An agent-based WCET analysis for Top-View Person Re-Identification

Marina Paolanti, Valerio Placidi,
Michele Bernardini, Andrea Felicetti, Rocco Pietrini, and
Emanuele Frontoni

Department of Information Engineering, Università Politecnica delle Marche,
Via Brecce Bianche 12, 60131, Ancona, Italy

Abstract. Person re-identification is a challenging task for improving and personalising the shopping experience in an intelligent retail environment. A new Top View Person Re-Identification (TVPR) dataset of 100 persons has been collected and described in a previous work. This work estimates the Worst Case Execution Time (WCET) for the features extraction and classification steps. Such tasks should not exceed the WCET, in order to ensure the effectiveness of the proposed application. In fact, after the features extraction, the classification process is performed by selecting the first passage under the camera for training and using the others as the testing set. Furthermore, a gender classification is exploited for improving retail applications. We tested all feature sets using k-Nearest Neighbors, Support Vector Machine, Decision Tree and Random Forest classifiers. Experimental results prove the effectiveness of the proposed approach, achieving good performance in terms of Precision, Recall and F1-score.

Keywords: Real-time; WCET; Person re-identification; RGB-D camera; Retail.

1 Introduction

Nowadays, camera are largely deployed in several sectors ranging from small business and large retail applications to home surveillance, environment monitoring and facility access applications. Identification cameras are widely employed in most public areas as shopping centers, airports, stations, office buildings and museums. In these situations, it is advisable to determine whether different instances or images of one person, captured at different times, belong to the same subject. Commonly, "person re-identification" (re-id) defines this kind of process. Re-id owns a great commercial value because of its wide range of potential applications and benefits.

During last years, research oriented to people behaviour analysis has been totally centered around person re-id, which is seen as the exploitation of many paradigms and approaches of pattern recognition [1]. In such conditions, algorithms need to be robust to address issues such as widely varying camera viewpoints and orientations, rapid changes in the appearance of clothing, occlusions, varied poses and different lighting conditions [2], [3].

Person re-id means modelling 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 [4], [3]. Labelled training data are required for metric learning approaches and new training data are needed whenever a camera setting changes [5].

Recently, person re-id is emerging as a very challenging task for improving and personalising the shopping experience in the intelligent retail environment. It is becoming a useful tool to properly recognise consumers in a store, to study returning consumers and to classify different shopper clusters and targets. Re-id can provide useful information for customer services and shopping space management. In fact, the increased development and change in consumer purchase behaviour have led the retailers to adapt their businesses, the products and services they provide, but also the way in which they communicate to the customers [6].

The use of RGB-D cameras can be strictly linked to this purpose, because it provides affordable and additional rough depth information coupled with visual images, offering sufficient accuracy and resolution for indoor applications. In the retail, this camera has already been successfully adopted with the aim to univocally identify customers and analyse their interactions with shoppers [7]. The usual choice is RGB-D camera placed in a top view configuration because of its greater suitability compared with a front view configuration, mostly adopted for gesture recognition or even for video gaming. The problem of occlusions is reduced by the choice of a top-view configuration, advantageously being privacy preserving since person's face cannot be recorded by the camera [8].

In a previous work, 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 [9]. We have chosen an Asus Xtion Pro Live RGB-D camera because it allows the acquisition of colour and depth information in an affordable and fast way [10]. The camera was installed on the ceiling above the area to be analysed. This dataset collects the data of 100 people, acquired across intervals of days and in different times. The camera has been located on the ceiling above the area of interest.

In this paper, the method applied within a real-time scenario is proposed. A software agent is supposed to recognize a subject when she/he passes under a camera more than once, in order to provide, at the same time, an instant and customized service for the single consumer. In the retail sector, the capacity to identify the consumer characteristics assumes a high relevance in order to offer personalized promotions, focused on the type of person (i.e., gender, age), the history of his preferences and shopping habits (i.e., fidelity card). In a supermarket where a varied offer is proposed, the goal is to identify the returning consumer through an RGB-D camera placed at the entrance. After that, suggestions and offers tailored to each consumer will be displayed on advertising screens located immediately after the entrance and notifications will be instantaneously sent on their smartphones. Within this context, a worst-case execution time (WCET) analysis for top-view person re-identification has been developed. The correctness of real time systems does not only depend on the accuracy of the results, but also on the delivery of the results within established time constraints [11]. To ensure that all deadlines are reached, real-time schedulers need to estimate the WCET of each process. Classification results should be correct not only in their accuracy but also in the time domain predefined by the user. A real-time task is characterized by a deadline, which is the maximum time within which it must complete its execution [12]. Depending on the consequences that may occur because of a missed deadline, a real-time task can be distinguished as hard, firm and soft category. A real-time task which belongs to the soft category is producing the results after its deadline, but still has some utility for the system, although causing a performance degradation. Soft tasks are typically related to system-user interactions. Such tasks as displaying ads on the screen or sending alerts are enclosed in this category. in addition an agent-based system that monitors the whole real time re-id procedure can manage several features such as:

- shopping chronology of each consumer connected with the personal fidelity card,
- selection of customized information to be shared to each consumer,
- entire messaging process for sending personal offers to advertisement screens or alerts on smartphones.

In any real-time control system, the algorithm of each task is known a priori and thus can be utilised to estimate its characteristics in terms of computational time [13]. Above all, it allows to estimate the WCET parameter, used by the operating system to know its schedulability within the specified timing deadlines. The various agent activities can be seen as parts of a team cooperating. In a real-time approach, a WCET analysis guarantees an efficient, instantaneous and prompt customer service.

Moreover, we introduce a method for person re-id based on a set of features extracted by RGB-D images, used to perform a classification process: the first passage under the camera is selected as training set, while returns to the initial position as the testing set. In addition, a gender classification focused on colour and length of the hair, is performed with the aim to improve retail applications on shopper clustering on different targets. In fact, recognising a customer is a crucial information for retailers who need to know who their potential customers are in order to adapt the market to them more effectively. We tested all feature sets using k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF) classifiers, as previously done in [14], [15], [16]. The performance evaluation demonstrates the effectiveness of the proposed approach, achieving good results in term of Precision, Recall and F1-score.

This paper is organized as follows: Section 2 provides a description of the approaches in the context of re-id (Subsection 2.1), a framework of the existing datasets (Subsection 2.2) and the characterization of the TVPR dataset. Section 3 gives details on the proposed methodology. It is followed by the process of evaluation of our dataset with some samples and key statistics of the dataset and the presentation of results (Section 4). The conclusions and future work in this direction are elaborated in Section 5.

2 Background

This section is an overview of the principal approaches for person re-id. In particular, Subsection 2.1 presents a review/summary of the works on person re-id, Subsection 2.2 describes the available datasets that have been used to test re-id models and Subsection 2.3 provides details on TVPR dataset for person re-id in a top-view configuration.

2.1 Previous works on person re-identification

In the field of pattern recognition, the re-id problem has gained considerable attention and several reviews and surveys are available, pointing out different aspects of this topic [17]. Four different strategies could be defined, depending on the camera setup and environmental conditions: biometric, geometric, appearance based and learning approaches.

In the biometric approaches, the person instances are matched together and are assigned to the same identity by the use of biometric features. The examples employed in a real situation are faces, gait, iris scans, fingerprints and so on [18], [19]. They are effective and reliable solutions, but these require a collaborative behaviour of the persons and suitable sensors. Thus, in the case of low resolution, poor views, such as the case with common settings for surveillance cameras, these techniques are not always applicable.

The geometric approaches consider the situations when more than one sensors or cameras collect simultaneously information of the same area, and geometric relations among the fields of view (epipolar lines, homographies and so on) and can be adopted to match the different detection data [20], [21], [22]. 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 [23], [24]. In these situations, the appearance based approaches are used. Re-id can be correctly done only if the appearance is preserved among the views. Exploiting dress colours and textures, perceived heights and other similar cues, is considered to be a soft-biometric approach. Occlusions, different sensor qualities, illumination changes, different viewpoints are some of the issues which make the appearance based re-id a difficult problem. Gray et al. for the first time considered the problem of appearance models for person recognition, reacquisition and tracking in [22], . They also claimed that these problems had been evaluated independently and there is a need for metrics that apply to complete systems [25], [26]. A standard protocol to compare results is described. It used the Cumulative Matching Curve (CMC) and presented the VIPeR dataset for re-id. In [27], an algorithm that learns a domain-specific similarity function using an ensemble of local features and the AdaBoost classifier is described. In [5], features are raw colour channels in many colour spaces and texture information captured by Schmid and Gabor filters. In fact, for person recognition background clutter highly affects descriptors of visual appearance. Otherwise, the background modelling is used in many person re-id approaches [23], [28], [29].

The re-id has even been considered as a learning problem. In [30], the authors have proposed a discriminative model. It is obtained with the use of Partial Least Squares (PLS). A robust Mahalanobis metric for Large Margin Nearest Neighbor classification with Rejection (LMNN-R) is created with the use of a metric learning framework in [31]. In [32], the approach proposed by the authors is a supervised technique and pairs of similar and dissimilar images and a relaxed RankSVM algorithm is used to rank probe images. The work described in [33] is another metric learning approach which learns a Mahalanobis distance from equivalence constraints derived from target labels.

In [34] is introduced a comparison model by the Probabilistic Distance Comparison (PRDC) approach. It aims at maximising the probability of a pair of correctly matched images having a smaller distance than that of an incorrectly matched pair. In [35], the same authors model person re-id as a transfer ranking problem. The main goal of this paper is to transfer similarity observations from a small gallery to a larger unlabelled probe set. Camera transfer approaches have also been described and these use images of the same person captured from different cameras to learn the associated metrics [36], [37]. The Multiple Component Dissimilarity (MCD) framework that allows one to turn a given appearance-based re-id method into a dissimilarity-based one is described in [38] .

2.2 Public available datasets

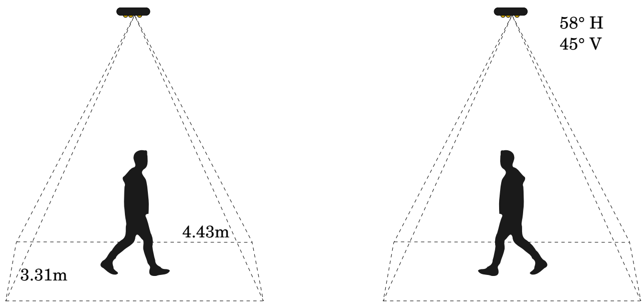
Different public datasets used to test re-id models are available. Currently, *VIPeR*[1], *iLIDS*,[2] *ETHZ*[3], *CAVIAR4REID*[4] are the most commonly used for re-id evaluations. Many aspects of the person re-id problem are covered by these datasets, such as occlusions, shape deformation, very low resolution images, illumination changes, image blurring, etc. [39]. The *ViPER* dataset [22] consists of images of people from two different camera views and it has only one image of each person per camera. The dataset has been collected for testing viewpoint invariant pedestrian recognition with 632 pedestrian images, normalized to 48×128 pixels, pairs taken from arbitrary viewpoints under varying illumination conditions. *iLIDS* was acquired in crowded public spaces [39] and it is used for tracking evaluation. This dataset collects 479 images of 119 people acquired from non-overlapping cameras. In [40] a modified version of the dataset of 69 individuals, is introduced, $iLIDS_{\geq 4}$, because *iLIDS* does not fit well in a multi-shot scenario. The average number of images per person is 4 and some individuals have only two images. In $iLIDS_{\geq 4}$ a subset of individuals with at least four images has been selected. The *ETHZ* dataset has images of people taken by a moving camera [41] and it contains three sequences and multiple images of a person from each sequence. It collects three sub-datasets: *ETHZ1* of 83 people and 4857 images, *ETHZ2* composed by 35 people and 1936 images, and *ETHZ3* of 28 and 1762 images. In [42], it has been introduced *CAVIAR4REID*, which is extracted from another multi-camera tracking dataset captured at an indoor shopping mall with two cameras with overlapping views in Lisbon. The dataset described in [42] contains multiple images of pedestrians. The images for each pedestrian were selected for maximizing appearance variations due to resolution changes, occlusions, light conditions, and pose changes. 72 individuals are identified (with images varying from 17×39 to 72×144) and 50 are captured by both views and 22 by just one camera. In [43], it is introduced another re-id dataset, which is composed by 79 people and 4 groups.

2.3 TVPR Dataset

The proposed system has been experimentally validated on TVPR (Top View Person Re-identification) dataset[5] for person re-id [9].

TVPR collects videos of 100 individuals recorded in several days from an RGB-D camera installed in a top-view configuration. The camera is positioned

[1] `https://vision.soe.ucsc.edu`
[2] `http://www.eecs.qmul.ac.uk`
[3] `https://data.vision.ee.ethz.ch/cvl/aess/dataset`
[4] `http://www.lorisbazzani.info/datasets`
[5] `http://vrai.dii.univpm.it/re-id-dataset`



(a) (b)

Fig. 1: System architecture.

on the ceiling of a laboratory at $4\,m$ above the floor and covers an area of $14.66\,m^2$ ($4.43\,m \times 3.31\,m$). The camera is above the surface which is to be analysed (Figure 1).

The 100 people of our dataset were acquired in 23 registration sessions. Each of the 23 folders has a video of one registration sessions. Acquisitions have been recorded in 8 days and the total registration time is about 2000 seconds.

Registrations are performed in an indoor scenario, where people pass under the camera. A big issue is environmental illumination. In each recording session, the illumination condition is not constant, because it varies in function of the different hours of the day and it also depends on natural illumination due to weather conditions.

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

3 Methodology and Framework

In this paper, the main goal is to ensure processing while maintaining the maximum frame rate of the camera. The camera captures depth and colour images, both with dimensions of 640×480 pixels, at a rate up to approximately 30 fps and illuminates the scene/objects with structured light based on infrared patterns. In particular, in order to carry out the assigned task in the real-time it is necessary to keep the entire processing time below 33 ms, which is the time that occurs between two consecutive frames. For estimating the computational time, TVPR video of four persons passing under the camera has been taken into account. The time that the program takes to extract the features is estimated by using the functions of the C++ “chrono” library.

The second step involves the processing of the data acquired from the RGB-D camera. 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; shoulders circumference, d_5; shoulders breadth, d_6; thoracic anteroposterior depth, d_7. The remaining two *colour-based features* are acquired by the colour image. In [9], we have also defined TVH the colour descriptor, TVD the depth descriptor and $TVDH$ the signature of a person.

For our experiments, we perform person re-id classification selecting the first passage under the camera for training and using a reset to the initial position as the testing set. We tested all feature sets using k-Nearest Neighbors (kNN) classifier [44], Support Vector Machine [45], [46], [47], Decision Tree [48] and Random Forest [49] and we evaluate performance in terms of precision, recall and F1-score.

Finally, a gender classification, based on colour and hair length, is carried out with the aim to improve retail applications. This aspect could be particularly useful in retail where new customers are certainly important, but returning customers should have greater weight. Recognising a customers gender is a crucial information for retailers who need to know who their potential customers are in order to adapt the market to them more effectively.

4 Results and discussion

The tests are performed on a notebook PC equipped with a processor Intel (R) Core (TM) i7-4510U CPU @ 2.00 GHz and 12 GB of RAM with Ubuntu 14.04 operating system. Figure 2a shows eight peaks corresponding to the time interval in which the person passes under the camera. During this time interval the features are extracted and the time spent for features extraction is estimated around 15 ms for frame. Spurious spikes are due to operating system processes running on the same machine.

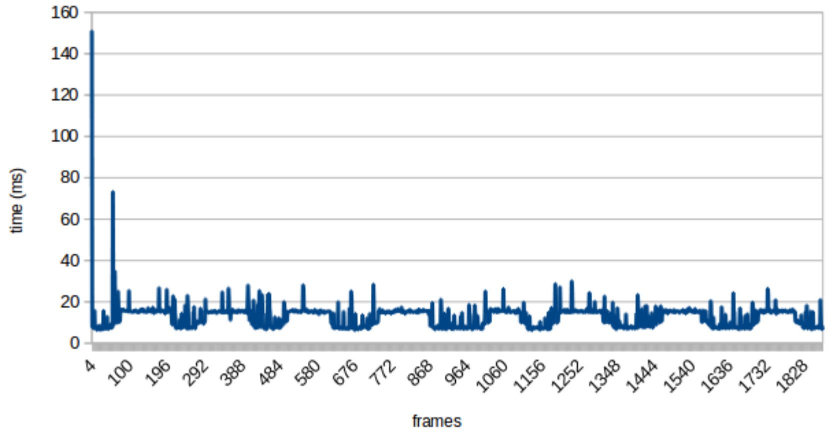
The next step corresponds to identify the person who passes again under the camera. The classification task is based on the predictor features extracted from each frame when the person passed through. At this point it would be enough to extract features only from a single frame for identifying the unique id of the person, but more frames are taken into account, greater will be the accuracy of the recognition of the correct person.

It is necessary that feature extraction and classification steps must be performed inside a time interval between two consecutive frames. Therefore it is resulting in less than 18 ms for the execution time of the classification step.

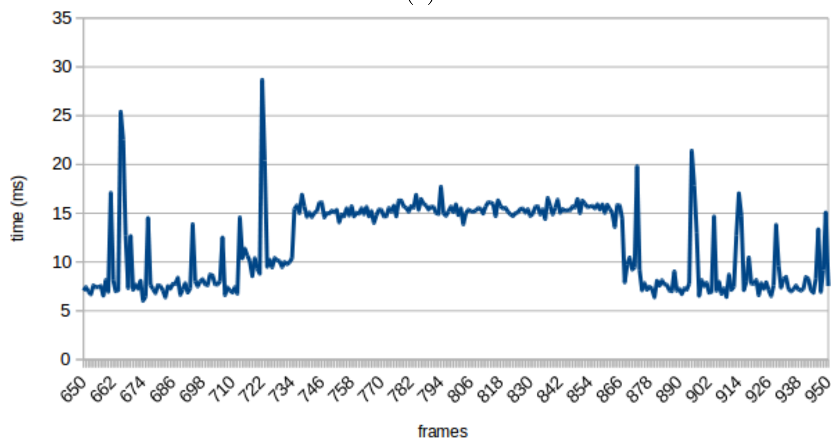
To evaluate our dataset, the performance results are reported in terms of recognition rate, using the CMC curves, as previously described in [9]. Figure 3 depicts a comparison between TVH and TVD in terms of CMC curves, to compare the ranks returned by using these different descriptors, where the horizontal axis is the rank of the matching score, the vertical axis is the probability of correct identification.

In particular, Figure 3a represents the CMC obtained for TVH. Figure 3b provides the CMC obtained for TVD. We compare these results with the average obtained by TVH and TVD. The average CMC is displayed in Figure 3d.

It can be assumed that the best performance is achieved when the combination of descriptors is used. It is possible to infer this aspect from Figure 3d where the combination of descriptors improve the results obtained by each of the descriptor separately. This result is due to the depth contribution that may be more informative. In fact, the depth outperforms the colour measure, giving the best performance for rank values higher than 15 (Figure 3b). Its better performance suggests the importance and potential of this descriptor.



(a)



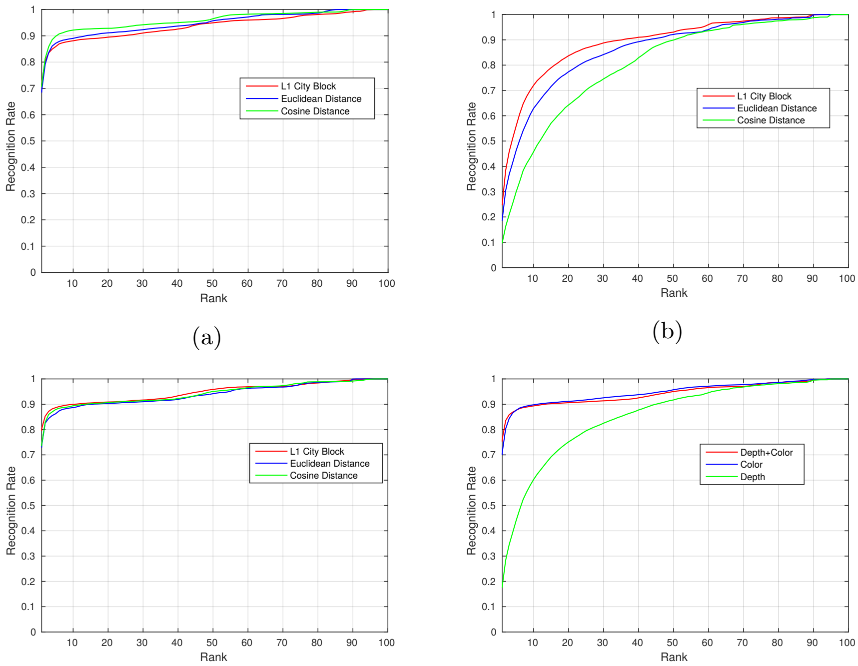
(b)

Fig. 2: (2a) describes the time occurring for the feature extraction frame by frame. (2b) shows a zoomed overview on several frames that correspond to a single person passing under the camera.

The classification process is performed with kNN, SVM, DT and RF classifiers. We carried out two experiments: a classic training/testing experiment and a gender classification, both based on TVPR dataset.

The task is solved using as a TVD descriptor an SVM with a quadratic degree of the polynomial kernel function, while the others descriptors are solved with SVM with a cubic degree of the polynomial kernel function. For the kNN classifier the "minkowski" as metric distance and "n neighbors = 5" has been chosen.

For the first case, we consider the first passage under the camera as training set and the return to the initial position as the testing set. The dataset is composed by 21685 instances divided in 11683 for training and 10002 for testing.



(a) (b)

(c) (d)

Fig. 3: The CMC curves obtained on TVPR Dataset.

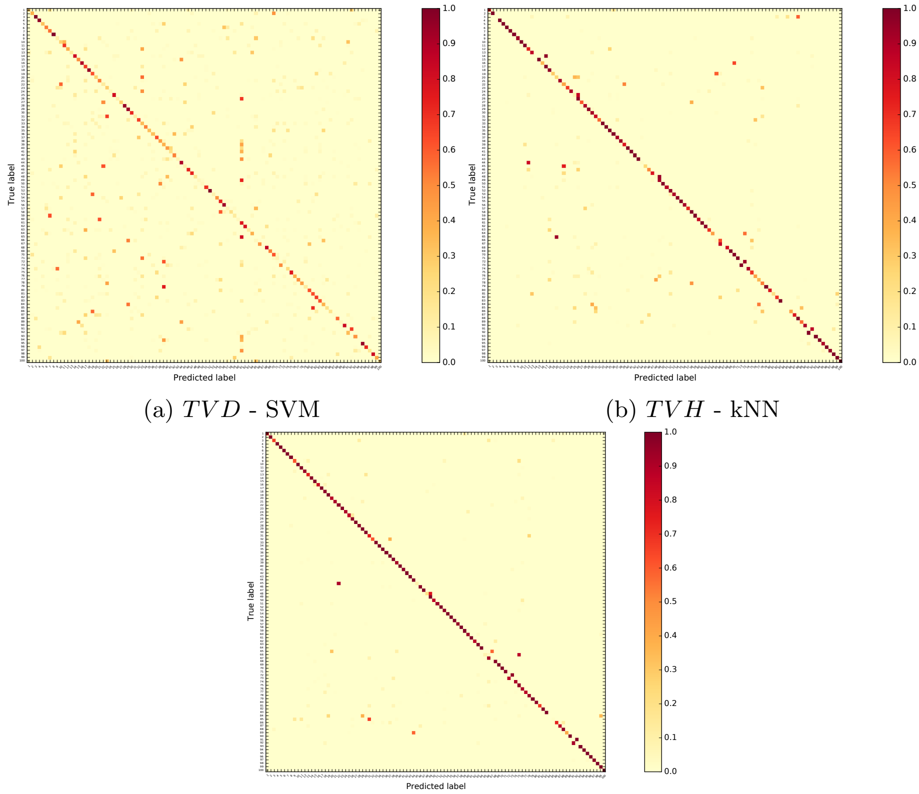
Table 1 reports, for each person of TVPR, the recognition results for kNN classifier with the TVDH descriptor.

The re-id classification performance of TVPR is summarized in Table 2 with a comparison among the descriptors TVH, TVD and TVDH. Figure 4 shows the best confusion matrices for the three descriptors: TVD with SVM classifier (Figure 4a, TVH with kNN classifier (Figure 4b) and TVDH with kNN classifier (Figure 4c).

In this case, we could observe high performance for our proposed approach to re-identify people. This accentuates the feasibility of utilizing colour as an effective cue in re-id scenarios. Moreover, by conducting the comparative study for the two descriptors TVD and TVH, we could observe the influence of colour for the re-id top view scenario. However, TVD descriptor is important for re-id, because it improves the overall precision as Figure 4c shows.

In this experiment, we try to classify gender considering the length of hair and colour. The results are summarized in Table 3. Figure 5 depicts the confusion matrix for the kNN classifier.

Results confirm the effectiveness and the suitability of the proposed approach. In fact, the class FSD "Female with dark and short hair" is confused, because females commonly have hair with considerable length. Same thing goes for class MLD "Male with dark and long hair", because generally short hair is an Italian male hairstyle. For the other class, classification overall precision is over 76%.



(a) TVD - SVM

(b) TVH - kNN

(c) $TVDH$ - SVM

Fig. 4: Confusion Matrices.

5 Conclusions and Future Works

In this paper, we describe a method for person re-identification based on features derived from both depth and colour. The experiments were conducted on TVPR dataset with an anthropometric and colour-based features set. The WCET of the whole process was estimated to ensure that computational time is within the constraints determined by the time necessary to send promotions to consumers in real time. Moreover, future development will ensure that execution time of all classification models is below 18 ms, and also that computational time falls within the useful time boundaries for the effectiveness of the proposed retail application. Person recognition is also handled using k-Nearest Neighbors classifier, Support Vector Machine, Decision Tree and Random Forest and we evaluate the performance in terms of Precision, Recall and F1-score. The classification is a classic training/testing experiment. Thus, a gender classification, based on colour and hair length, is carried out with the aim to improve retail applications. This approach is useful for different purposes in retail field. First of all, the study of returning customers and the identification of their shopping patterns allows predictive analytics to recommend products and offer personalized pricing or

Table 1: Classification results for each person of TVPR for kNN classifier with the TVDH descriptor.

ID	Precision	Recall	F1-S	Sup.	ID	Precision	Recall	F1-S	Sup.
1	0.90	0.85	0.87	53	51	0.84	0.20	0.33	103
2	0.70	0.74	0.72	43	52	0.58	1.00	0.73	110
3	1.00	0.91	0.95	54	53	0.99	0.87	0.93	100
4	0.90	1.00	0.95	69	54	1.00	0.94	0.97	101
5	0.93	0.98	0.95	86	55	0.99	1.00	0.99	94
6	1.00	0.95	0.98	109	56	0.92	0.97	0.94	67
7	0.85	0.98	0.91	63	57	0.99	1.00	1.00	105
8	1.00	1.00	1.00	102	58	1.00	1.00	1.00	76
9	1.00	1.00	1.00	86	59	1.00	1.00	1.00	93
10	1.00	1.00	1.00	85	60	0.96	1.00	0.98	91
11	1.00	1.00	1.00	84	61	0.94	1.00	0.97	120
12	1.00	1.00	1.00	101	62	0.96	0.94	0.95	126
13	1.00	1.00	1.00	73	63	1.00	1.00	1.00	65
14	1.00	1.00	1.00	82	64	1.00	0.88	0.94	68
15	0.96	1.00	0.98	73	65	0.93	0.99	0.96	145
16	0.75	0.62	0.68	73	66	1.00	1.00	1.00	125
17	1.00	1.00	1.00	116	67	0.00	0.00	0.00	98
18	0.88	0.99	0.93	113	68	0.03	0.04	0.03	112
19	0.95	0.96	0.95	93	69	0.00	0.00	0.00	101
20	1.00	0.98	0.99	93	70	1.00	1.00	1.00	157
21	0.90	1.00	0.95	94	71	1.00	1.00	1.00	163
22	0.99	0.84	0.90	91	72	0.98	0.98	0.98	121
23	0.99	1.00	0.99	98	73	0.00	0.00	0.00	82
24	0.79	0.97	0.87	107	74	0.00	0.00	0.00	149
25	0.73	1.00	0.85	77	75	0.96	0.91	0.93	107
26	0.71	0.88	0.79	94	76	0.48	0.96	0.64	114
27	0.98	0.91	0.94	140	77	0.76	0.91	0.83	78
28	0.23	0.97	0.37	31	78	0.99	0.88	0.93	179
29	1.00	0.98	0.99	123	79	0.71	0.94	0.81	64
30	0.97	0.86	0.92	169	80	1.00	0.97	0.98	131
31	0.86	0.97	0.91	171	81	1.00	0.68	0.81	62
32	1.00	1.00	1.00	151	82	1.00	0.99	0.99	83
33	0.91	0.97	0.94	111	83	1.00	1.00	1.00	77
34	0.74	1.00	0.85	112	84	0.00	0.00	0.00	80
35	0.94	0.99	0.96	134	85	0.12	0.01	0.02	76
36	0.50	0.75	0.60	84	86	1.00	0.73	0.85	49
37	0.95	0.61	0.74	88	87	1.00	0.88	0.93	72
38	0.99	1.00	1.00	102	88	0.91	0.96	0.94	84
39	1.00	1.00	1.00	97	89	1.00	0.41	0.58	139
40	1.00	1.00	1.00	77	90	0.00	0.00	0.00	103
41	0.65	1.00	0.79	72	91	0.00	0.00	0.00	100
42	0.83	0.99	0.90	101	92	1.00	1.00	1.00	152
43	0.89	0.92	0.90	98	93	1.00	1.00	1.00	99
44	0.99	1.00	1.00	130	94	0.98	1.00	0.99	100
45	1.00	0.97	0.98	100	95	1.00	1.00	1.00	92
46	1.00	1.00	1.00	118	96	1.00	0.97	0.99	110
47	1.00	1.00	1.00	101	97	1.00	1.00	1.00	157
48	0.59	1.00	0.74	116	98	0.74	1.00	0.85	87
49	1.00	0.09	0.16	113	99	1.00	1.00	1.00	91
50	0.99	1.00	1.00	100	100	0.95	0.67	0.78	93
					AVG	**0.85**	**0.85**	**0.83**	10002

promotions. Customer analytics are also the most useful instrument to address both consumer and enterprise needs. The experimental results demonstrate the effectiveness and suitability of our approach that achieves high accuracy and performs better without having to rely on the data annotation required in the other existing approaches. Further investigation will be devoted to improving our approach by extracting other informative features and setting up a full neural network for the real time processing of video images. Future works include also the evaluation of the necessary resources for the design of CNN layers.

In the field of retail, the long term goal of this work is to integrate this re-identification system with an audio framework, and to use other types of RGB-D cameras such as time of flight (TOF) ones. The system can additionally be integrated 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, that can alert one to a possible uncontrolled circumstance. It would also be interesting to evaluate both colour

Table 2: Training/Testing Classification results for TVD, TVH and TVDH descriptors.

	Classifier	Precision	Recall	F1-Score
TVD	KNN	0.35	0.32	0.31
	SVM	**0.48**	**0.43**	**0.42**
	Decision Tree	0.37	0.34	0.33
	Random Forest	0.46	0.43	0.42
TVH	KNN	**0.75**	**0.73**	**0.71**
	SVM	0.70	0.67	0.64
	Decision Tree	0.49	0.46	0.45
	Random Forest	0.71	0.70	0.68
TVDH	KNN	0.81	0.80	0.79
	SVM	**0.85**	**0.85**	**0.83**
	Decision Tree	0.52	0.50	0.48
	Random Forest	0.74	0.71	0.69

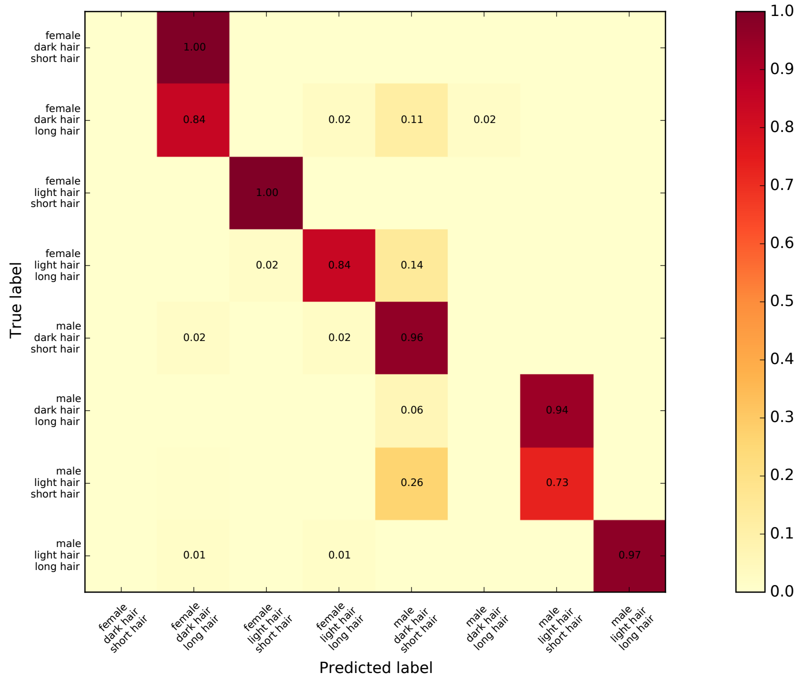


Fig. 5: Gender Classification Confusion Matrix with kNN classifier.

and depth images in a way that does not decrease the performance of the system when the colour image is being affected by changes in pose and/or illumination.

Table 3: Gender Classification results with kNN classifier.

Class	Gender	Hair Type	Precision	Recall	F1-S	Sup.
FSD	Female	Short Dark	0.00	0.00	0.00	101
FLD	Female	Long Dark	0.93	0.84	0.88	3036
FSL	Female	Short Light	0.92	1.00	0.96	157
FLL	Female	Long Light	0.76	0.84	0.80	708
MSD	Male	Short Dark	0.89	0.96	0.92	5222
MLD	Male	Long Dark	0.00	0.00	0.00	98
MSL	Male	Short Light	0.82	0.73	0.77	612
MLL	Male	Long Light	1.00	0.97	0.99	68
			0.87	**0.88**	**0.88**	10002

6 Acknowledgement

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 behaviour analysis in domestic and commercial environments" (HBA 2.0 – Human Behaviour Analysis).

References

1. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. ACM Computing Surveys (CSUR) **46**(2) (2013) 29
2. Chahla, C., Snoussi, H., Abdallah, F., Dornaika, F.: Discriminant quaternion local binary pattern embedding for person re-identification through prototype formation and color categorization. Engineering Applications of Artificial Intelligence **58** (2017) 27–33
3. Hariri, W., Tabia, H., Farah, N., Benouareth, A., Declercq, D.: 3d facial expression recognition using kernel methods on riemannian manifold. Engineering Applications of Artificial Intelligence **64** (2017) 25–32
4. Farou, B., Kouahla, M.N., Seridi, H., Akdag, H.: Efficient local monitoring approach for the task of background subtraction. Engineering Applications of Artificial Intelligence **64** (2017) 1–12
5. Lisanti, G., Masi, I., Bagdanov, A.D., Del Bimbo, A.: Person re-identification by iterative re-weighted sparse ranking. IEEE transactions on pattern analysis and machine intelligence **37**(8) (2015) 1629–1642
6. Paolanti, M., Liciotti, D., Pietrini, R., Mancini, A., Frontoni, E.: Modelling and forecasting customer navigation in intelligent retail environments. Journal of Intelligent & Robotic Systems (2017) 1–16
7. 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 (2014) 146–157
8. 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: International Conference on Image Analysis and Processing, Springer (2017) 207–218
9. 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 (2017)

10. 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 Recognition Letters (2016)
11. Calvaresi, D., Cesarini, D., Sernani, P., Marinoni, M., Dragoni, A.F., Sturm, A.: Exploring the ambient assisted living domain: a systematic review. Journal of Ambient Intelligence and Humanized Computing **8**(2) (2017) 239–257
12. Calvaresi, D., Marinoni, M., Sturm, A., Schumacher, M., Buttazzo, G.: The challenge of real-time multi-agent systems for enabling iot and cps. In: Proceedings of the International Conference on Web Intelligence, ACM (2017) 356–364
13. Sernani, P., Calvaresi, D., Calvaresi, P., Pierdicca, M., Morbidelli, E., Dragoni, A.F.: Testing intelligent solutions for the ambient assisted living in a simulator. In: Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments, ACM (2016) 71
14. Paolanti, M., Kaiser, C., Schallner, R., Frontoni, E., Zingaretti, P.: Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks. In: International Conference on Image Analysis and Processing, Springer (2017) 402–413
15. 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: Mobile Robots (ECMR), 2017 European Conference on, IEEE (2017) 1–6
16. Sturari, M., Paolanti, M., Frontoni, E., Mancini, A., Zingaretti, P.: Robotic platform for deep change detection for rail safety and security. In: Mobile Robots (ECMR), 2017 European Conference on, IEEE (2017) 1–6
17. Messelodi, S., Modena, C.M.: Boosting fisher vector based scoring functions for person re-identification. Image and Vision Computing **44** (2015) 44–58
18. Havasi, L., Szlávik, Z., Szirányi, T.: Eigenwalks: Walk detection and biometrics from symmetry patterns. In: IEEE International Conference on Image Processing 2005. Volume 3., IEEE (2005) III–289
19. Fischer, M., Ekenel, H.K., Stiefelhagen, R.: Interactive person re-identification in tv series. In: Content-Based Multimedia Indexing (CBMI), 2010 International Workshop on, IEEE (2010) 1–6
20. Calderara, S., Prati, A., Cucchiara, R.: Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance. Computer Vision and Image Understanding **111**(1) (2008) 21–42
21. Javed, O., Shafique, K., Rasheed, Z., Shah, M.: Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views. Computer Vision and Image Understanding **109**(2) (2008) 146–162
22. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS). Volume 3., Citeseer (2007)
23. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE (2010) 2360–2367
24. Alahi, A., Vandergheynst, P., Bierlaire, M., Kunt, M.: Cascade of descriptors to detect and track objects across any network of cameras. Computer Vision and Image Understanding **114**(6) (2010) 624–640
25. Gandhi, T., Trivedi, M.M.: Panoramic appearance map (pam) for multi-camera based person re-identification. In: 2006 IEEE International Conference on Video and Signal Based Surveillance, IEEE (2006) 78–78
26. Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Volume 2., IEEE (2006) 1528–1535
27. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: European conference on computer vision, Springer (2008) 262–275

28. Bazzani, L., Cristani, M., Perina, A., Farenzena, M., Murino, V.: Multiple-shot person re-identification by hpe signature. In: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE (2010) 1413–1416
29. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recognition Letters **33**(7) (2012) 898–903
30. Schwartz, W.R., Davis, L.S.: Learning discriminative appearance-based models using partial least squares. In: Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on, IEEE (2009) 322–329
31. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: Asian conference on Computer vision, Springer (2010) 501–512
32. Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC. Volume 2. (2010) 6
33. Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2288–2295
34. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE transactions on pattern analysis and machine intelligence **35**(3) (2013) 653–668
35. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: Computer vision and pattern recognition (CVPR), 2011 IEEE conference on, IEEE (2011) 649–656
36. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: European Conference on Computer Vision, Springer (2012) 381–390
37. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: European Conference on Computer Vision, Springer (2012) 780–793
38. Satta, R., Fumera, G., Roli, F.: Fast person re-identification based on dissimilarity representations. Pattern Recognition Letters **33**(14) (2012) 1838–1848
39. Gong, S., Cristani, M., Yan, S., Loy, C.C.: Person re-identification. Volume 1. Springer (2014)
40. Bazzani, L., Cristani, M., Murino, V.: Sdalf: modeling human appearance with symmetry-driven accumulation of local features. In: Person Re-Identification. Springer (2014) 43–69
41. Ess, A., Leibe, B., Gool, L.V.: Depth and appearance for mobile scene analysis. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE (2007) 1–8
42. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC. Volume 1. (2011) 6
43. Barbosa, I.B., Cristani, M., Del Bue, A., Bazzani, L., Murino, V.: Re-identification with rgb-d sensors. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer (2012) 433–442
44. Duda, R.O., Hart, P.E., et al.: Pattern classification and scene analysis. Volume 3. Wiley New York (1973)
45. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3) (1995) 273–297
46. Vladimir, V.N., Vapnik, V.: The nature of statistical learning theory (1995)
47. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory, ACM (1992) 144–152
48. Quinlan, J.R.: C4. 5: programs for machine learning. Elsevier (2014)
49. Breiman, L.: Random forests. Machine learning **45**(1) (2001) 5–32