A Robust Model for Person Re-Identification in Multimodal Person Localization

Thi Thanh Thuy Pham
Faculty of Information Technology
University of Technology and Logistics
Bacninh, Vietnam
Email: thanh-thuy.pham@mica.edu.vn

Thi Lan Le
and Trung Kien Dao
and Duy Hung Le
International Research Institute MICA
University of Science and Technology
Hanoi, Vietnam

Van Toi Nguyen
The University of Information
and Communication Technology
under Thai Nguyen University
ThaiNguyen, Vietnam

***Abstract*—Determining person ID (Identity) is one of the crucial steps in indoor human localization system. It is more exactly stated as person Re-ID (Re-Identification) problem because for each user's position, the user ID at the first occurrence needs to be shown correspondingly at the later times of localization. In this paper, a multimodal person localization system of WiFi and camera is proposed, with the analysis for the key role of appearance-based person Re-ID in fusing different information sources. A new model for person appearance representation based on kernel descriptor is proposed to tackle the challenges of person Re-ID in camera network. Additionally, a dataset for the real scenario of the proposed multimodal person localization is also established, in which we set a database for vision-based human Re-ID evaluation. The experiments on the benchmark datasets and our dataset show the outperforming results in comparison with other state-of-the-art approaches.**

Keywords–Multimodal person localization; Person Re-ID; KDES descriptor; Camera; WiFi.

I. INTRODUCTION

Person Re-ID and positioning are two key problems in a typical human localization system. In case of multi-object localization, we need to identify the person who is localized, therefore we know the determined positions belong to which objects. Person Re-ID in camera network is a hard problem and increasingly attracted many researchers. Three basic steps need to be done for vision-based person Re-ID problem. People detection in consecutive frames is firstly executed, then feature extraction within the detected regions and feature descriptor is generated, finally object matching is done for Re-ID. Each step has its own challenges and these affect strongly to the system performance. In general, they include (1) illumination conditions that are different by time and space; (2) pose, scale and appearance variation of people at distinctive camera FOVs (Fields of View). This is considered as the most challenging, because the human appearance features are mainly used in the human re-identification system; (3) occlusions in which people are obscured by each other or obstacles in the environment; (4) re-identification scenarios involving closed set Re-ID (the identified objects are included in both gallery and probe sets) or open set Re-ID (the objects may not be contained in the gallery set).

Many approaches are proposed for vision-based person Re-ID problem, however most of them are oriented to (1) build a distinctive feature descriptor for each object and then apply an effective object classifier for that or (2) design potential distance metrics from data. In this paper, we concentrate on establishing a robust feature descriptor which improves the original KDES (Kernel Descriptor) of [1], and applying multi-class SVM as relative ranking for person Re-ID in camera network. This is proven to be more robust than original KDES or other state-of-the-art methods in solving vision-based person Re-ID problem.

The rest of the paper is organized as follow. In Section II, the related works on vision-based human Re-ID are presented. Section III indicates a combined system of WiFi and visual signals for human localization, in which appearance-based person Re-ID problem in camera network is solved by improved KDES. Some experimental results on benchmark datasets and our dataset are shown in Section IV. Conclusion and future directions will be finally denoted.

II. RELATED WORK

Design of a robust person descriptor is the most decisive step for vision-based person Re-ID problem. Many kinds of features are utilized for this, in which human body appearance is the simplest and the most popular one. Color, texture, and shape are features that can be extracted for human appearance. In [2][3], color histogram is used for feature descriptor. There are two ways to represent the image of detected people with color histogram: global color histogram and local color histogram. A single histogram is used in the first method for the whole image, while in the second way, image is divided into some parts and concatenating the part-based color histograms is done to give a final result. Most reported person Re-ID works pay attention on the second solution, such as in [4], a weighted color histogram derived from MSCR (Maximally Stable Colour Regions) and structured patches are combined for visual description. In [5][6], color histogram on different color models is calculated and syndicated with texture features to make person descriptor more robust. Shape features are also extracted for appearance model. However, they are unstable because of non-rigid objects as people; so, in [7], color and texture features are associated with shape feature to enhance the effectiveness of person descriptor. Local region descriptors, such as SIFT (Scale-Invariant Feature Transform), SURF (Speeded Up Robust Features) and GLOH (Gradient Location and Orientation Histogram) are evaluated in [8] for person Re-ID in image sequences. The results show that GLOH and SIFT outperform both shape context and SURF descriptor. Additionally, a large number of visual features are exploited

 ISBN: 978-1-61208-418-3

for person Re-ID problem, such as Haar-like features, HOG (Histogram of Oriented Gradients), edges, covariance, interest points, etc.

The next step in human Re-ID process is classification, with two scenarios of single-shot and multi-shot being reported. The first case is more simple with one-to-one matching between a pair of probe and gallery image for each person, whereas in the second scenario, each object has multiple images, either in the gallery or the probe set. In general, the purpose of classification in person Re-ID is finding out the most similar candidate for a target or ranking the candidates based on a standard distance minimization strategy, which is known as distance metric. This metric can be chosen independently (non-learning based method) [9] or learned from the data (learning-based method) [10] in order to minimize intra-class variation whilst maximize extra-class variation. They typically include histogram-based Bhattacharyya distance, K Nearest Neighbor classifiers, L1-Norm, diffusion distance [11]. Additionally, some later proposed methods, such as LMNN-R (Large Margin Nearest Neighbor) distance metric in [12] or PRDL (Probabilistic Relative Distance Learning) in [13] are more robust.

To get an ID ranking list, distance scores between true and wrong matches can be compared directly or relatively (ranking the scores that show the correspondence of each likely match to the probe image). The relative ranking treated by either Boosting as RankBoost in [14] or kernel-based learning, such as RankSVM [15], primal-based RankSVM [16] or Ensemble RankSVM [6].

III. PROPOSED SYSTEM

A. Overview of multimodal person localization system

Object localization is known as a problem of determining the object position in the environment. For each user in multi-user localization system, two problems of positioning (where the user is) and identifying (who the user is) must be solved simultaneously. A general diagram for object localization is illustrated in Figure 1. In this figure, the input cues can come from different sensors, such as optical, radio frequency, ultra sound, inertia, DC Electromagnetic sensors, etc. From the input cues, localization and Re-ID are executed simultaneously to give the output for object position and ID. Multimodal object localization is defined as a problem of multi-cue combination for input or fusion of different positioning methods. As proven by Vinyals et al. [17] and Dao et al. [18], compounding of different models gives better positioning results than applying a single model. Teixeira et al. [19] proposed to use the motion signature taken from wearable accelerometer for identifying people in camera network.

Figure 1. Flowchart of object localization system.

Our research aims at developing a multimodal person localization system by using both WiFi and camera systems. This offers some benefits in comparison with single-method systems. (1) System setting cost is limited because of available WiFi infrastructure and uncrowded-deployed cameras. (2) Positioning range is easily broaden by simply adding more APs (Access Point) in the environment. (3) Computational expense is much lower for WiFi-based than vision-based positioning system. (4) The positioning accuracy is provided in accordance with the application-specific demands. Although the camera-based system brings more impressed positioning results, but not every where in building needs high localizing accuracy. (5) Sampling frequency is improved for the WiFi-based system, because it has lower sampling rate (about one signal measure per second) than vision-based system (approximately 15 fps). (6) The information for person Re-ID becomes richer. One object can be identified simultaneously by both WiFi and camera systems. These ID cues can be used in the way of supporting one another in multi-model object localization system. For example, at a certain time, one object is localized and identified by WiFi system, with the position of P_{WiFi} and the identity of ID_{WiFi} (the MAC address of mobile device) respectively. At the same time, this object is also determined by P_{cam} and ID_{cam} from camera system. However, P_{WiFi} is not as accurate as P_{cam}, whilst ID_{WiFi} is clearer than ID_{cam}. Therefore, by using both of these systems, the object can be localized by P_{cam} and identified by ID_{WiFi}.

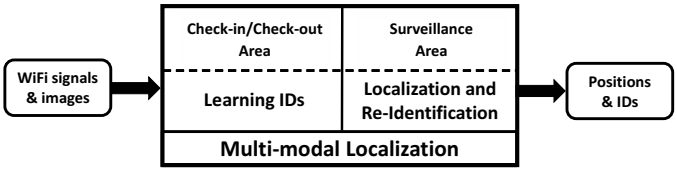

Figure 2. Multimodal localization system fusing WiFi signals and images.

Figure 2 shows a framework for our multi-model human localization system using both WiFi signals and camera network. The framework indicates that the proposed system is implemented in two subregions of the whole positioning area: check-in/check-out region and surveillance region. In the first region, learning ID cues is executed. Person holding a WiFi-integrated device will one by one come in and come out of the first region. At the entrance of the first region, the person's ID will be learned individually by the images captured from cameras and MAC address of WiFi-enable equipment held by that person. One camera, which is in front door of check-in gate, captures human face and then a face recognition program is executed. Another camera acquires human body images at different poses and learning phase of appearance-based ID is done for each person. In short, in the first region, we get three types of signature for each person (N_i): face-based ID (ID^i_F), WiFi-based ID (ID^i_{WF}), and appearance-based ID (ID^i_{Apr}). Depending on different circumstances, we can map among signatures of (ID^i_F, ID^i_{WF}), (ID^i_F, ID^i_{Apr}), (ID^i_{Apr}, ID^i_{WF}) and utilize them for person localization and identification in the surveillance region. The user will end up his route at the exit gate and he will be checked out by other camera. This camera acquires human face for person Re-ID, and based on this, the user will be removed from the localization system. By using check-in/check-out region, we can (1) control the human appearance changes (the difference in cloth colors) at each time people come in the positioning area, (2) decrease the computing cost by eliminating the checked-out users from the system, (3) map between different ID cues for the same person.

 ISBN: 978-1-61208-418-3

In the surveillance region, two problems of person localization and Re-ID will be solved concurrently by combining visual and WiFi information. Figure 3 demonstrates a surveillance region which contains WiFi range and camera FOVs. In this region, the WiFi range covers some visual ranges (the camera FOVs: FOV of C_1, FOV of C_2,.., FOV of C_n). This means the user always move within WiFi range but switch from one camera FOV to others.

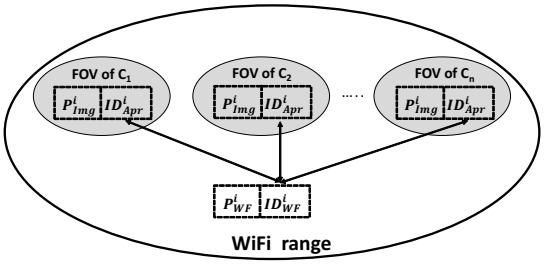

Figure 3. Surveillance region with WiFi range and disjoint cameras' FOVs.

In each camera FOV and for an individual, we calculate image-based and WiFi-based positions (P^i_{img}, P^i_{WF}) and ID^i_{Apr}. From ID^i_{Apr}, we know ID^i_{WF} correspondingly by ID mapping result taken from the first region. Outside the camera FOV, there only exits the information of P^i_{WiFi}, ID^i_{WF}, and ID^i_{Apr} respectively. When people switch from one camera FOV to others, their positions and IDs will be updated in the WiFi-available region. The localization accuracy then be tuned by combination of WiFi-based and vision-based systems.

From the above analysis, we see that finding ID^i_{Apr} plays a key role in the proposed multimodal person localization system. It is used to link the object trajectories from one camera range to others through the intermediate positioning range of WiFi. Therefore, ID^i_{Apr} must be shown at each frame captured from different cameras in the surveillance area. That means the appearance-based person Re-ID problem needs to be solved. In this circumstance, it belongs to multi-shot person Re-ID problem, with multiple images for each detected person at different resolutions, lighting conditions, and poses are processed.

B. Vision-based person re-identification

1) The system overview: The flowchart of vision-based person Re-ID system is illustrated in Figure 4. It includes three stages of (1) person detection, (2) feature extraction, and (3) classification. In the first stage, from the input frames, the ROI (Region of Interest) of person can be determined by using the state-of-the-art methods. The features are then extracted from these regions and feature descriptors are created in the second stage. Finally, a classifier is applied to learn the person model and predict the corresponding ID.

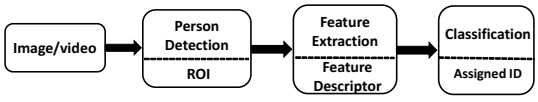

Figure 4. A diagram of vision-based person Re-ID system.

In this section, we present in detail the second stage since it is the main contribution of our paper. For this, we propose a new person appearance representation model based on KDES. This descriptor is firstly proposed by Bo et al. [1] and has been proved to be robust for recognition of different objects, such as hand pose recognition [20].

2) KDES-based person representation: The basic idea of the representation based on kernel methods is to compute the approximate explicit feature map for kernel match function (see Figure 5). In other words, the kernel match functions are approximated by explicit feature maps. This enables efficient learning methods for linear kernels to be applied to the non-linear kernels. This approach was introduced in [1][21]. Given a match kernel function $k(x, y)$, the feature map $\varphi(.)$ for the kernel $k(x, y)$ is a function mapping a vector x into a feature space so as $k(x, y) = \varphi(x)^\top \varphi(y)$. Suppose that we have a set of basis vectors $B = \{\varphi(v_i)\}_{i=1}^{D}$, the approximation of feature map $\varphi(x)$ can be:

$$\phi(x) = Gk_B(x) \tag{1}$$

where G is defined by: $G^\top G = K_{BB}^{-1}$ and K_{BB} is $D \times D$ matrix with $\{K_{BB}\}_{ij} = k(v_i, v_j)$. k_B is a $D \times 1$ vector with $\{k_B\}_i = k(x, v_i)$.

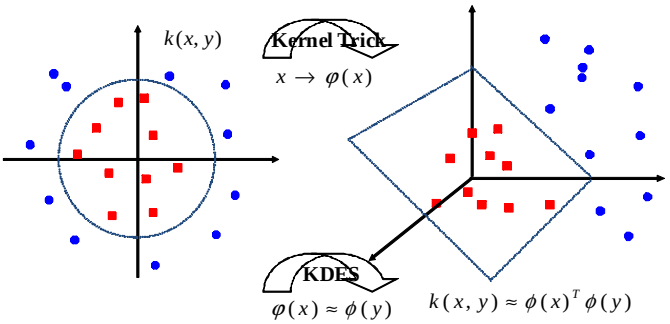

Figure 5. The basic idea of representation based on kernel methods.

Feature extraction is then done at three levels of pixel, patch and the whole image of detected person. At pixel level, a normalized gradient vector is computed for each pixel of the image. The normalized gradient vector at a pixel z is defined by its magnitude $m(z)$ and normalized orientation $\omega(z) = \theta(z) - \overline{\theta}(P)$, where $\theta(z)$ is orientation of gradient vector at the pixel z, and $\overline{\theta}(P)$ is the dominant orientation of the patch P that is the vector sum of all the gradient vectors in the patch. This normalization will make patch-level features invariant to rotation. In practice, the normalized orientation of a gradient vector will be:

$$\widetilde{\omega}(z) = [sin(\omega(z))\ cos(\omega(z))] \tag{2}$$

At the second level, the image with different resolutions will be divided into a grid of a fix number of cells as in [20], instead of size-fixed cells as in [1]. A patch is then set by 2×2 cells and two adjacent patches along x-axis or y-axis are overlapped at two cells. This division results to size-adaptive patches to the different image resolutions, and nearly the same feature vectors for the scale-varied images of intraclass are created (see Figure 6). In our work, this technique is utilized for KDES extraction because of a large variation of person size caused by different distances from pedestrian to the stationary camera. For each patch, we compute patch features based on

 ISBN: 978-1-61208-418-3

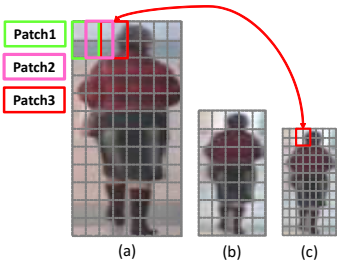

Figure 6. Illustration of size-adaptive patches (a, c) and size-fixed patches (a, b) which is mentioned in [1].

a given definition of match kernel. The gradient match kernel is constructed from three kernels: gradient magnitude kernel $k_{\widetilde{m}}$, orientation kernel k_o, and position kernel k_p.

$$K_{gradient}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} k_{\widetilde{m}}(z,z') k_o(\widetilde{\omega}(z), \widetilde{\omega}(z')) k_p(z,z') \tag{3}$$

where P and Q are patches of two different images need to measure the similarity. z and z' denote the 2D positions of a pixel in the image patch P and Q respectively. $\varphi_o(.)$ and $\varphi_p(.)$ are the feature maps for the gradient orientation kernel k_o and position kernel k_p respectively. Then, the approximate feature over the image patch P is constructed as:

$$\overline{F}_{gradient}(P) = \sum_{z \in P} \widetilde{m}(z) \phi_o(\widetilde{\omega}(z)) \otimes \phi_p(z) \tag{4}$$

where $\otimes$ is a Kronecker product, $\phi_o(\widetilde{\omega}(z))$ and $\phi_p(z)$ are approximate feature maps (1) for the kernel k_o and k_p respectively.

The last level is finished by creating a complete descriptor for the whole image. As in [22], a pyramid structure is used to combine patch features. Given an image, the final representation is built based on features extracted from lower levels using EMK (Efficient Match Kernels) proposed in [1]. First, the feature vector for each cell of the pyramid structure is computed. The final descriptor is the concatenation of feature vectors of all cells.

Let C be a cell that has a set of patch-level features $X = \{x_1, ..., x_p\}$, then the feature map on this set of vectors is defined as:

$$\overline{\phi}_S(X) = \frac{1}{|X|} \sum_{x \in X} \phi(x) \tag{5}$$

where $\phi(x)$ is approximate feature map (1) for the kernel $k(x,y)$. The feature vector on the set of patches, $\overline{\phi}_S(X)$, is extracted explicitly.

Given an image, let L be the number of spatial layers to be considered. In this case, $L = 3$. The number of cells in layer l-th is (n_l). $X(l,t)$ is a set of patch-level features that fall within the spatial cell (l,t) (cell t-th in the l-th level). A patch is fallen in a cell when its centroid belongs to the cell. The feature map on the pyramid structure is:

$$\begin{aligned} \overline{\phi}_P(X) \quad = \quad & [w^{(1)} \overline{\phi}_S(X^{(1,1)}); ...; w^{(l)} \overline{\phi}_S(X^{(l,t)}); \\ & ...; w^{(L)} \overline{\phi}_S(X^{(L,n_L)})] \end{aligned} \tag{6}$$

In (6), $w^{(l)} = \frac{\frac{1}{n_l}}{\sum_{l=1}^{L} \frac{1}{n_l}}$ is the weight associated with level l.

Once the KDES computed, multiclass SVM is applied to train the model for each person. For each detected instance, a list of ranked objects will be generated based on the probability of SVM.

IV. Experimental Results

This section will present the testing datasets and the results obtained for vision-based person Re-ID. The CMC (Cumulative Match Curve) is employed as the performance evaluation method for person Re-ID problem. The CMC curve represents the expectation of finding correct match in the top n matches.

A. Testing datasets

In our experiments, two multi-shot benchmark datasets of CAVIAR4REID and i-LIDS are used. We also build our own dataset in the context of multimodal person localization. The CAVIAR4REID dataset includes 72 pedestrians, in which 50 of them are captured from two camera views and the remaining 22 from one camera view. i-LIDS dataset contains 119 individuals, with the images captured from multi-camera network. Both of them, especially CAVIAR4REID, are challenging because of broad changes in resolution, lighting condition, occlusion, and human pose. Concerning to our dataset, we build it for multimodal person localization evaluation. A database for testing appearance-base person Re-ID is also established in this.

Figure 7 shows the 8th floor plan of our office building. It is set as the testing environment for our combined person localization system. At the entrance, people hold smart phones or tablets go one by one through the check-in gate, then move inside the surveillance area, and finish their routes by going out check-out gate.

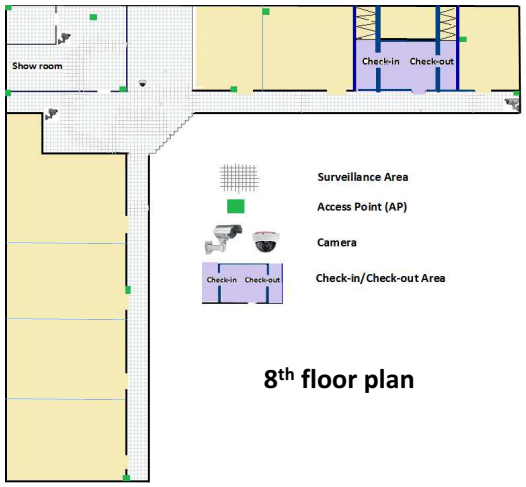

Figure 7. Testing environment.

In the check-in and check-out area, we set three cameras. Two of them are used at the entrance. One camera captures human face in order to check-in user by face recognition. The remaining camera acquires human body images at different poses. This will help the system learn appearance-based signature of the checked-in user. The third camera is used to capture human face at the exit, and based on this, the system will check out or release the user from its process. In the surveillance

area, four cameras with non-overlapping FOVs are deployed along the hallway and in a room. People are detected, localized, and re-identified at each frame captured from these cameras. Besides this, 11 APs are established throughout the testing environment. RSSIs (Received Signal Strength Indicators) and the MAC address are consecutively scanned and sent from mobile device to the server to calculate the position and ID of the device holder.

In short, a total of seven AXIS IP cameras and eleven APs are deployed throughout the testing environment of the 8^{th} storey floor plan. These cameras and APs are fixed at certain distances from the floor ground (about 1.6m-2.2m for cameras and 2m-2.8m for APs). They are configured with static IP addresses. The camera frame rate is set to 20 fps and image resolution is 640x480.

The dataset for human Re-ID includes 25 people with different routes in the testing environment. Each person spends from 3 to 5 minutes for his route. An approximation of 800 values of RSSIs are scanned, about 2000 frames are captured for each camera in the surveillance area. All captured frames are processed as real Re-ID scenario of multimodal pedestrian localization system. Firstly, the images of person body at different poses are extracted from video sequence of the entrance camera. They are later used for training phase of appearance-based person identification. In the surveillance area, each frame from the sequences of four cameras is processed for human detection, so each of 25 pedestrians will have the appearance images (the bounding boxes of each person) at different views. The image filename identifies the video sequence to which it belongs, the camera ID, frame number, time (hour, minute, second), and the person ID: `VVV_WW_XXXXX_YYYYYY_ZZZ.jpg` (E.g., `025_01_01260_153702_012.jpg` means the appearance image belongs to video sequence 25; it is captured from the camera 01; frame number 1260; the time is 15h:37m:02s; the person ID is 12). These images are utilized for testing phase of person Re-ID system. An example in the dataset for person Re-ID in camera network is shown in Figure 8. The images on the top are used for training phase of appearance-based person identification. They are captured by a camera at the check-in/check-out region. The images for testing phase are shown at the bottom. They come from four different cameras in the surveillance region.

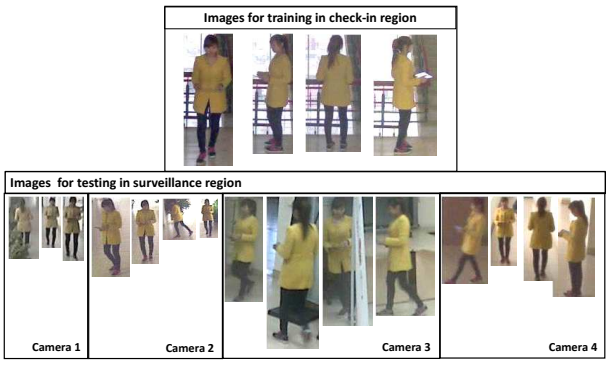

Figure 8. The instances in dataset for vision-based person Re-ID.

In comparison with other person Re-ID benchmark datasets, such as iLIDS, ETHZ, PRID 2011, CAVIAR4REID and VIPeR, our dataset contains multiple images for each person. These images are captured from many cameras (4 cameras) at different FOVs. This makes more variations for intraclass images in terms of resolution, illumination, pose and scale. In addition, it is set for real scenario of our proposed multimodal person localization system.

B. Person re-identification results

We compare the results of our proposed method with original KDES [1] and other state-of-the-art approaches on multi-shot datasets of CAVIAR4REID and iLIDs. The experimental settings are kept the same as in [23], with modified version of iLIDS dataset is selected (including only 69 individuals, with at least 4 images for each) and 72 pedestrians for CAVIAR4REID dataset, in which 50 different individuals are captured under both views. Each view has 10 images for each pedestrian. The outperforming results of the proposed method are shown in Figure 9. For CAVIAR4REID dataset, rank-1 recognition rate of AHPE (Asymmetry-based Histogram Plus Epitome) [23] is much lower than our method. It is only about 8 %, compared with 67.76 % of the original KDES [1] and 73.81 % for our method. However, both KDES and our method gain nearly the same figures from rank-13 and backward. For iLIDS dataset, the gap between rank-1 recognition rates of the proposed method with AHPE [23] or SDALF (Symmetry-Driven Accumulation of Local Features) [24] is approximately 20 %, and about 7 % with the original KDES [1]. This gap is slightly decreased for KDES but significantly reduced for AHPE and SDALF after rank-15.

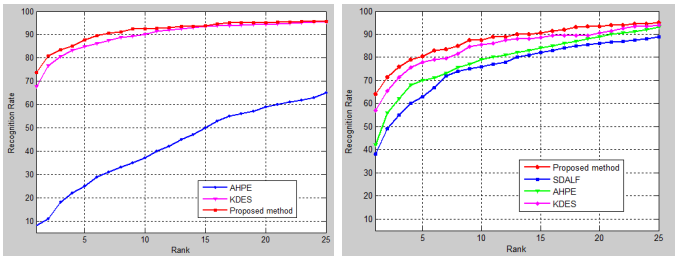

Figure 9. The results of proposed method against AHPE [23], SDALF [24] and KDES [1] on (a) CAVIAR4REID dataset and (b) iLIDS dataset.

Other experiments with iLIDs dataset are presented in Figure 10-a in comparison with other methods reported in [25]. The highest result for rank-1 belongs to RDC (Relational Divergence Classification) as mentioned in [25], but it is roughly 14% lower than our proposed method (66.18%). KDES [1] is tested on this dataset with 61.76% for rank-1, which is approximately 5% smaller than our method at the first 7 ranks.

The state-of-the-art SDALF [26] and the proposed method for person Re-ID are also tested on our dataset, with the gallery images (about 50 images for each class) from the entrance camera and the probe images (from 60 to 193 images for each person) from four cameras in the surveillance area (see Figure 10-b). The testing result is 73.13% at rank-1, compared with the original KDES of 67.16% and 30% for SDALF. The deviation between two recognition rates of our method and KDES gradually declines and almost reaches to the same value as SDALF after rank-21.

 ISBN: 978-1-61208-418-3

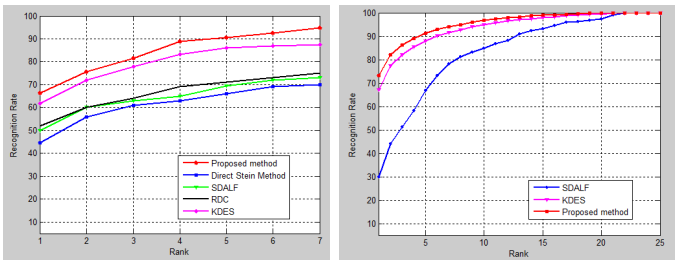

Figure 10. The comparative results with (a) reported methods in [25] and (b) the results are tested on our dataset.

The experimental results obtained with three different datasets have proved the better performance of the proposed method in comparison with the original KDES and other state-of-the-art methods. Based on these results, we will use the proposed method for person Re-ID in our multimodal human localization system.

V. Conclusion and Future Work

In this paper, person Re-ID problem in camera network achieves state-of-the-art performance on the benchmark datasets and our dataset by applying a robust person appearance representation based on KDES. The visual person ID can be used in connective and complementing manner of different types of information in the proposed multimodal pedestrian localization system of WiFi and camera. The experimental results are promising, and based on this, a multimodal method, which uses particle filter and integrated data association algorithm, will be promoted in the future work to increase the performance of the combined person Re-ID and localization system.

Acknowledgement

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.04-2013.32.

References

[1] L. Bo, X. Ren, and D. Fox, "Kernel descriptors for visual recognition," in Advances in Neural Information Processing Systems (NIPS), Vancouver, Canada, 2010, pp. 244–252.

[2] L. F. Teixeira and L. Corte-Real, "Video object matching across multiple independent views using local descriptors and adaptive learning," Pattern Recognition Letters, vol. 30, no. 2, 2009, pp. 157–167.

[3] D. N. T. Cong, C. Achard, L. Khoudour, and L. Douadi, "Video sequences association for people re-identification across multiple non-overlapping cameras," in Image Analysis and Processing (ICIAP). Springer, 2009, pp. 179–189.

[4] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification." in BMVC, vol. 2, no. 5. Citeseer, 2011, p. 6.

[5] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, and V. Murino, "Semi-supervised multi-feature learning for person re-identification," in Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on. IEEE, 2013, pp. 111–116.

[6] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking." in BMVC, vol. 2, no. 5, 2010, p. 6.

[7] N. Martinel, C. Micheloni, and C. Piciarelli, "Learning pairwise feature dissimilarities for person re-identification," in Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on. IEEE, 2013, pp. 1–6.

[8] M. Bauml and R. Stiefelhagen, "Evaluation of local features for person re-identification in image sequences," in Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on. IEEE, 2011, pp. 291–296.

[9] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on. IEEE, 2010, pp. 435–440.

[10] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 649–656.

[11] D. Figueira and A. Bernardino, "Re-identification of visual targets in camera networks: A comparison of techniques," in Image Analysis and Recognition. Springer, 2011, pp. 294–303.

[12] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in Computer Vision–ACCV 2010. Springer, 2011, pp. 501–512.

[13] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, no. 3, 2013, pp. 653–668.

[14] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," The Journal of machine learning research, vol. 4, 2003, pp. 933–969.

[15] D. Simonnet, M. Lewandowski, S. A. Velastin, J. Orwell, and E. Turkbeyler, "Re-identification of pedestrians in crowds using dynamic time warping," in Computer Vision–ECCV 2012. Workshops and Demonstrations. Springer, 2012, pp. 423–432.

[16] O. Chapelle and S. S. Keerthi, "Efficient algorithms for ranking with svms," Information Retrieval, vol. 13, no. 3, 2010, pp. 201–215.

[17] O. Vinyals, E. Martin, and G. Friedland, "Multimodal indoor localization: An audio-wireless-based approach," in Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on. IEEE, 2010, pp. 120–125.

[18] T. K. Dao, H. L. Nguyen, T. T. Pham, E. Castelli, V. T. Nguyen, and D. V. Nguyen, "User localization in complex environments by multimodal combination of gps, wifi, rfid, and pedometer technologies," The Scientific World Journal, vol. 2014, 2014.

[19] T. Teixeira, D. Jung, G. Dublon, and A. Savvides, "Identifying people in camera networks using wearable accelerometers," in Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2009, p. 20.

[20] V. T. Nguyen, T. L. Le, T. H. Tran, R. Mullot, and V. Courboulay, "A New Hand Representation Based on Kernels for Hand Posture Recognition," in The 11th IEEE International Conference on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 2015.

[21] S. Maji, A. C. Berg, and J. Malik, "Efficient classification for additive kernel svms," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, no. 1, 2013, pp. 66–77.

[22] L. Bo and C. Sminchisescu, "Efficient match kernel between sets of features for visual recognition," in Advances in neural information processing systems, 2009, pp. 135–143.

[23] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," Pattern Recognition Letters, vol. 33, no. 7, 2012, pp. 898–903.

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

[25] A. Alavi, Y. Yang, M. Harandi, and C. Sanderson, "Multi-shot person re-identification via relational stein divergence," arXiv preprint arXiv:1403.0699, 2014.

[26] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," Computer Vision and Image Understanding, vol. 117, no. 2, 2013, pp. 130–144.

 ISBN: 978-1-61208-418-3