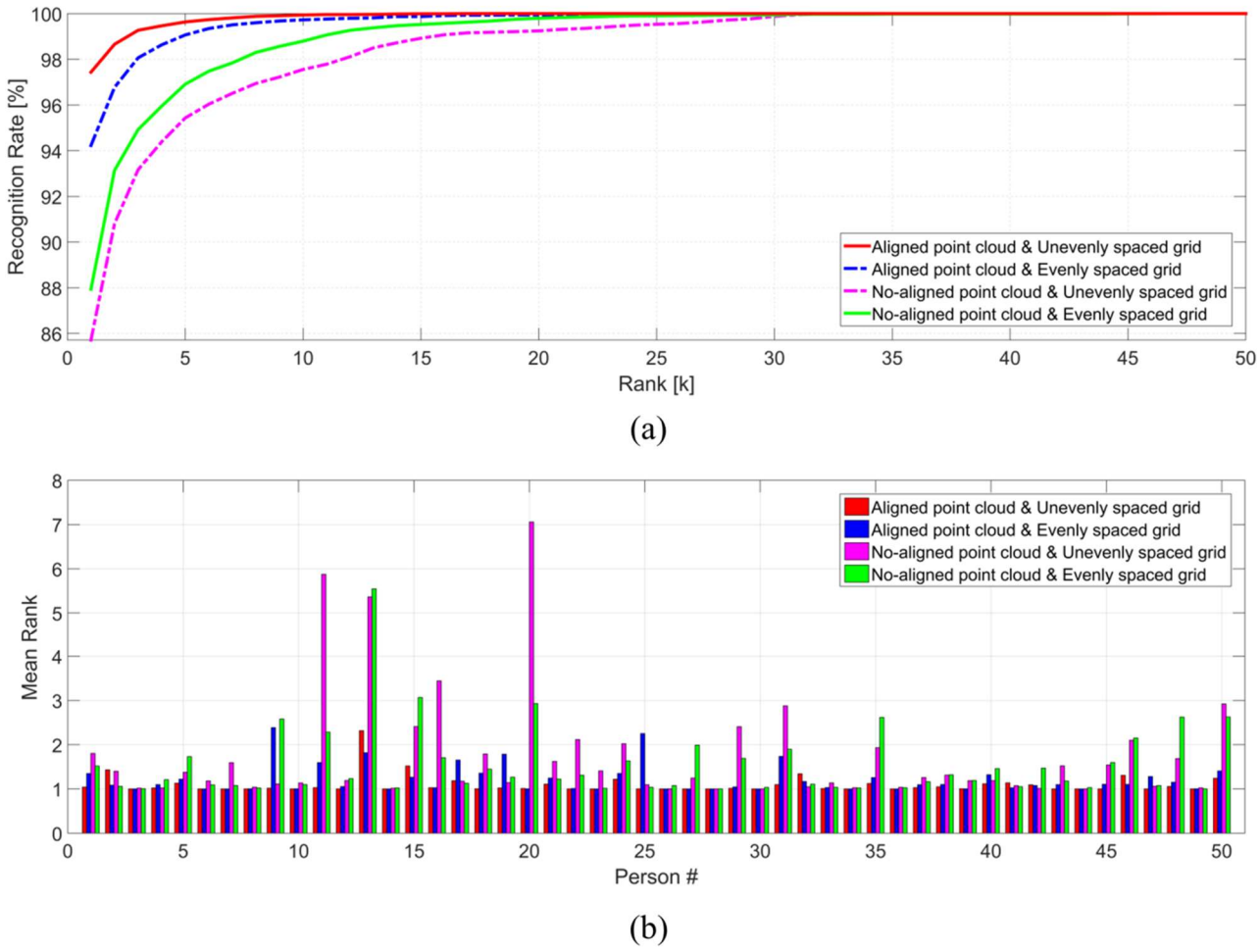


Fig. 11: (a) Cumulative matching characteristics and (b) average ranking histograms. The proposed method *APC-USG* (red curve) presents the best recognition performance in comparison with other cases.

Fig. 11(a) shows the CMC curves obtained in the considered cases. It is worth noticing that applying the proposed method (*Aligned Point Cloud-Unevenly Spaced grid*, or *APC-USG*), the rank-1 recognition rate reaches 97.44%, which is about 3.2% above to the rate of the *Aligned Point Cloud-Evenly Spaced Grid* case (*APC-ESG*). In case of no alignment, the performance further deteriorates in both cases of evenly- and unevenly-spaced grid. The rank-1 recognition scores are indeed, 87.94% and 85.71%, respectively.

Tab. 2 summarizes the statistics related to the CMC curves of Fig. 11(a). It also presents the values of normalized Area Under Curves (nAUC), which describes the area under the CMC curves divided by the total area of the ideal CMC curve, and of the rank-100%, which determines the lowest rank value at which the CMC curve reaches the 100%.

Tab. 2: Statistics for each CMC curve of Fig. 11(a).

Method	Rank-1 [%]	nAUC [%]	Rank-100%
APC-USG (proposed)	**97.44**	**99.90**	**17**
APC-ESG	94.23	99.71	44
NAPC-USG	85.71	98.66	34
NAPC-ESG	87.94	99.18	47

Fig. 11(b) reports the average ranking histograms, which specifies the average rank needed to identify a person correctly. It allows to investigate how the performance changes for each person of the dataset. As can be observed from the histograms, the *APC-USG* method shows the best mean ranks. In this regard, only person #13 is properly recognized at rank 2. This is due to the highly comparable appearance of subjects #13 and #35, who dress similar clothes. Nevertheless, the very low average ranks obtained for all people prove that the proposed method (*APC-USG*) is very effective and robust for person re-identification. Moreover the presented comparisons with respect to the other considered cases confirm once more that the alignment of point cloud and the non-equally spaced partitioning of the grid are very successful.

The performance of the proposed method are further shown in Fig. 12, where the confusion matrix related to *APC-USG* method and the ROC curves are presented. As shown in Fig. 12(a), most of the counts are located on the diagonal, which is much brighter than the other locations. In the proposed case, the confusion matrix indicates that only few predicted IDs have been incorrectly associated. Specifically, the highest count among the incorrect matches is found equal to 23 (see the red ellipse in Fig. 12(a)), which is negligible in comparison with the total number of items of the source set (10602 items). The ROC curves for all the previously described cases are shown in Fig. 12(b). They are obtained by plotting the true positive rate (TPR) against to the false positive rate (FPR).

It is possible to notice that the ROC curve (red line in Fig. 12(b)) of the presented method (*APC-USG*) is very close to an ideal curve, proving once more the reliability of this approach. Conversely, the other cases provide ROC curves always worse than the *APC-USG* one.

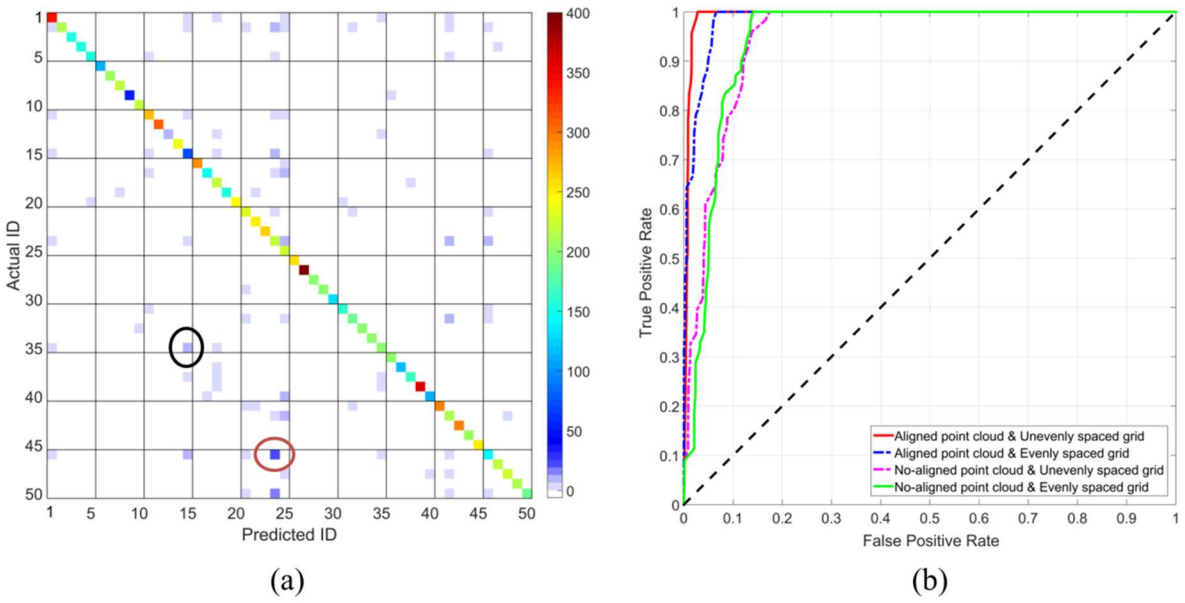


Fig. 12: (a) Confusion matrix (*APC-USG* method) and (b) Receiver Operating Characteristics of the different considered cases. The red ellipse in (a) indicates the couple of users that have been mainly mismatched.

For the sake of completeness, an example of re-identification error is shown in Fig. 13, with the RGB images of two mismatched people #15 and #35 (highlighted with a black ellipse in Fig. 12(a)). These two people wear similar clothes and have an almost comparable shape. These aspects might affect negatively the proposed algorithm. However, these erroneous recognitions occur rarely, as widely proven in the confusion matrix of Fig. 12(a).

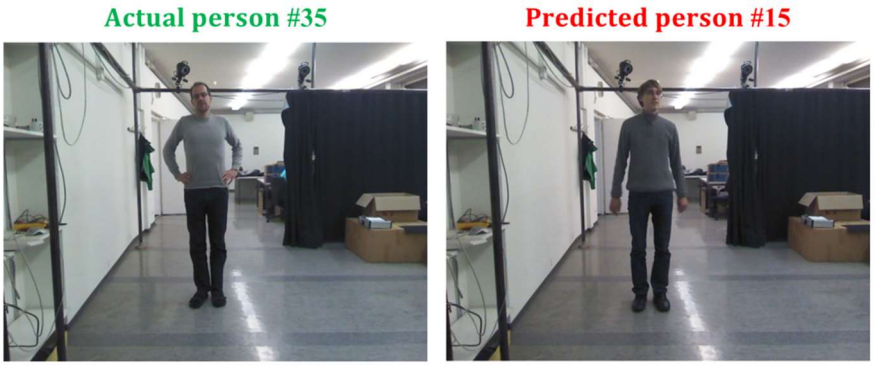


Fig. 13: Example of a re-identification error corresponding to the black ellipse in Fig. 12(a). .

Further tests have been performed in order to investigate other typologies of features for building the person signatures. In details, the main curvatures referred to the 3D points belonging to a single cell of partition grid have been extracted. By exploiting a part of the method presented in [53,54], the curvature vector for a cell under analysis, which collects the Gaussian and the mean curvatures, has been computed and then stitched to the color-based descriptor. Nevertheless, the addition of curvature information to the color one, has not brought any relevant advantage during the re-identification steps, leading even to worse results. A decrease of recognition score of about 17% has been observed with respect to the results reported in Tab. (2).

Similarly, the deep features have been also taken into account. The power of Deep Learning (DL) in computer vision applications is undoubted. The ability of extracting the best descriptive features for accomplishing a specific task, makes DL very powerful. Nevertheless, additional experiments based on deep feature extraction have proven the outperformance of our approach although of a small margin (about 5%).

Additional experiments have been run in order to further demonstrate the capability of the proposed approach in managing new entering users, who have not been already recognized or labeled.

As aforementioned, when the signature of a person to be recognized is compared with the reference database, a sorted list of users is given back. The first item of such a list indicates the most similar user among the labeled ones whereas, the last element points out the least similar one. The list is sorted according to the Euclidean distance values returned back after the signature comparison step. Under this assumption, the entrance of a new user, who has not been previously recognized, can lead to high distance values during the comparisons. Consequently, the introduction of an opportunely tuned threshold enables to assert if the user under investigation is a new one or not. In this regard, in case the distance value is over the determined threshold, a new class has to be introduced into the reference database. On the contrary, the signature of the recognized user is averaged with the one related to the person under analysis.

In the absence of an available open dataset, we have performed the tests using the BIWI RGBD ID dataset. In details, 5 users have been randomly removed from the reference set. Consequently, the reference database has been built by using the other remaining 45 users. We have considered all the source set related to the 50 users to validate the method. After a tuning step of threshold value, it has been possible to observe that in very limited cases a new user has been identified as an already cataloged one. In fact, the probability of recognizing the new users is found equal to 99.63%. In addition, an average recognition score of about 94% is further achieved as soon as the new users are investigated and inserted as new classes. The choice of the threshold value depends on the similarity between the new users and the other elements contained in the database, and represents a tradeoff between the ability of discerning new entrance and the possibility of generate false new candidates with people already observed.

4.2.2 Tests on the KinectREID and RGBD-ID datasets

Other experiments have been performed on the KinectREID and the RGBD-ID datasets. In this case, back-view images of people are not discarded from the database in order to compare the proposed approach with the method presented in [29].

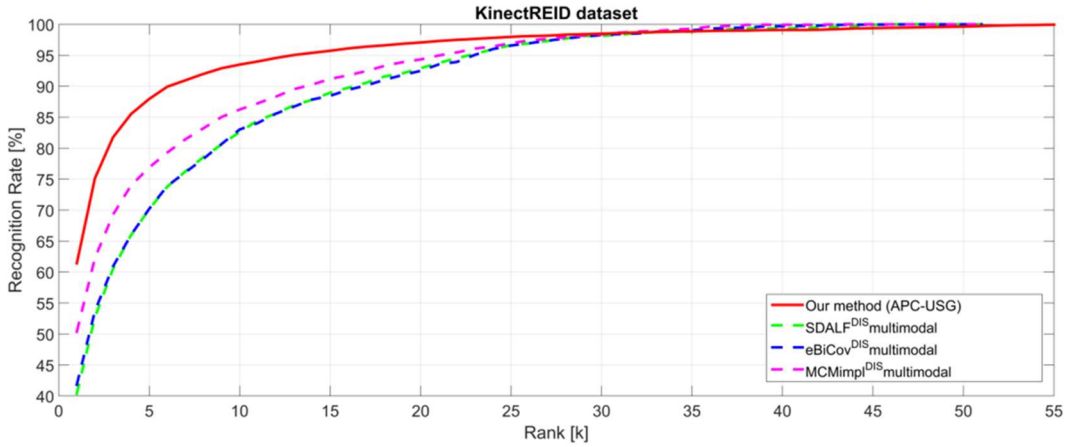
In Fig. 14 the CMC curves of the methods presented in [29] are shown together with those of the proposed one. In this case a k-fold cross-validation partitioning with $k = 4$ has been set. Only the CMC curves related to the worst recognition score among the four iterations are reported in Fig. 14.

As observable, the proposed approach (*APC-USG*) outperforms the other methods on the KinectREID dataset by considering the first ranks. A rank-1 recognition rate of 61.41% is achieved by the *APC-USG*: it improves the $MCMimpl^{DIS}$*multimodal* results of about 11% at rank-1.

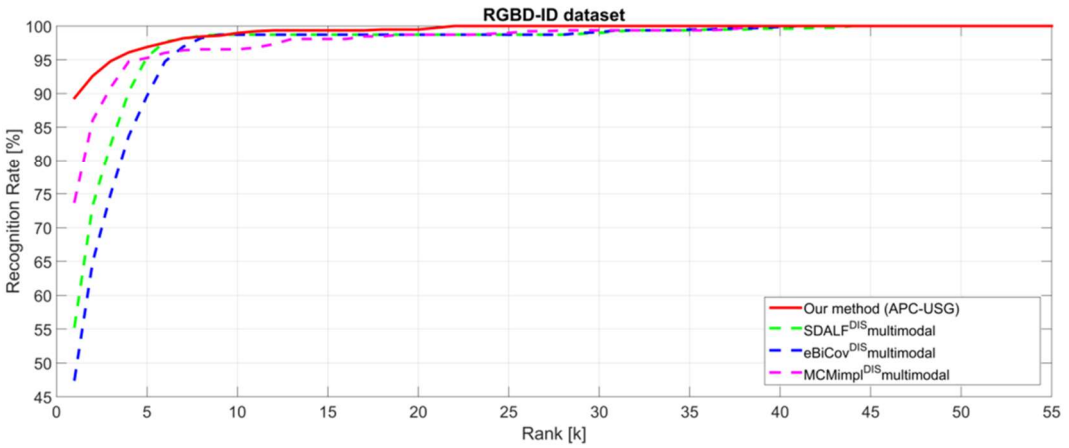
The recognition performance on the RGBD-ID dataset is reported in Fig. 14(b). Similarly to the previous case, the *APC-USG* performs well in comparison with other methods. In this regard, better recognition rates are obtained, especially at the first ranks (from rank-1 to rank-7) which are the most relevant for re-identification tasks. The rank-1 rate of *APC-USG* is equal to 89.34%, against to the 73.85% of $MCMimpl^{DIS}$*multimodal*. An improvement of about 15% is achieved. The

SDALF^{DIS}*multimodal* and *eBiCov*^{DIS}*multimodal* perform much worse by considering the first six ranks. Further statistics of CMC curves on both datasets are summarized in Tab. 3.

As reported in the Tab. 3, the proposed method shows the best recognition metrics with respect to other algorithms (see the bold values), except for the rank-100% on the KinectREID dataset. In details, the *APC-USG* needs 52 ranks for ensuring the 100% of recognition, whereas the *MCMimpl*^{DIS}*multimodal* requires less ranks (43 ranks). Nevertheless, this aspect is negligible since the first ranks are the most significant for people re-identification. Moreover, the proposed approach further improves the nAUC values of about 4% and 1% for the KinectREID and the RGBD-ID dataset, respectively, with reference to the best results attainable with the other methods under analysis.



(a)



(b)

Fig. 14: CMC curves related to (a) the KinectREID and (b) the RGBD-ID datasets. The superscript *DIS* in the legend, refers to the MCD (Multiple Component Dissimilarity) descriptor described in [29].

The recognition rates related to the BIWI RGBD-ID dataset are much higher than the scores obtained analyzing the other datasets considered in this subsection. This is due to the different RGB-D cameras used to create the datasets. Specifically, the KinectREID and RGBD-ID datasets are built by using the Microsoft Kinect v1, based on structured-light projection, whereas the BIWI RGBD-ID dataset is built by using the Microsoft Kinect v2, which is based on time-of-flight technology [55]. The latter camera ensures more accurate measurements than the former in terms of depth and image resolution. As expected, better camera performance can produce more accurate input point clouds of increasing sizes (i.e. number of samples). This clearly enables the best recognition rates of the BIWI RGBD-ID dataset with respect to the others. Furthermore, different viewpoints and variable ambient light conditions are taken into account in the KinectREID and RGBD-ID datasets, which also consider the back views of the subjects under evaluation. It further downs the recognition rates with respect to the BIWI RGBD-ID one.

Tab. 3: Statistics for each CMC curve reported in Fig. 14. Bold values are the best for each column.

Method	KinectREID dataset			RGBD-ID dataset		
	Rank-1 [%]	**nAUC [%]**	**Rank-100%**	**Rank-1 [%]**	**nAUC [%]**	**Rank-100%**
APC-USG (proposed)	**61.41**	**96.99**	52	**89.34**	**99.49**	**22**
SDALF^{DIS}multimodal	40.37	91.08	46	55.38	97.91	44
eBiCov^{DIS}multimodal	41.81	91.10	47	47.56	97.32	41
MCMimpl^{DIS}multimodal	50.37	92.97	**43**	73.85	98.32	39

For the sake of completeness, we report also the outcomes related to [35], where the deep learning approach has been explored for the people re-identification. The Multi-Modal Uniform Deep Lerning (MMUDL) method shows very outstanding recognition rates. Specifically, on the reduced RGBD-ID dataset, it achieves the 100% of recognition at first rank whereas, it reaches the 97.0 % on the KinectREID dataset. Instead when the complete RGBD-ID dataset is considered, the re-identification rate decreases to 76.7%.

By comparing these outcomes with ours, one can notice that the MMUDP outperforms our approach. However, as before mentioned in the discussion of related works, the MMUDP suffers to

manage new situations. In this regard, the CNN extract features that strictly depend on the set of images on which have been trained. When a new camera or a different point of view is considered, the CNN needs to be retrained. On the contrary, our method does not require any learning steps or ad-hoc tuning, and it manages all the new situations in the same way of old ones, still ensuring good recognition rates.

A final consideration regards the whole processing time of the propose method. The average time spent for analyzing the data referred to one person of the source set is found equal to about 162 ms. This value includes the needed time for the data loading (about 35 ms), the preliminary data processing (about 42 ms) and the extraction of person signature (about 85 ms). The re-projections of the person signatures in the DB require about 150 s by considering 1000 instances in the database.

By implementing the code on general-purpose programming languages (e.g. C, C++, C# etc.) and by using the parallel computing, it is possible to reduce dramatically the required processing time thus enabling the real-time usage of the method even with larger database.

5. Conclusions

In this paper an accurate method for people re-identification has been presented. Input data, returned by RGB-D cameras, are used to compute a color-based descriptor for the subjects under investigation, by taking advantage of the formulation of a skeleton standard posture.

After a preliminary step of point cloud preprocessing, 3D point clouds are aligned exploiting the skeletal information. Then, the skeleton joints of the SSP are used to define a partition grid having cells of different sizes. This grid properly divides the color point cloud in several cells. Each cell is then investigated to extract statistical information about the color distribution. This information is then arranged into a one-dimensional array, which constitutes the person signature of the user under analysis. The same partition grid of the person to be identified is used for re-projecting the reference signatures of the labelled people of a known database. Final comparisons in the signature space allow for the labelling of the person under analysis to a specific ID.

The validity of presented method in terms of recognition rate has been proven by performing different experiments on three publicly available datasets. Recognition rates of 97.84% (subset of BIWI RGBD-ID), 61.97% (KinectREID) and 89.71% (RGBD-ID) have been obtained by applying the proposed method, demonstrating its outperformance with respect to the current state-of-the-art. Future activities will involve the evaluation of more challenging datasets, wherein occlusions and overcrowding problems may occur. Also the use of heterogeneous sensors, based on different technologies, will be subject of further investigations. Additional features based on edges, shapes and textural information will be further analyzed in order to explore new promising research directions.

Acknowledgements

The authors would like to thank Mr. Michele Attolico and Mr. Giuseppe Bono for their valuable advices about this work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Leo, M., Mosca, N., Spagnolo, P., Mazzeo, P. L., D'Orazio, T., & Distante, A. (2008, July). Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval (pp. 525-534). ACM.

[2] Munaro, M., & Menegatti, E. (2014). Fast RGB-D people tracking for service robots. Autonomous Robots, 37(3), 227-242.

[3] Aziz, K. E., Merad, D., & Fertil, B. (2011, August). People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation. In Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on (pp. 303-308). IEEE.

[4] D'Orazio, T., & Guaragnella, C. (2015). A survey of automatic event detection in multi-camera third generation surveillance systems. International Journal of Pattern Recognition and Artificial Intelligence, 29(01).

[5] Fosty, B., Crispim-Junior, C. F., Badie, J., Bremond, F., & Thonnat, M. (2013, November). Event recognition system for older people monitoring using an RGB-D camera. In ASROB-Workshop on Assistance and Service Robotics in a Human Environment.

[6] Cicirelli, G., Attolico, C., Guaragnella, C., & D'Orazio, T. (2015). A kinect-based gesture recognition approach for a natural human robot interface. International Journal of Advanced Robotic Systems, 12(3).

[7] D'Orazio, T., Marani, R., Renó, V., & Cicirelli, G. (2016). Recent trends in gesture recognition: how depth data has improved classical approaches. Image and Vision Computing, 52, 56-72.

[8] Mosca N., Renò, V., Marani, R., Nitti, M., D'Orazio, T., & Stella, E. (2018). Human Walking Behavior detection with a RGB-D sensors network for ambient assisted living applications. CEUR Workshop Proceedings, 3rd Italian Workshop on Artificial Intelligence for Ambient Assisted Living. Bari, Italy, Vol. 2061, (pp. 17-29), 2018.

[9] Dantcheva, A., Velardo, C., D'angelo, A., & Dugelay, J. L. (2011). Bag of soft biometrics for person identification. Multimedia Tools and Applications, 51(2), 739-777.

[10] Jain, A., Flynn, P., & Ross, A. A. (Eds.). (2007). Handbook of biometrics. Springer Science & Business Media.

[11] Xtion PRO. On-line available: https://www.asus.com/3D-Sensor/Xtion_PRO/ (Accessed 2017-09-19).

[12] Kinect v1. On-line available: https://msdn.microsoft.com/en-us/library/hh855355.aspx (Accessed 2017-09-19).

[13] Kinect v2. On-line available: https://developer.microsoft.com/en-us/windows/kinect/hardware (Accessed 2017-09-19).

[14] Bumblebee 2. On-line available: https://www.ptgrey.com/stereo-vision-cameras-systems (Accessed 2017-09-19).

[15] SDK Windows. On-line available: https://www.microsoft.com/en-us/download/details.aspx?id=44561 (Accessed 2017-09-19).

[16] OpenNI. On-line available: http://openni.ru/files/nite/index.html (Accessed 2017-09-19).

[17] Skeleton tracking. On-line available: https://msdn.microsoft.com/en-us/library/hh973074.aspx (Accessed 2017-09-19).

[18] Bedagkar-Gala, A., & Shah, S. K. (2014). A survey of approaches and trends in person re-identification. Image and Vision Computing, 32(4), 270-286.

[19] Mazzon, R., Tahir, S. F., & Cavallaro, A. (2012). Person re-identification in crowd. Pattern Recognition Letters, 33(14), 1828-1837.

[20] T. D'Orazio and G. Cicirelli, "People Re-Identification and Tracking from Multiple Cameras: a Review", IEEE International Conference on Image Processing (ICIP 2012), Sept 30 – Oct 3, 2012, Orlando, Florida, USA

[21] Lorenzo-Navarro, J., Castrillón-Santana, M., & Hernández-Sosa, D. (2013). On the use of simple geometric descriptors provided by RGB-D sensors for re-identification. Sensors, 13(7), 8222-8238.

[22] Liciotti, D., Paolanti, M., Frontoni, E., Mancini, A., & Zingaretti, P. (2016, December). Person Re-identification Dataset with RGB-D Camera in a Top-View Configuration. In International Workshop on Face and Facial Expression Recognition from Real World Videos (pp. 1-11). Springer, Cham.

[23] Barbosa, I., Cristani, M., Del Bue, A., Bazzani, L., & Murino, V. (2012). Re-identification with rgb-d sensors. In Computer Vision–ECCV 2012. Workshops and Demonstrations (pp. 433-442). Springer Berlin/Heidelberg.

[24] Gharghabi, S., Shamshirdar, F., Shangari, T. A., & Maroofkhani, F. (2015, June). People re-identification using 3D descriptor with skeleton information. In Informatics, Electronics & Vision (ICIEV), 2015 International Conference on (pp. 1-5). IEEE.

[25] Munaro, M., Fossati, A., Basso, A., Menegatti, E., & Van Gool, L. (2014). One-shot person re-identification with a consumer depth camera. In Person Re-Identification (pp. 161-181). Springer London.

[26] Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2197-2206).

[27] Satta, R., Pala, F., Fumera, G., & Roli, F. (2013, February). Real-time Appearance-based Person Re-identification Over Multiple KinectTM Cameras. In VISAPP (2) (pp. 407-410).

[28] Martinel, N., Micheloni, C., & Piciarelli, C. (2013, October). Learning pairwise feature dissimilarities for person re-identification. In Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on (pp. 1-6). IEEE.

[29] 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(4), 788-799.

[30] Oliver, J., Albiol, A., & Albiol, A. (2012, November). 3D descriptor for people re-identification. In Pattern Recognition (ICPR), 2012 21st International Conference on (pp. 1395-1398). IEEE.

[31] Baltieri, D., Vezzani, R., & Cucchiara, R. (2015). Mapping appearance descriptors on 3d body models for people re-identification. International Journal of Computer Vision, 111(3), 345-364.

[32] Baltieri, D., Vezzani, R., & Cucchiara, R. (2011, December). 3dpes: 3d people dataset for surveillance and forensics. In Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding (pp. 59-64). ACM.

[33] Ahmed, E., Jones, M., & Marks, T. K. (2015). An improved deep learning architecture for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3908-3916).

[34] Schumann, A., Stiefelhagen, R. (2017). Person Re-identification by Deep Learning Attribute Complementary Information. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops , (pp. 1435-1443), July 2017.

[35] Ren, L., Lu, J., Feng, J., & Zhou, J. (2017). Multi-modal uniform deep learning for RGB-D person re-identification. Pattern Recognition, 72, 446-457.

[36] BIWI RGBD-ID. On-line available: http://robotics.dei.unipd.it/reid/ (Accessed 2017-04-28).

[37] KinectREID. On-line available: http://pralab.diee.unica.it/it/PersonReIdentification (Accessed 2017-04-28).

[38] RGBD-ID. On-line available: http://pavis.iit.it/datasets/rgbd-id (Accessed 2017-04-28).

[39] Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M. (2010, June). Person re-identification by symmetry-driven accumulation of local features. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 2360-2367). IEEE.

[40] Ma, B., Su, Y., & Jurie, F. (2014). Covariance descriptor based on bio-inspired features for person re-identification and face verification. Image and Vision Computing, 32(6), 379-390.

[41] Satta, R., Fumera, G., Roli, F., Cristani, M., & Murino, V. (2011). A multiple component matching framework for person re-identification. Image Analysis and Processing–ICIAP 2011, 140-149.

[42] Rusu, R. B., Marton, Z. C., Blodow, N., Dolha, M., & Beetz, M. (2008). Towards 3D point cloud based object maps for household environments. Robotics and Autonomous Systems, 56(11), 927-941.

[43] Möller, T., & Hughes, J. F. (1999). Efficiently building a matrix to rotate one vector to another. Journal of graphics tools, 4(4), 1-4.

[44] Robertson, A. R. (1977). The CIE 1976 Color Difference Formulae. Color Research & Application, 2(1), 7-11.

[45] Harold, R. M. (2001). An introduction to appearance analysis. GATFWORLD, 13(3), 5-12.

[46] Renò, V., Marani, R., Nitti, M., Mosca, N., D'Orazio, T., & Stella, E. (2017). A Powerline-Tuned Camera Trigger For AC Illumination Flickering Reduction. IEEE Embedded Systems Letters, 99.

[47] Mathworks Matlab. On-line available: https://ch.mathworks.com/ (Accessed 2017-09-25).

[48] Lim, J. S. (1990). Two-dimensional signal and image processing. Englewood Cliffs, NJ, Prentice Hall, 1990.

[49] Zuiderveld, Karel. "Contrast Limited Adaptive Histograph Equalization." Graphic Gems IV. San Diego: Academic Press Professional, 1994. 474–485.

[50] Hartley, R., & Zisserman, A. (2003). Multiple view geometry in computer vision. Cambridge university press, (pp. 153-155).

[51] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE.

[52] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, pp. 149-151.

[53] Martino, F., Patruno, C., Marani, R., & Stella, E. Signature Extraction from 3D Point Clouds using Frame Theory for Environmental Modeling. In Proc. of 8th International Conference on Sensing Technology (ICST), Liverpool, UK, (pp. 593-598), 2014,.

[54] Martino, F., Patruno, C., Marani, R., & Stella, E. An Application of the Frame Theory for Signature Extraction in the Analysis of 3D Point Clouds. In Next Generation Sensors and Systems (pp. 289-310). Springer, Cham, 2016.

[55] Sarbolandi, H.; Lefloch, D.; Kolb, A. Kinect Range Sensing: Structured-Light versus Time-of-Flight Kinect. Comput. Vision and Image Understanding 2015, 139, 1–20.