sensors

MDPI

Article

Person Re-Identification with RGB-D Camera in Top-View Configuration through Multiple Nearest Neighbor Classifiers and Neighborhood Component Features Selection

Marina Paolanti *, Luca Romeo, Daniele Liciotti, Rocco Pietrini, Annalisa Cenci, Emanuele Frontoni and Primo Zingaretti

Department of Information Engineering, Universitá Politecnica delle Marche, I-60131 Ancona, Italy; l.romeo@univpm.it (L.R.); d.liciotti@pm.univpm.it (D.L.); r.pietrini@pm.univpm.it (R.P.); a.cenci@pm.univpm.it (A.C.); e.frontoni@univpm.it (E.F.); p.zingaretti@univpm.it (P.Z.)

* Correspondence: m.paolanti@pm.univpm.it

Received: 30 August 2018 ; Accepted: 11 October 2018 ; Published: 15 October 2018

Abstract: Person re-identification is an important topic in retail, scene monitoring, human-computer interaction, people counting, ambient assisted living and many other application fields. A dataset for person re-identification TVPR (Top View Person Re-Identification) based on a number of significant features derived from both depth and color images has been previously built. This dataset uses an RGB-D camera in a top-view configuration to extract anthropometric features for the recognition of people in view of the camera, reducing the problem of occlusions while being privacy preserving. In this paper, we introduce a machine learning method for person re-identification using the TVPR dataset. In particular, we propose the combination of multiple k-nearest neighbor classifiers based on different distance functions and feature subsets derived from depth and color images. Moreover, the neighborhood component feature selection is used to learn the depth features' weighting vector by minimizing the leave-one-out regularized training error. The classification process is performed by selecting the first passage under the camera for training and using the others as the testing set. Experimental results show that the proposed methodology outperforms standard supervised classifiers widely used for the re-identification task. This improvement encourages the application of this approach in the retail context in order to improve retail analytics, customer service and shopping space management.

Keywords: RGB-D camera; person re-identification; machine learning; K-nearest neighbors; retail

1. Introduction

Camera installations are widespread in several domains, from small business and large retail applications, to home surveillance applications, environment monitoring, facility access, sports venues and mass-transit. Identification cameras are widely employed in most public places like malls, office buildings, airports, stations and museums. In these applications, it is desirable to identify different instances or images of the same person, recorded at different moments, as belonging to the same subject. This kind of process, commonly known as "person re-identification" (re-id), has a wide range of applications and is of great commercial value.

Research in people behavior analysis has been thoroughly focused on person re-id during the last decade, which has seen the exploitation of many paradigms and approaches of pattern recognition [1–3]. In challenging situations, algorithms need to be robust to be able to deal with issues such as widely-varying camera viewpoints and orientations, rapid changes in the appearance of clothing, occlusions, varying poses and various lighting conditions [4,5].

The first studied re-id problem was related to vehicle tracking and traffic analysis, where objects move in well-defined paths, have almost uniform colors and are rigid. Features like color, speed, size and lane position are generally embedded in Bayesian frameworks. However, person re-id requires more elaborate methods in order to deal with the widely-varying degrees of freedom of a person's appearance [6].

Much of the research on person re-id has been devoted to modeling human appearance. In fact, descriptors of image content have been proposed in order to discriminate identities while compensating for appearance variability due to changes in illumination, pose and camera viewpoint. Re-id is also a learning problem in which either metrics or discriminative models are actually learned [5,7]. Labeled training data are required for metric learning approaches, and new training data are needed whenever a camera setting changes [8].

Recently, person re-id has emerged as a very interesting tool for detection and tracking of people under occlusion or partial camera coverage. In a retail environment, re-id can provide useful information for improving customer service and shopping space management. In fact, changes in consumer purchase behavior led retailers to adapt their businesses, the products and services provided, as well as the way they communicate with customers. In the retail field, person re-id becomes a useful tool to recognize consumers in a store properly, to study returning consumers and to classify different shopper clusters and targets. The customer interactions such as (i) the level of attraction (i.e., attraction that the shelf is creating for consumers), (ii) the attention (i.e, the time consumers spend in front of a brand display) and (iii) the action (i.e., the number of consumers that enter the store and interact with particular merchandise) can be closely monitored through RGB-D cameras. This solution provides affordable and additional rough depth information coupled with visual images, offering sufficient accuracy and resolution for indoor applications. A distributed RGB-D camera has already been successfully applied in the retail field to identify customers univocally and to analyze behaviors and interactions with shoppers [9,10]. The choice of the RGB-D camera in a top-view configuration is preferred due to its greater suitability compared with a front view configuration, usually adopted for gesture recognition or even for video gaming. The top-view configuration reduces the problem of occlusions and has the advantage of being privacy preserving because a person's face is not recorded by the camera [11]. Top-view people counting applications are the most accurate (with accuracy up to 99%) even in very crowded scenarios (more than three people per square meter) [12]. The point of view of the camera in the top-view configuration is also the only one that allows measuring anthropometric features of the people passing by and interactions among shoppers and products on the shelf at the same time [13,14]. However, this configuration may lead to an important limitation: it does not allow one to retrieve features related to the front view that are widely employed in other state-of-the-art approaches (e.g., [15,16]), in which the subject identification can be highly discriminative. Hence, the proposed approach including the feature extraction and the classification stage was designed according to this challenging setup.

Currently, several datasets using RGB-D technology are available for the study of person re-id and cover many aspects of this problem, such as shape deformation, occlusions, illumination changes, very low resolution images and image blurring [17]. The most popular are VIPeR [18], the iLIDSmulti-camera tracking scenario [19], ETHZ [20], CAVIAR4REID [21] and [22]. However, since these datasets are not collected in a top-view configuration, they are not suitable for our purposes.

In this regard, we have built a new dataset for person re-id that uses an RGB-D camera in a top-view configuration: the TVPR (Top View Person Re-identification) dataset [23], using an Asus Xtion Pro Live RGB-D camera, which allows the acquisition of color and depth information in an affordable and fast way [24]. The camera was installed on the ceiling above the area to be analyzed. This dataset includes the data of 100 people, acquired across intervals of days and at different times.

Differently from [23], the main goal of the paper comprises the introduction of the feature extraction and classification stage for the re-id task in a top-view configuration scenario using a set of features extracted by the color and depth images. The overall system comprises the recording

stage, the pre-processing/feature extraction stage and the classification stage. Thus, we have tested the approach using the TVPR dataset [23] with respect to other state-of-the-art classifiers in order to measure the reliability and the effectiveness of our approach. In particular, we propose an ensemble method, named Multiple K-Nearest Neighbor (MKNN), based on the combination of different k-Nearest Neighbor (K-NN) classifiers. The problem of combining different K-NN has been addressed in [25–27] respectively for different feature subsets and different distance functions. The main contributions of this work with respect to the existing literature are: (i) the adoption of different distance functions for each single K-NN based on the nature of the feature descriptors, (ii) the introduction of Neighborhood Component Feature Selection (NCFS) for the anthropometric features, (iii) the overall combination method and (iv) the application of the following methodology on the TVPR dataset collected by the authors in a previous work [23]. The motivation for the usage of the specific method, i.e., MKNN, arose from the need to exploit the informative power of depth and RGB input properly combining the different nature of each feature. Although the authors combined different existing classifiers in an ensemble strategy, the way these classifiers were chosen and combined represents the main advantage of the proposed classification stage. The experimental results demonstrated the effectiveness of the proposed approach, encouraging its application in public contexts and in different real-world applications (e.g., safety and security in crowded environments, access control), where the top-view configuration allows reducing the problem of occlusions and privacy.

Each K-NN is trained by different distance functions and feature subsets. The neighborhood component feature selection is applied to the depth features to find the optimal weights, while cosine distance and Spearman's rank correlation are applied to measure the similarity between two RGB feature points. Instead of the standard majority vote method, we propose a variation of the Bayesian approach for combining the decision of different K-NN. The performance evaluation encourages the reliability and the effectiveness of the proposed approach. The MKNN methodology decreases the generalization error compared to the baseline K-NN method, outperforming supervised classifiers used for the re-id task (i.e., K-Nearest Neighbors (K-NN) [28], Decision Tree (DT) [29] and Random Forest (RF) [30,31]).

The paper is organized as follows: Section 2 provides a description of the approaches in the context of re-id (Section 2.1) and the characterization of the TVPR dataset (Section 2.2). Section 3 gives details on the proposed methodology for the feature extraction stage and the machine learning model implemented. Section 4 provides the experimental results and comparison with respect to baseline classifiers. The conclusions and future work in this direction are proposed in Section 5.

2. Background

This section presents an overview of the main approaches in the context of person re-id. In particular, Section 2.1 provides a review/summary of the literature on person re-id methods, and Section 2.2 gives details on the TVPR dataset for person re-id in a top-view configuration.

2.1. Previous Works on Person Re-Identification

Over the past few years, in the field of object recognition, the re-id problem has received considerable attention, and various reviews and surveys are available, pointing out different aspects of this topic [32,33]. Among the proposed approaches, four different classes could be defined, mainly depending on the camera setup and environmental conditions: biometric, geometric, appearance-based and learning approaches.

In the biometric approaches, the different person instances are matched together and are assigned to the same identity by the use of biometric features. The examples adopted in the real situation involve gait, faces, fingerprints, iris scans, and so on [34,35]. They are reliable and effective solutions, but these require a collaborative behavior of the people and suitable sensors. Thus, in the case of low resolution, poor views and a non-collaborative public, as in the case with common settings for surveillance cameras, these techniques are not often applicable.

The geometric approaches occur when more than one camera or sensor simultaneously collects information of the same area, and geometric relations among the fields of view (homographies, epipolar lines, and so on) can be adopted to match the data [18,36,37]. The geometric relations, when available, guarantee strong matches or, at least, a stiff candidate selection.

In the general case, only the appearance of the different items can be adopted [38,39]. In the appearance-based approaches, re-id can be correctly done only if the appearance is preserved among the views. It consists of exploiting dress colors and textures, perceived heights and other similar cues and can be considered a soft-biometric approach. Occlusions, illumination changes, different sensor qualities and different viewpoints are some of the challenging issues that make the appearance-based re-id difficult to implement. In [18], Gray et al. for the first time considered the problem of appearance models for person recognition, reacquisition and tracking. Until then, these problems had been evaluated independently, so they called for metrics that apply to complete systems [40,41]. A standard protocol to compare the results is proposed. This is done using the Cumulative Matching Curve (CMC) and introducing the VIPeR dataset for re-id. In [42], an algorithm was proposed that learns a domain-specific similarity function using an ensemble of local features and the AdaBoost classifier. Features are raw color channels in many color spaces and texture information captured by Schmid and Gabor filters [8]. Background clutter highly affects the descriptors of visual appearance for person recognition, and thus, the background modeling is used in many person re-id approaches [38,43,44].

The re-id has even been reinterpreted as a learning problem. In [45], the authors proposed a discriminative model based on the use of Partial Least Squares (PLS). In [46], a robust Mahalanobis metric for Large Margin Nearest Neighbor classification with Rejection (LMNN-R) was obtained with the use of a metric learning framework. Accordingly, in [47], the authors introduced a metric learning approach that learns a Mahalanobis distance from equivalence constraints derived from target labels. A comparison model aimed to maximize the probability of a pair of correctly matched images having a smaller distance than that of an incorrectly matched pair. The model was introduced as the Probabilistic Distance Comparison (PRDC) approach [48]. In [49], the same authors modeled person re-id as a transfer ranking problem, with the main goal of transferring similarity observations from a small gallery to a larger unlabeled probe set. Camera transfer approaches have also been introduced using images of the same person captured from different cameras to learn the associated metrics [50,51]. The Multiple Component Dissimilarity (MCD) framework was defined in [52] to turn a given appearance-based re-id method into a dissimilarity-based one. A supervised technique based on SVM is the approach presented in [53]. Pairs of similar and dissimilar images and a relaxed RankSVM algorithm [54] were used to rank probe images. The main issue with running RankSVM on large datasets is its very expensive computational load due to a large amount of inequality constraints. The authors in [29] used a decision tree to perform a fast matching between descriptors. In this case, the association of the query to one of the models is done by a voting approach. Dimensionality reduction was performed in [30] on image feature vectors through random projection. Afterwards, they built an ensemble of random forests, trained by feature vectors randomly projected onto different subspaces. Random forest was also employed in [31] to learn the similarity function of pairs of person images using color features.

The main differences with our work lay in:

- An RGB-D camera in a top view configuration motivated by the enhancement of the applicability of the proposed approach in crowded public environments is employed. The top-view configuration reduces the problem of occlusions and has the advantage of being privacy preserving because a person's face is not recorded by the camera [55]. However, this challenging configuration does not allow one to retrieve features related to the front view, which can be highly discriminative for the subject identification. Hence, the proposed approach including the feature extraction and the classification stage was designed according to this challenging setup

- The ensemble classifier was built taking into account the different nature of each feature. The model ensures a higher interpretability with respect to other black box models, allowing one to localize which features contribute to the final prediction.
- The computation time of the training stage is reasonably fast and would be practically feasible for real-world application.

2.2. TVPR Dataset and Related Applications

TVPR (Top View Person Re-identification) dataset (http://vrai.dii.univpm.it/re-id-dataset) for person re-id [23] contains videos of 100 individuals recorded over several days from an RGB-D camera installed in a top-view configuration. The camera was installed on the ceiling of a laboratory at 4 m above the floor and covered an area of 14.66 m^2 (4.43 m $\times$ 3.31 m). The camera was positioned above the surface where the analyses took place (Figure 1).

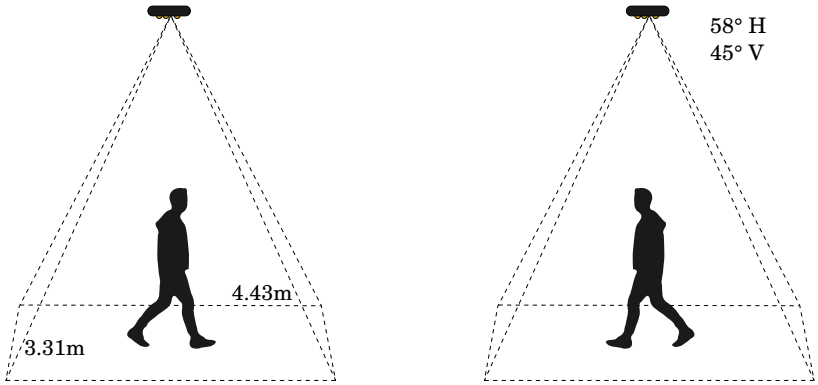


Figure 1. System architecture. (**a**) represents the first passage under the camera as training set, (**b**) is the the returning in the initial position considered as testing set.

The 100 people of our dataset were acquired in 23 registration sessions. Each of the 23 folders contains a video of one registration session. Acquisitions have been performed over eight days, and the total recording time was about 2000 s.

Registrations were made in an indoor scenario, where people passed under the camera installed on the ceiling. A big issue was environmental illumination. In the recording sessions, the illumination condition was not constant, but it varied as a function of the different hours of the day and also depended on natural illumination due to weather conditions. Snapshots of the video acquisitions, in our scenario, are depicted in Figure 2, where examples of person registration with artificial light are given.

Each person during a registration session walked with an average gait within the recording area in one direction and subsequently turned back and repeated over the same route in the opposite direction. This methodology is used for a better split of the TVPR in the training set (the first passage of the person under the camera) and the testing set (when the person passes a second time under the camera).

Although in the previous datasets presented in the literature, data were gathered using the RGB-D technology, they were not actually suitable for our purposes. The main motivating factors for our top-view dataset are due to some related applications that will be described below.

First, the top-view configuration provides the reliable and occlusion free counting of persons, which is crucial in many applications. Most of the previous works can only count moving people from a single camera, and they fail to count still people or situations when occlusions are very frequent and when there is a crowd. Possible applications can be: safety and security in crowded environments, people flow analysis and access control, as well as counting [56–58]. Actual tracking accuracy of top-view cameras overperforms all other tracking methods in crowded environments, with accuracies

up to 99%. When there are special security applications or the system is working in usually crowded scenarios, the proposed architecture with the top-view configuration is the only suitable one.

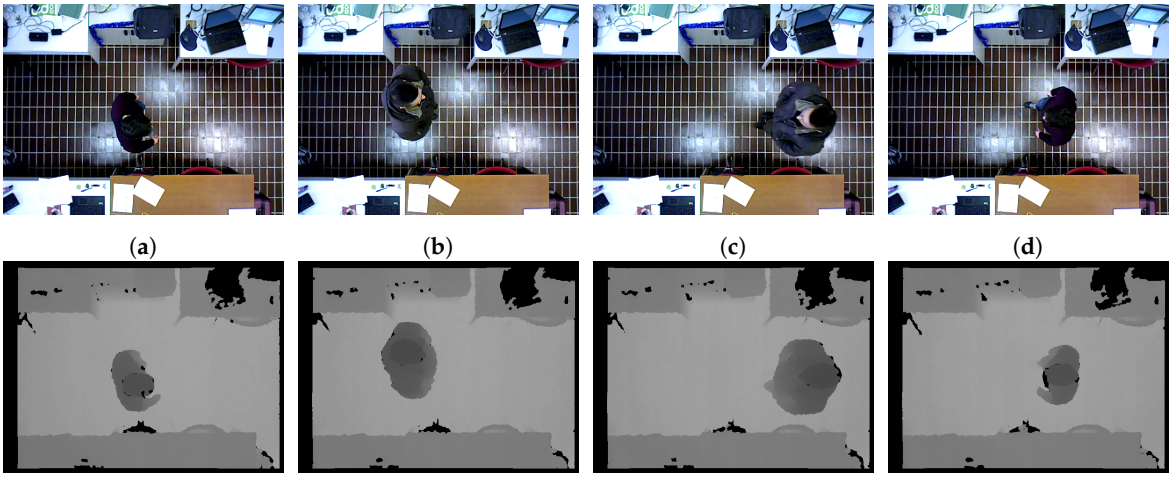


Figure 2. Snapshots of a registration session of the recorded data, in an indoor scenario, with artificial light. People passed under the camera installed on the ceiling. The sequence (**a**–**e**), (**b**–**f**) corresponds to the sequence (**d**–**h**), (**c**–**g**), respectively, training and testing set of the classes 8–9 for the registration session g003.

Second, the scope of this specific configuration and analysis is also the interaction detection between people and the environment with the many possible applications for the field of intelligent retail environment such as shopper analytics, in addition to the field of Human Behavior Analysis (HBA) for Ambient Assisted Living (AAL) [59–62].

Third, another possible application of this specific top-view configuration is fall detection and HBA in smart homes, from high-reliability fall detection to occlusion-free HBA at home for elders in AAL environments [55,63].

All these applications have relevant outcomes from the current research, with the ability to identify users or shoppers while performing tracking, interaction analysis or HBA. Furthermore, all these scenarios can gather data using low-cost sensors and processing units, ensuring scalability and mass usage. Finally, the proposed architecture can be certified on a EU basis privacy by design approach.

3. Methodology and Framework

Figure 3 shows the overview of the proposed approach comprised of data recording, feature extraction and the classification stage.

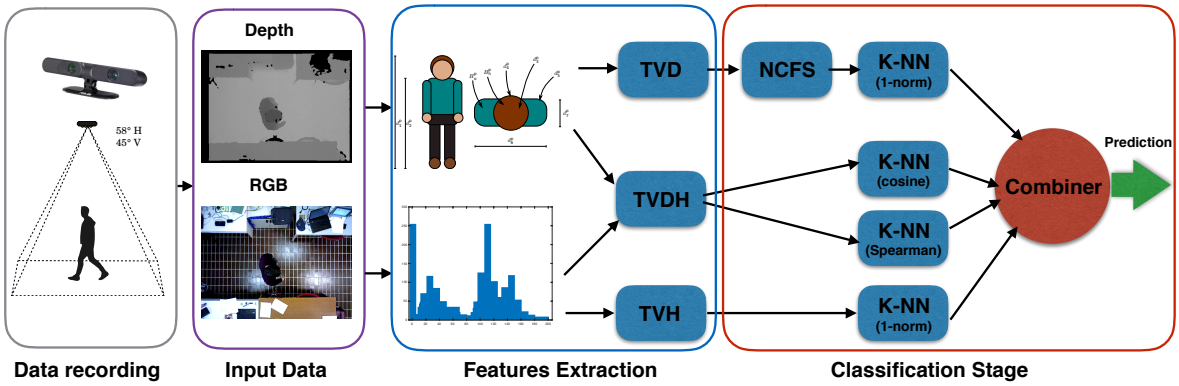


Figure 3. Overview of the proposed approach comprised of data recording, feature extraction and classification stage. NCFS, Neighborhood Component Feature Selection.

3.1. Pre-Processing and Feature Extraction

The first step involves the processing of the data acquired from the RGB-D camera. The camera captures the depth and color images, both with dimensions of 640 × 480 pixels, at a rate up to approximately 30 fps. The scene/objects are illuminated with structured light based on infrared patterns. People were detected from the top-view configuration using the same algorithm employed in [64].

Seven out of the nine features selected are anthropometric features extracted from the depth image: distance between floor and head, d_1; distance between floor and shoulders, d_2; area of head surface, d_3; head circumference, d_4; shoulder circumference, d_5; shoulder breadth, d_6; thoracic anteroposterior depth, d_7. The remaining two color-based features are acquired by the color image. We also define the color descriptor *TVH*:

$$TVH = \{H_h^p, H_o^p\} \tag{1}$$

and the depth descriptor *TVD*:

$$TVD = \{d_1^p, d_2^p, d_3^p, d_4^p, d_5^p, d_6^p, d_7^p\} \tag{2}$$

Finally, *TVDH* is the signature of a person defined as:

$$TVDH = \{d_1^p, d_2^p, d_3^p, d_4^p, d_5^p, d_6^p, d_7^p, H_h^p, H_o^p\} \tag{3}$$

Color is an important visual attribute for both computer vision and human perception. It is one of the most widely-used visual features in image/video retrieval. To extract these two features, we used HSV histograms. Local histograms have proven to be largely adopted and are very effective. The signature of a person is also composed by two color histograms computed for head/hair and outerwear: H_h^p, H_o^p in Equation (1), such as in [65], with $n = 10$ bin quantization, for both the H channel and S channel.

Figure 4 depicts the set of features considered: anthropometric and color-based.

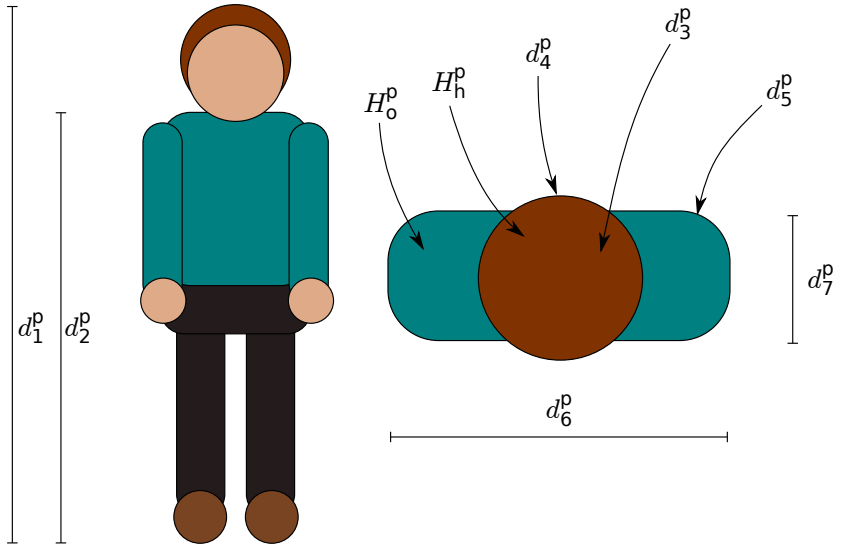


Figure 4. Anthropometric and color-based features.

3.2. Classification Stage

The classification stage is depicted in Figure 3. We propose an ensemble classification approach, named Multiple K-Nearest Neighbor (MKNN), where the primary classification stage is represented by different K-NN classifiers according to the nature of the feature descriptors. The overall prediction

is performed averaging the computed posterior probability of each K-NN classifier, in order to provide the optimal decision rule.

3.2.1. Predictive Model for TVD Descriptors

Since the TVD descriptors represent anthropometric features, we decided to adopt the 1-norm distance as a discriminative function of the K-NN model and the well-known Neighborhood Component Feature Selection (NCFS) approach [66] in order to learn the optimal feature weighting vector by maximizing the approximate regularized leave-one-out classification error. The application of NCFS allows decreasing the sensitivity of K-NN to irrelevant features [25]. In order to perform feature selection and decrease overfitting, we further introduce the regularization parameter λ, which controls the magnitude of the weighting vector. The optimal lambda found (i.e., $\lambda = 5 \times 10^{-4}$) was selected by previously implementing a grid-search and optimizing the macro-f1 score in the validation set. For further explanation about NCFS, the reader can refer to [66,67].

3.2.2. Predictive Model for TVH Descriptors

The cosine and the correlation metric are widely used in the literature to measure the similarity among different HSV descriptors [68,69]. Then, we implement two K-NN models with cosine and Spearman rank correlation, respectively, as the distance function.

The cosine distance between two HSV histogram features is defined as:

$$d_{cosine} = 1 - \frac{TVH_{test_i} \cdot TVH'_{train_j}}{\|TVH_{test_i}\| \left\|TVH_{train_j}\right\|} \tag{4}$$

while the Spearman rank correlation-based distance is defined as:

$$d_{spearman} = 1 - \frac{(rg\ TVH_{test_i} - rg\ \overline{TVH}_{test_i}) \cdot (rg\ TVH_{train_j} - rg\ \overline{TVH}_{train_j})'}{\left\|(rg\ TVH_{test_i} - rg\ \overline{TVH}_{test_i})\right\| \left\|(rg\ TVH_{train_j} - rg\ \overline{TVH}_{train_j})'\right\|} \tag{5}$$

where TVH_{test} and TVH_{train} are converted to ranks $rg\ TVH_{test}$ and $rg\ TVH_{train}$, while $\overline{TVH}$ is the sample mean.

3.2.3. Predictive Model for TVDH Descriptors

For the single K-NN model of the TVDH descriptors, we consider the 1-norm metric, to measure the distance between two different TVDH feature vectors.

3.3. Combiner

We introduce the approach for combining the prediction of the single K-NN model. Assuming $\{y_{p_1}, y_{p_2}, y_{p_3}, y_{p_4}\}$ are the predictions of the TVD, TVH and TVDH unseen sample, respectively (i.e., x_i), if we use the majority vote to determine the final label of y_{p_i}, the result will be:

$$\arg\max_{y \in 1 \ldots 100} \sum_{l=1}^{4} \delta(y, y_{p_l}) \tag{6}$$

where $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ otherwise. The Majority Vote (MV) approach does not take into account the posterior probability and does not always provide the best prediction results. The standard Bayesian approach [70]. finds the most probable hypothesis $\{y \in 1 \ldots 100\}$ given the observed data $\{y_{p_1}, y_{p_2}, y_{p_3}, y_{p_4}\}$:

$$\arg\max_{y} P(y | \{y_{p_1}, y_{p_2}, y_{p_3}, y_{p_4}\}) \tag{7}$$

according to Bayes' theorem, the maximally probable hypothesis becomes:

$$\arg\max_{y} P(\{y_{p_1}, y_{p_2}, y_{p_3}, y_{p_4}\}|y)P(y) \tag{8}$$

The Bayesian approach selects the model with the highest posterior probability and then proceeds as if the selected model had generated the data.

Differently from the Bayesian approach, we compute the average of the posterior probability (i.e., $P(\bar{y})$) of the 4 hypotheses as follows:

$$P(\bar{y}) = \sum_{l=1}^{4} P(y|y_{p_l}) = \sum_{l=1}^{4} P(y_{p_l}|y)P(y) \tag{9}$$

and the final prediction is:

$$y_p = \arg\max_{\bar{y} \in 1...100} P(\bar{y}) \tag{10}$$

Our ensemble methodology is based on Bayesian Model Averaging (BMA), which is an application of Bayesian inference to the problems of combined prediction of different classifiers. Although this choice can lead to overfitting in some situations [71], it provides straightforward model choice criteria and less risky predictions [72–74]. The BMA ignores the uncertainty in model selection, leading to over-confident inferences and decisions [73].

4. Results

The baseline results are reported in Section 4.1 in terms of the Cumulative Match Curve (CMC). In Sections 4.2 and 4.3, however, we show the results of the proposed MKNN approach for re-id classification. The authors compare the performance of the proposed methodology with respect to single K-NN classifiers and other supervised machine learning algorithms widely used in the re-id literature. We have also performed the computation time comparison related to the training stage.

4.1. Baseline Results

The baseline performance of the TVPR dataset was evaluated in terms of recognition rate, using the CMC curves, as previously described in [23]. Figure 5 depicts a comparison among the *TVH*, *TVD* and *TVDH* predictors in terms of CMC curves, to compare the ranks returned by using these different descriptors, where the horizontal axis is the rank of the matching score and the vertical axis is the probability of correct identification.

In particular, Figure 5a,b represents respectively the CMC obtained using the *TVH* and *TVD* descriptors for three different distances: one-norm (L1 city block), two-norm (euclidean) and cosine. Figure 5c provides the CMC computed using both *TVH* and *TVD* descriptors (i.e., *TVDH*), while Figure 5d is the averaged CMC over the three considered distances for the color (i.e., average of CMC curves in Figure 5a), depth (i.e., average of CMC curves in Figure 5b) and depth + color (i.e., average of CMC curves in Figure 5c). Although it can be assumed that the best performance was achieved when using the combination of descriptors (*TVDH*), the contribution of the depth was small, and the CMC curves in Figure 5a,c are very similar. However, the depth information can be informative for the re-id task (see Figure 5b). These baseline results suggest the need for a methodology to combine the different nature of descriptors, exploiting the importance and potential of the depth information. In this context, our approach aimed to exploit the informative power of depth and RGB input, properly combining the different nature of each feature.

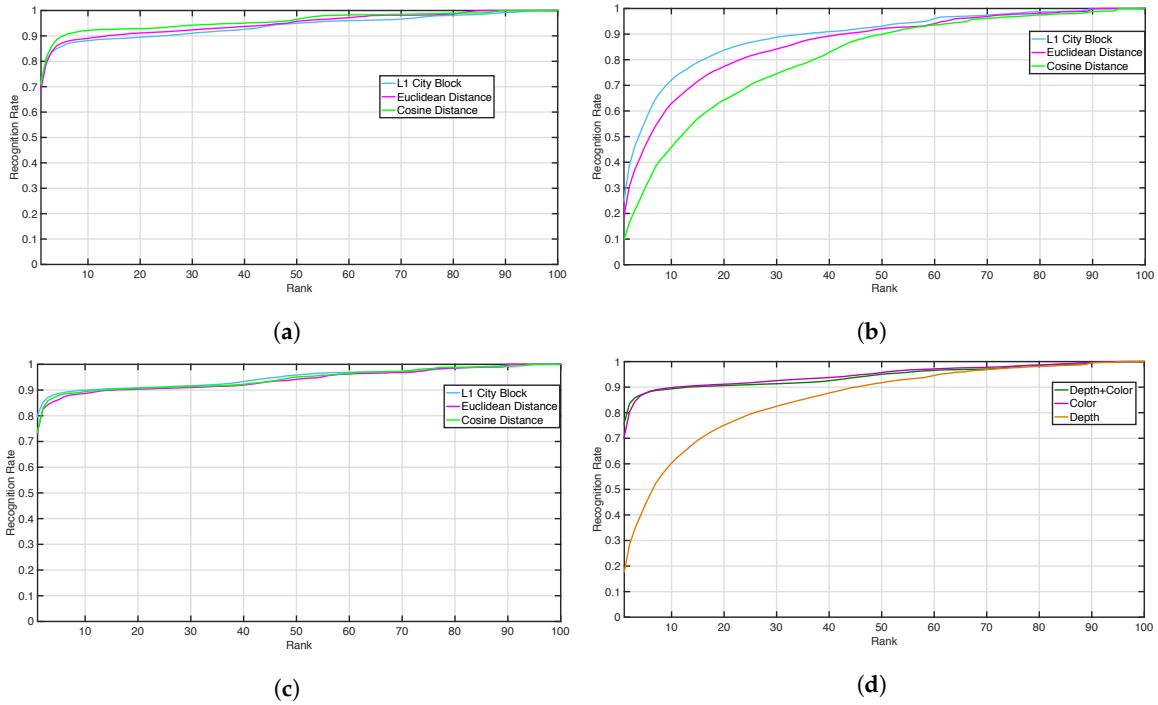


Figure 5. The baseline Cumulative Matching Curve (CMC) curves obtained on the Top View Person Re-Identification (TVPR) dataset. (**a**,**b**) shows respectively the CMC obtained using the *TVH* and *TVD* descriptors for three different distance: one-norm (L1 city block, cyan), two-norm (euclidean, purple) and cosine (green). (**c**) provides the CMC computed using both the *TVH* and *TVD* descriptors (i.e., *TVDH*), while (**d**) is the averaged CMC over the three considered distance for the color (i.e., average of CMC curves in (a), purple), depth (i.e., average of CMC curves in (b), orange) and depth + color (i.e., average of CMC curves in (c), green).

4.2. Results of the Proposed Approach

We considered the first passage under the camera as the training set and the return to the initial position as the testing set. The dataset was composed of 21,685 instances divided into 11,683 for training and 10,002 for testing. The performance of the proposed MKNN method is reported in Table 1 in terms of macro-F1 score, precision and recall. We also report the results of the single K-NN classifier for each descriptor (i.e., TVH, TVD, TVDH) and each different distance (i.e., cosine, Spearman's rank correlation and one-norm). We have highlighted in bold the single K-NN used for designing the proposed MKNN method. The optimal number of neighbors is five, and it has been chosen since it maximizes the macro-F1 score in the validation set. Additionally, we have reported the results of different combiner approaches (i.e., MV, Bayesian and BMA). The proposed BMA-MKNN approach performed favorably over the other methods.

According to the nature of the descriptors, the cosine distance was the most consistent measure in order to achieve the best performance for the TVH input, while the K-NN with one-norm achieved the best performance considering the TVDH input. The proposed MKNN methodology outperformed all single K-NN classifiers. In particular, the MKNN improved the performance of TVD-KNN, TVH-KNN and TVDH-KNN by 84.44%, 12% and 2.5%, respectively. Figure 6 shows the CMC curve of the MKNN compared with respect to the CMC curves of the single weak learner fed with TVH, TVD and TVDH. The ranking returned by MKNN showed better performance than the single classifier. This result outlines the advantage of the proposed approach in order to exploit the discriminative power of the depth information for the re-id task. In addition, the introduced BMA approach performed favorably over the MV and Bayesian methods.

Table 1. Classification results for single K-NN and Multiple K-Nearest Neighbor (MKNN) algorithms. BMA, Bayesian Model Averaging.

	Classifier	Distance	Precision	Recall	Macro-F1 Score
TVD	**KNN + NCFS**	1-norm	0.49	0.46	0.45
	KNN	1-norm	0.38	0.36	0.34
TVH	**KNN**	cosine	0.77	0.76	0.74
	KNN	Spearman	0.75	0.73	0.71
	KNN	1-norm	0.76	0.76	0.74
TVDH	**KNN**	1-norm	0.83	0.82	0.81
	KNN	2-norm	0.81	0.80	0.78
	MKNN (MV)		0.83	0.83	0.81
	MKNN (Bayesian)		0.81	0.80	0.78
	MKNN (BMA)		**0.86**	**0.85**	**0.83**

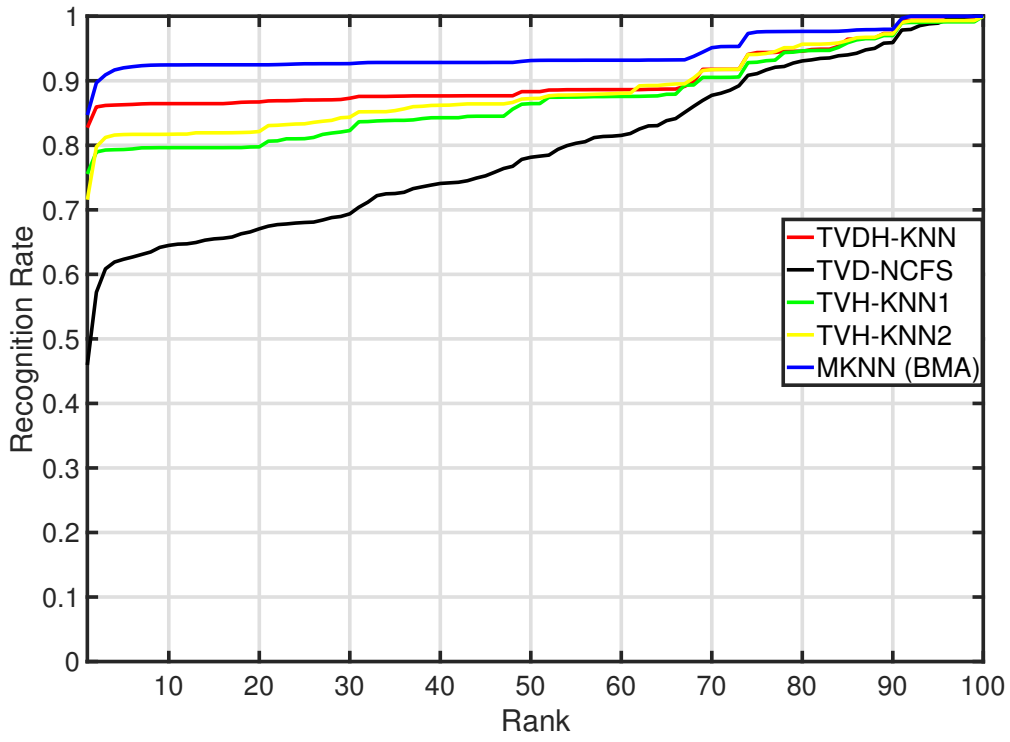
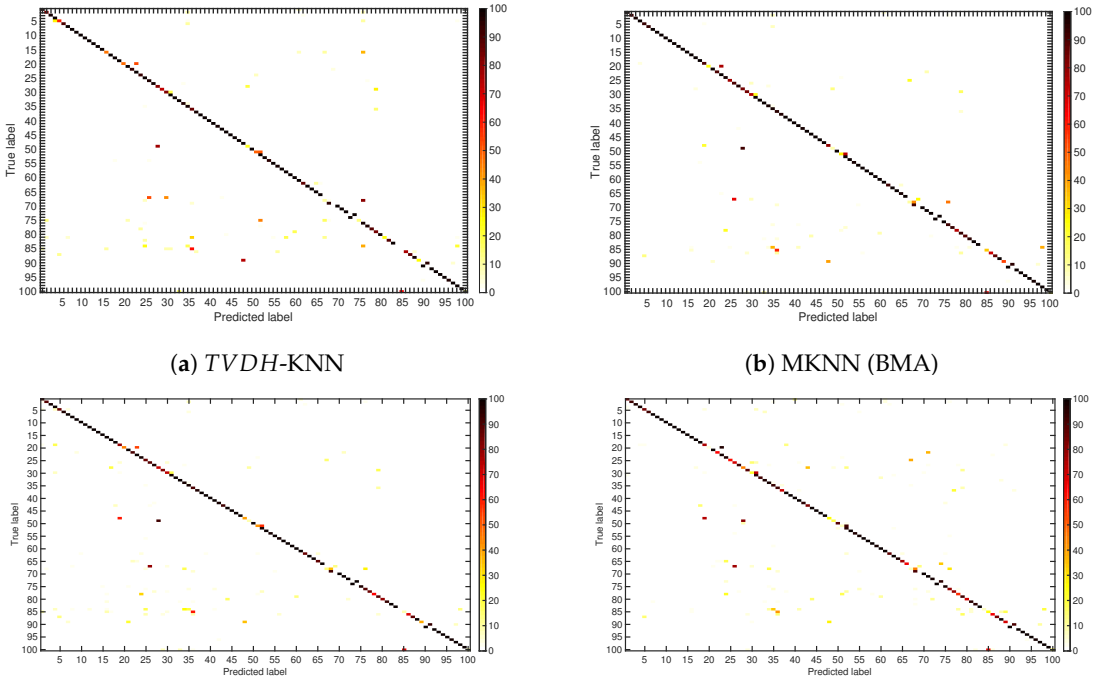


Figure 6. The CMC curves of the MKNN and the standard K-NN methods.

In order to highlight the misclassification error, we disclose in Figure 7 the confusion matrices of the TVDH-KNN, MKNN (BMA), MKNN (MV) and MKNN (Bayesian). The MKNN (BMA) shows a lower number of misclassified id-subject with respect to TVDH-HNN, MKNN (MV) and MKNN (Bayesian).

We summarize in Figure 8 the macro-f1 score for the MKNN and the TVDH-KNN for each class (subjects). The macro-f1 score is the same for 32 out of 100 subjects, while the MKNN achieves higher performance than TVDH-KNN in 42 out of 100 subjects. This result suggests how the MKNN (BMA) recognizes 10% of subjects with a higher recognition rate with respect to TVDH-KNN.



(**a**) $TVDH$-KNN (**b**) MKNN (BMA)

(**c**) MKNN (MV) (**d**) MKNN (Bayesian)

Figure 7. Confusion matrices of $TVDH$-KNN, MKNN (BMA), MKNN (MV) and MKNN (Bayesian).

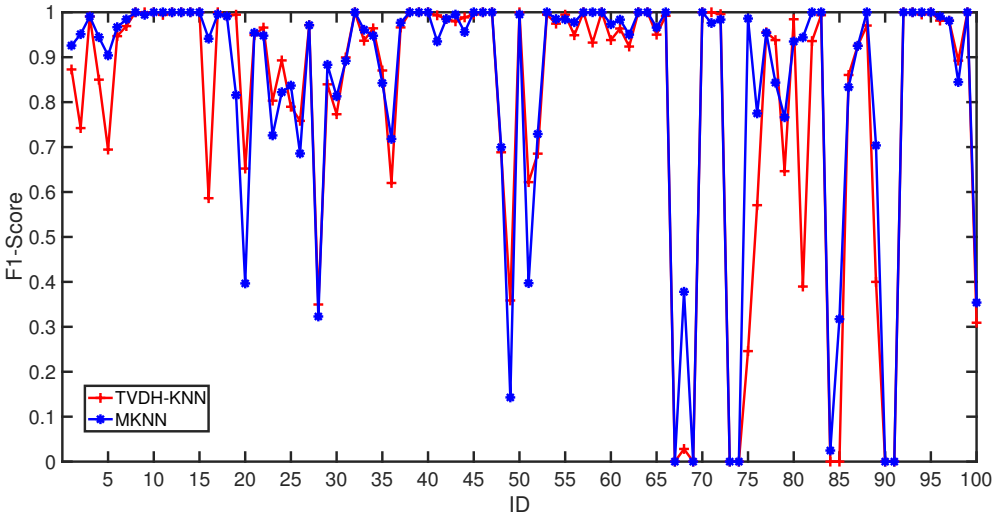


Figure 8. The macro-F1 for each subject for the MKNN and standard K-NN method.

The implemented NCFS for the TVD descriptors allowed decreasing the generalization error of the standard K-NN classifier while increasing the sparsity, as well as the interpretability of the model. Moreover, also the increase of K-NN performance in terms of precision, recall and macro-f1 score can be seen in Table 1. The optimal weighting vector found by the NCFS algorithm is shown in Figure 9. The feature with the highest predictive power is the thoracic anteroposterior depth (d_7), while the less relevant TVD descriptors are the distance between floor and shoulders (d_2), the area of the head surface (d_3) and the shoulder circumference (d_5).

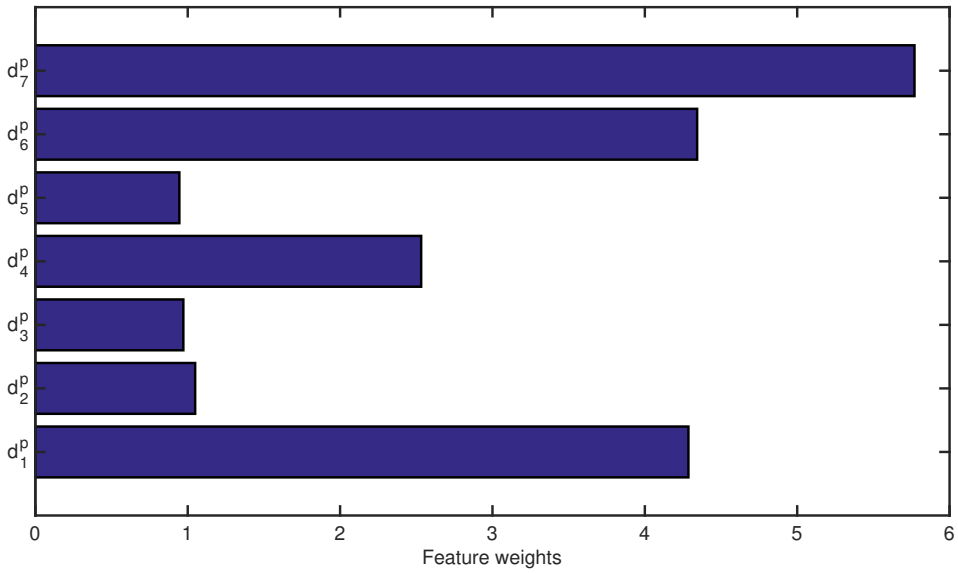


Figure 9. The optimal feature weights for TVD descriptors found by the NCFS algorithm.

4.3. Comparison with the Standard Supervised Machine Learning Algorithm

Table 2 shows the comparison between our approach and standard supervised learning algorithms widely adopted in the re-id scenario such as DT [29], bagged tree, RF [30,31], adaptive boosting (AdaBoost), linear programming boosting (LPBoost) and totally corrective boosting (TotalBoost). The considered inputs for the DT, bagged tree, RF, AdaBoost, LPBoost and TotalBoost classifiers are the TVDH descriptors.

Table 2. Comparison of MKNN with respect to the standard supervised learning approach. LPBoost, linear programming boosting.

Classifier	Input	Precision	Recall	F1-Score
KNN	TVDH	0.83	0.82	0.81
DT	TVDH	0.52	0.50	0.47
Bagged Tree	TVDH	0.83	0.81	0.80
RF	TVDH	0.74	0.72	0.70
AdaBoost	TVDH	0.65	0.60	0.58
LPBoost	TVDH	0.57	0.52	0.49
TotalBoost	TVDH	0.69	0.62	0.61
MKNN (BMA)		**0.86**	**0.85**	**0.83**

The MKNN outperformed all standard methods, achieving an improvement of 76.60%, 3.75%, 18.57%, 43.10%, 69.39% and 36.07% with respect to DT, bagged tree, RF, AdaBoost, LPBoost and TotalBoost. The K-NN may perform better than DT and RF when the number of training samples is not huge compared to the number of classes. The advantage of our ensemble strategy lies in the way we have built and combined each classifier. In particular, each weak learner was built according to the different nature of the features in order to extract the discriminative information of each subject. Differently from our approach, the other boosting and bagged strategies combined different weak learners in an automatic fashion without taking into account the different descriptors (i.e., TVH and TVD).

Table 3 shows the computation time expressed in seconds (s) for the training stage of all methodologies. MKNN (BMA) was reasonably fast and would be practically feasible for the re-id task.

Table 3. Computation time training stage.

Classifier	Training Time (s)
KNN	0.02
DT	1.31
Bagged Tree	12.14
RF	113.21
AdaBoost	31.14
LPBoost	375.94
TotalBoost	576.24
MKNN (BMA)	6.94

5. Conclusions and Future Works

In this paper, we describe a method for person re-identification based on features derived from both depth (anthropometric features) and color. Different from other approaches, the experiments were conducted on the TVPR dataset where the RGB-D images were collected in a top-view setting, reducing the problems of occlusions, while preserving the privacy issue [55].

Person recognition is handled by using the proposed ensemble method, named Multiple K-Nearest Neighbor (MKNN), based on the combination of different K-NN classifiers. Each K-NN is built with a different distance function based on the nature of the feature descriptors, and the neighborhood component feature selection is introduced for the anthropometric features. The experimental results demonstrate how the proposed methodology outperforms standard supervised classifiers (i.e., k-NN, DT, bagged tree, RF and boosting methods). Moreover, the computation time analysis of the training stage suggests that the proposed MKNN method is reasonably fast, encouraging the application of the proposed approach for the person re-identification task in the retail scenario. This improvement may be explained by the fact that our approach is consistent to model and combine the nature and information of different descriptors (i.e., TVH and TVD), weighting the importance of the anthropometric features. Further investigation will be devoted to improve our approach by extracting other informative features and setting up the proposed approach for the real-time processing of video images in the retail scenario. In the field of retail applications, the long-term goal of this work is to merge the developed re-identification system with an audio framework and the use of other types of RGB-D cameras, such as Time Of Flight (TOF) ones. The system can be integrated additionally as a source of high semantic level information in a networked ambient intelligence scenario, to provide cues for different problems, such as detecting abnormal speed and dimension outliers, alerting one to a possible uncontrolled circumstance. It would also be interesting to evaluate both color and depth images in a way that it does not decrease the performance of the system when the color image is being affected by changes in pose and/or illumination.

Author Contributions: Introduction, M.P., E.F. and P.Z.; Background, M.P. and A.C.; Methodology and Framework, M.P., L.R., R.P. and D.L.; Results, M.P., L.R., R.P. and D.L.; Conclusion and Future Works, E.F. and P.Z. All the authors contributed to the development and set up of the general idea.

Acknowledgments: This work was supported by FIT (Fondo speciale rotativo per l'Innovazione Tecnologica), Programme Title "Study, design and prototyping of an innovative artificial vision system for human behavior analysis in domestic and commercial environments" (HBA 2.0 (Human Behaviour Analysis)).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Vezzani, R.; Baltieri, D.; Cucchiara, R. People reidentification in surveillance and forensics: A survey. *ACM Comput. Surv. CSUR* **2013**, *46*, 29. [CrossRef]
2. Soleymani, R.; Granger, E.; Fumera, G. Progressive Boosting for Class Imbalance and Its Application to Face Re-Identification. *Expert Syst. Appl.* **2018**, *101*, 271–291. [CrossRef]

3. Panniello, U.; Hill, S.; Gorgoglione, M. Using context for online customer re-identification. *Expert Syst. Appl.* **2016**, *64*, 500–511. [CrossRef]
4. Chahla, C.; Snoussi, H.; Abdallah, F.; Dornaika, F. Discriminant quaternion local binary pattern embedding for person re-identification through prototype formation and color categorization. *Eng. Appl. Artif. Intell.* **2017**, *58*, 27–33. [CrossRef]
5. Hariri, W.; Tabia, H.; Farah, N.; Benouareth, A.; Declercq, D. 3D facial expression recognition using kernel methods on Riemannian manifold. *Eng. Appl. Artif. Intell.* **2017**, *64*, 25–32. [CrossRef]
6. Baltieri, D.; Vezzani, R.; Cucchiara, R. 3D Body Model Construction and Matching for Real Time People Re-Identification. In Proceedings of the Eurographics Italian Chapter Conference, Genova, Italy, 18–19 November 2010; pp. 65–71.
7. Farou, B.; Kouahla, M.N.; Seridi, H.; Akdag, H. Efficient local monitoring approach for the task of background subtraction. *Eng. Appl. Artif. Intell.* **2017**, *64*, 1–12. [CrossRef]
8. Lisanti, G.; Masi, I.; Bagdanov, A.D.; Del Bimbo, A. Person re-identification by iterative re-weighted sparse ranking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1629–1642. [CrossRef] [PubMed]
9. Paolanti, M.; Liciotti, D.; Pietrini, R.; Mancini, A.; Frontoni, E. Modelling and forecasting customer navigation in intelligent retail environments. *J. Intell. Robot. Syst.* **2018**, *91*, 165–180. [CrossRef]
10. Paolanti, M.; Sturari, M.; Mancini, A.; Zingaretti, P.; Frontoni, E. Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning. In Proceedings of the 2017 European Conference on IEEE Mobile Robots (ECMR), Paris, France, 6–8 September 2017; pp. 1–6.
11. Liciotti, D.; Paolanti, M.; Frontoni, E.; Zingaretti, P. People Detection and Tracking from an RGB-D Camera in Top-View Configuration: Review of Challenges and Applications. In Proceedings of the International Conference on Image Analysis and Processing, Catania, Italy, 11–15 September 2017; pp. 207–218.
12. Liciotti, D.; Paolanti, M.; Pietrini, R.; Frontoni, E.; Zingaretti, P. Convolutional Networks for semantic Heads Segmentation using Top-View Depth Data in Crowded Environment. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
13. Frontoni, E.; Mancini, A.; Zingaretti, P.; Placidi, V. Information management for intelligent retail environment: The shelf detector system. *Information* **2014**, *5*, 255–271. [CrossRef]
14. Pierdicca, R.; Liciotti, D.; Contigiani, M.; Frontoni, E.; Mancini, A.; Zingaretti, P. Low cost embedded system for increasing retail environment intelligence. In Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Torino, Italy, 29 June–3 July 2015; pp. 1–6.
15. Wu, A.; Zheng, W.; Lai, J. Robust Depth-Based Person Re-Identification. *IEEE Trans. Image Process.* **2017**, *26*, 2588–2603. [CrossRef] [PubMed]
16. Wang, Z.; Hu, R.; Liang, C.; Yu, Y.; Jiang, J.; Ye, M.; Chen, J.; Leng, Q. Zero-shot person re-identification via cross-view consistency. *IEEE Trans. Multimed.* **2016**, *18*, 260–272. [CrossRef]
17. Gong, S.; Cristani, M.; Yan, S.; Loy, C.C. *Person Re-Identification*; Springer: Berlin, Germany, 2014; Volume 1.
18. Gray, D.; Brennan, S.; Tao, H. Evaluating appearance models for recognition, reacquisition, and tracking. In Proceedings of the IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), Rio de Janeiro, Brazil, 14 October 2007.
19. Wang, T.; Gong, S.; Zhu, X.; Wang, S. Person re-identification by video ranking. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 688–703.
20. Ess, A.; Leibe, B.; Van Gool, L. Depth and appearance for mobile scene analysis. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio De Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
21. Cheng, D.S.; Cristani, M.; Stoppa, M.; Bazzani, L.; Murino, V. Custom Pictorial Structures for Re-identification. In Proceedings of the BMVC 2011, Dundee, Scotland, 30 August–1 September 2011; p. 6.
22. Barbosa, I.B.; Cristani, M.; Del Bue, A.; Bazzani, L.; Murino, V. Re-identification with rgb-d sensors. In Proceedings of the Computer Vision—ECCV 2012. Workshops and Demonstrations, Florence, Italy, 7–13 October 2012; pp. 433–442.
23. Liciotti, D.; Paolanti, M.; Frontoni, E.; Mancini, A.; Zingaretti, P. Person Re-Identification Dataset with RGB-D Camera in a Top-View Configuration. In *Video Analytics for Face, Face Expression Recognition, and Audience Measurement*; Springer: Berlin, Germany, 2017.
24. Sturari, M.; Liciotti, D.; Pierdicca, R.; Frontoni, E.; Mancini, A.; Contigiani, M.; Zingaretti, P. Robust and affordable retail customer profiling by vision and radio beacon sensor fusion. *Pattern Recognit. Lett.* **2016**, *81*, 30–40. [CrossRef]

25. Bay, S.D. Nearest neighbor classification from multiple feature subsets. *Intell. Data Anal.* **1999**, *3*, 191–209. [CrossRef]
26. Siahroudi, S.K.; Moodi, P.Z.; Beigy, H. Detection of evolving concepts in non-stationary data streams: A multiple kernel learning approach. *Expert Syst. Appl.* **2018**, *91*, 187–197. [CrossRef]
27. Bao, Y.; Ishii, N.; Du, X. Combining Multiple k-Nearest Neighbor Classifiers Using Different Distance Functions. In *Intelligent Data Engineering and Automated Learning—IDEAL 2004*; Yang, Z.R., Yin, H., Everson, R.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 634–641.
28. Zheng, L.; Wang, S.; Tian, L.; He, F.; Liu, Z.; Tian, Q. Query-Adaptive Late Fusion for Image Search and Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
29. Hamdoun, O.; Moutarde, F.; Stanciulescu, B.; Steux, B. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In Proceedings of the ICDSC 2008 Second ACM/IEEE International Conference on Distributed Smart Cameras, Palo Alto, CA, USA, 7–11 September 2008; pp. 1–6.
30. Li, Y.; Wu, Z.; Radke, R.J. Multi-shot re-identification with random-projection-based random forests. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), Big Island, HI, USA, 6–9 January 2015; pp. 373–380.
31. Du, Y.; Ai, H.; Lao, S. Evaluation of color spaces for person re-identification. In Proceedings of the 2012 21st International Conference on Pattern Recognition (ICPR), Tsukuba Science City, Japan, 11–15 November 2012; pp. 1371–1374.
32. Bedagkar-Gala, A.; Shah, S.K. A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **2014**, *32*, 270–286. [CrossRef]
33. Messelodi, S.; Modena, C.M. Boosting Fisher vector based scoring functions for person re-identification. *Image Vis. Comput.* **2015**, *44*, 44–58. [CrossRef]
34. Havasi, L.; Szlávik, Z.; Szirányi, T. Eigenwalks: Walk detection and biometrics from symmetry patterns. In Proceedings of the IEEE International Conference on Image Processing 2005, Genoa, Italy, 11–14 September 2005; pp. III–289.
35. Fischer, M.; Ekenel, H.K.; Stiefelhagen, R. Interactive person re-identification in TV series. In Proceedings of the 2010 International Workshop on Content-Based Multimedia Indexing (CBMI), Grenoble, France, 23–25 June 2010; pp. 1–6.
36. Calderara, S.; Prati, A.; Cucchiara, R. Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance. *Comput. Vis. Image Understand.* **2008**, *111*, 21–42. [CrossRef]
37. Javed, O.; Shafique, K.; Rasheed, Z.; Shah, M. Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views. *Comput. Vis. Image Understand.* **2008**, *109*, 146–162. [CrossRef]
38. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 2360–2367.
39. Alahi, A.; Vandergheynst, P.; Bierlaire, M.; Kunt, M. Cascade of descriptors to detect and track objects across any network of cameras. *Comput. Vis. Image Understand.* **2010**, *114*, 624–640. [CrossRef]
40. Gandhi, T.; Trivedi, M.M. Panoramic appearance map (pam) for multi-camera based person re-identification. In Proceedings of the 2006 IEEE International Conference on Video and Signal Based Surveillance, Sydney, Australia, 22–24 November 2006; pp. 78–78.
41. Gheissari, N.; Sebastian, T.B.; Hartley, R. Person reidentification using spatiotemporal appearance. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1528–1535.
42. Gray, D.; Tao, H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Computer Vision–ECCV 2008*; Springer: Berlin, Germany, 2008; pp. 262–275.
43. Bazzani, L.; Cristani, M.; Perina, A.; Farenzena, M.; Murino, V. Multiple-shot person re-identification by hpe signature. In Proceedings of the 2010 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 23–26 August 2010; pp. 1413–1416.
44. Bazzani, L.; Cristani, M.; Perina, A.; Murino, V. Multiple-shot person re-identification by chromatic and epitomic analyses. *Pattern Recognit. Lett.* **2012**, *33*, 898–903. [CrossRef]

45. Schwartz, W.R.; Davis, L.S. Learning discriminative appearance-based models using partial least squares. In Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing, Rio de Janeiro, Brazil, 11–15 October 2009; pp. 322–329.
46. Dikmen, M.; Akbas, E.; Huang, T.S.; Ahuja, N. Pedestrian recognition with a learned metric. In Proceedings of the Asian Conference on Computer Vision, Queenstown, New Zealand, 8–12 November 2010; pp. 501–512.
47. Köstinger, M.; Hirzer, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Large scale metric learning from equivalence constraints. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2288–2295.
48. Zheng, W.S.; Gong, S.; Xiang, T. Reidentification by relative distance comparison. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 653–668. [CrossRef] [PubMed]
49. Zheng, W.S.; Gong, S.; Xiang, T. Person re-identification by probabilistic relative distance comparison. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 649–656.
50. Avraham, T.; Gurvich, I.; Lindenbaum, M.; Markovitch, S. Learning implicit transfer for person re-identification. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 381–390.
51. Hirzer, M.; Roth, P.M.; Köstinger, M.; Bischof, H. Relaxed pairwise learned metric for person re-identification. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 780–793.
52. Satta, R.; Fumera, G.; Roli, F. Fast person re-identification based on dissimilarity representations. *Pattern Recognit. Lett.* **2012**, *33*, 1838–1848. [CrossRef]
53. Prosser, B.; Zheng, W.S.; Gong, S.; Xiang, T.; Mary, Q. Person Re-Identification by Support Vector Ranking. In Proceedings of the BMVC 2010, Aberystwyth, UK, 30 August–2 September 2010; Volume 2, p. 6.
54. Chapelle, O.; Keerthi, S.S. Efficient Algorithms for Ranking with SVMs. *Inf. Retr.* **2010**, *13*, 201–215. [CrossRef]
55. Liciotti, D.; Massi, G.; Frontoni, E.; Mancini, A.; Zingaretti, P. Human activity analysis for in-home fall risk assessment. In Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015; pp. 284–289.
56. Zhang, X.; Yan, J.; Feng, S.; Lei, Z.; Yi, D.; Li, S.Z. Water filling: Unsupervised people counting via vertical kinect sensor. In Proceedings of the 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS), Beijing, China, 18–21 September 2012; pp. 215–220.
57. Wateosot, C.; Suvonvorn, N. Top-view Based People Counting Using Mixture of Depth and Color Information. In Proceedings of the ACIS 2013: The Second Asian Conference on Information Systems, Phuket, Thailand, 31 October–2 November 2013.
58. Nalepa, J.; Szymanek, J.; Kawulok, M. Real-time people counting from depth images. In *Beyond Databases, Architectures and Structures*; Springer: Berlin, Germany, 2015; pp. 387–397.
59. Liciotti, D.; Contigiani, M.; Frontoni, E.; Mancini, A.; Zingaretti, P.; Placidi, V. Shopper Analytics: A Customer Activity Recognition System Using a Distributed RGB-D Camera Network. In *Video Analytics for Audience Measurement*; Springer: Berlin, Germany, 2014; pp. 146–157.
60. Mancini, A.; Frontoni, E.; Zingaretti, P.; Placidi, V. Smart vision system for shelf analysis in intelligent retail environments. In Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Portland, OR, USA, 4–7 August 2013.
61. Migniot, C.; Ababsa, F. 3D human tracking from depth cue in a buying behavior analysis context. In *Computer Analysis of Images and Patterns*; Springer: Berlin, Germany, 2013; pp. 482–489.
62. Marquardt, N.; Hinckley, K.; Greenberg, S. Cross-device interaction via micro-mobility and f-formations. In Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology, Cambridge, MA, USA, 7–10 October 2012; pp. 13–22.
63. Kepski, M.; Kwolek, B. Fall detection using ceiling-mounted 3d depth camera. In Proceedings of the 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, 5–8 January 2014; Volume 2, pp. 640–647.

64. Liciotti, D.; Frontoni, E.; Mancini, A.; Zingaretti, P. Pervasive System for Consumer Behaviour Analysis in Retail Environments. In *Video Analytics. Face and Facial Expression Recognition and Audience Measurement*; Nasrollahi, K., Distante, C., Hua, G., Cavallaro, A., Moeslund, T.B., Battiato, S., Ji, Q., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 12–23.
65. Baltieri, D.; Vezzani, R.; Cucchiara, R. Learning articulated body models for people re-identification. In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; pp. 557–560.
66. Yang, W.; Wang, K.; Zuo, W. Neighborhood Component Feature Selection for High-Dimensional Data. *J. Comput.* **2012**, *7*, 161–168. [CrossRef]
67. Goldberger, J.; Hinton, G.E.; Roweis, S.T.; Salakhutdinov, R.R. Neighbourhood components analysis. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2005; pp. 513–520.
68. Mistry, Y.; Ingole, D.; Ingole, M. Content based image retrieval using hybrid features and various distance metric. *J. Electr. Syst. Inf. Technol.* **2017**. [CrossRef]
69. Stricker, M.A.; Orengo, M.Q. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III*; International Society for Optics and Photonics: Bellingham, WA, USA, 1995; Volume 2420, pp. 381–393.
70. Wang, J.; Jean-Daniel, Z. Solving the Multiple-Instance Problem: A Lazy Learning Approach. In Proceedings of the 17th International Conference on Machine Learning, Stanford, CA, USA, 29 June–2 July 2000; pp. 1119–1125.
71. Domingos, P.M. Bayesian Averaging of Classifiers and the Overfitting Problem. In Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford, CA, USA, 29 June–2 July 2000; pp. 223–230.
72. Wasserman, L. Bayesian Model Selection and Model Averaging. *J. Math. Psychol.* **2000**, *44*, 92–107. [CrossRef] [PubMed]
73. Hoeting, J.A.; Madigan, D.; Raftery, A.E.; Volinsky, C.T. Bayesian model averaging: A tutorial. *Stat. Sci.* **1999**, *14*, 382–401.
74. Fragoso, T.M.; Bertoli, W.; Louzada, F. Bayesian model averaging: A systematic review and conceptual classification. *Int. Stat. Rev.* **2018**, *86*, 1–28. [CrossRef]