Copyright ⓒ2017 KSII

Viewpoint Invariant Person Re-Identification for Global Multi-Object Tracking with Non-Overlapping Cameras

Jeonghwan Gwak [1,†], Geunpyo Park [1,‡] and Moongu Jeon [1,*]
[1] School of Electrical Engineering and Computer Science (EECS)
Gwangju Institute of Science and Technology (GIST)
Gwangju 61005 - South Korea
[e-mail: †james.han.gwak@gmail.com, mgjeon@gist.ac.kr]
‡First author equivalent
*Corresponding author: Moongu Jeon

Received May 18, 2016; revised January 2, 2017; accepted February 19, 2017; published April 30, 2017

Abstract

Person re-identification is to match pedestrians observed from non-overlapping camera views. It has important applications in video surveillance such as person retrieval, person tracking, and activity analysis. However, it is a very challenging problem due to illumination, pose and viewpoint variations between non-overlapping camera views. In this work, we propose a viewpoint invariant method for matching pedestrian images using orientation of pedestrian. First, the proposed method divides a pedestrian image into patches and assigns angle to a patch using the orientation of the pedestrian under the assumption that a person body has the cylindrical shape. The difference between angles are then used to compute the similarity between patches. We applied the proposed method to real-time global multi-object tracking across multiple disjoint cameras with non-overlapping field of views. Re-identification algorithm makes global trajectories by connecting local trajectories obtained by different local trackers. The effectiveness of the viewpoint invariant method for person re-identification was validated on the VIPeR dataset. In addition, we demonstrated the effectiveness of the proposed approach for the inter-camera multiple object tracking on the MCT dataset with ground truth data for local tracking.

Keywords: Viewpoint invariance, person re-identification, global multi-object tracking, non-overlapping cameras

This work was supported by the ICT R&D Program of MSIP/IITP (Grant No. B0101-15-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis) and the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2016M3C7A1905477, NRF-2014M3C7A1046050).

https://doi.org/10.3837/tiis.2017.04.014 **ISSN : 1976-7277**

1. Introduction

With the purpose of security, e.g., for detecting and prosecuting crimes, the number of cameras used for video surveillance is on an increase trend. Consequently, the demand of the algorithms for intelligent video analysis is increasing because they can enhance the efficiency and the effectiveness of video surveillance and support for crime prevention. Person re-identification is the task to recognize a pedestrian who has been observed over a network of cameras with non-overlapping views. It is an important problem because it can make algorithms for intelligent video analysis applied to a single camera operate in non-overlapping multiple cameras.

Although the interest in person re-identification is increasing and many proposals for person re-identification have been proposed [22–31], no method that has satisfactory performance has yet been proposed. This is mainly due to the difficulties and challenges of person re-identification operating on a camera network with non-overlapping views. In particular, there are significant viewpoint, pose, and illumination changes between non-overlapping cameras, and they make person re-identification very difficult. Many existing approaches take a patch-based method to re-identify persons invariantly to pose and viewpoint [1–4]. These approaches divide pedestrian images into patches and extract features from each patch, then the similarity between two images is computed using the similarity between the features of the patches. These approaches have the pose and viewpoint invariant property because these approaches are based on the local features of the local patches instead of the global features. Nonetheless, these approaches are not viewpoint invariant enough since they can still give wrong matching results when images of different pedestrians observed from different viewpoints have similar appearances.

In this work, we propose a viewpoint invariant method for person re-identification using orientation of pedestrian. Our proposed method enhances the viewpoint invariance of the patch-based method by using the orientation of the pedestrian in the image. Assuming that a person body is the cylindrical shape, the proposed method estimates angle of each patch based on the orientation of the pedestrian. Then, patches have cylindrical coordinates and the similarity between patches is computed not only based on features but also based on angles. The proposed method shows the viewpoint invariant property by matching pedestrian images based on exact coordinate of each local feature.

Furthermore, we deal with a practical problem of applying the person re-identification to inter-camera multi-object tracking. In this work, the term *global multi-object tracking* is defined to denote the tracking of moving objects across multiple camera views in a camera network. In contrast, the term *local multi- object tracking* is defined to denote the tracking of moving objects in a single camera view. Person re-identification algorithm makes local multi-object trackers (where each operates only for its designated camera) to be a global multi-object tracker (that operates for its camera network) by connecting local trajectories of the same objects from local multiple object trackers to make one global trajectory. In order to do this, we need to solve some issues that have not been dealt with in existing methods because most existing methods consider still image dataset, but not video data in real scenarios. In order to do this, we need to deal with two issues. The first one is to select pedestrian images that are good for person re-identification among bounding boxes from the local trackers, which is referred to as a sample selection problem. Suitable sample selection method is needed because it is beneficial to use very useful some pedestrian images from all frames obtained by

as local tracking results so as to reduce the processing time. The second one is to determine whether the object observed in a camera network has ever appeared before. This problem is called the novelty detection.

In this work, we propose a framework (for inter-camera multi- object tracking using person re-identification) to handle the two problems. For the sample selection, we compute confidence of each pedestrian image based on the ratio of three body parts consisting of head, upper body and lower body. We used asymmetric axes in [5] to divide the three body parts from a pedestrian image. Then, pedestrian images with high confidence for person re-identification are selected. For the novelty detection, we use two Gaussian distributions [5] which are computed based on the similarity of false matching and correct matching. When the matching result of an object is given, if the probability of false matching is higher than the probability of correct matching, the object is regarded as the novel object.

To this end, the main contributions of this work are in the following two aspects. The first one is to devise a simple but effective viewpoint invariant person re-identification method using the orientation of each pedestrian to exploit spatial location of each extracted feature on a 3D body model. The second one is to propose a framework to handle global multi-object tracking for multiple disjoint cameras with non-overlapping camera views by adopting the proposed viewpoint invariant person re-identification approach facilitated by the novelty detection and sample selection.

This work is organized as follows. Related work in the field of person re-identification is introduced in Section 2. Section 3 describes the overall procedure and the detail of the proposed viewpoint invariant person re-identification method. Then, Section 4 shows the global multi-target tracking based on the person re-identification approach. Experimental results that demonstrate the effectiveness of the proposed methods are reported in Section 5. Finally, conclusions and future work are discussed in Section 6.

2. Related Work on Person Re-Identification

Most existing person re-identification methods [4–11] take the appearance-based approach by using only images of an object taken from cameras. On the other hand, some other methods [12–14] utilize not only the appearance but also spatial and temporal relationships between cameras to improve the performance. The methods first learn spatial and temporal relationships between cameras, and then the relationship information is used to predict the location (on the camera network) where objects reappear. These methods can distinguish objects that it is difficult to be discriminated from the other objects when appearance is solely used, but there is an issue of how to learn spatial and temporal information between cameras.

2.1 Spatial and Temporal Relationship-based Approaches

In some existing methods adopting spatial and temporal information between cameras, object tracking is used to model spatial and temporal relationship between cameras [12, 13]. To describe a probability that an object appears with respect to time and position, a probabilistic model between camera views is constructed. Also, illumination changes between non-overlapping camera views can be overcome using brightness transfer function (BTF). By clustering start/end points of object tracking, entry/exit zones are found [15] and visible links and invisible links are estimated [12, 16]. Conversely, relationships between cameras can be also obtained without object tracking [14]. First, patterns of pixels over time are analyzed. Then, camera views are divided into regions, and connections between regions are estimated based on the pattern analysis.

2.2 Appearance-based Approaches

Appearance-based methods can be divided into (i) those that utilize novel feature descriptors considering special challenges of person re-identification and (ii) those that deal with learning algorithms for the distance metric. Appearance-based methods basically use the similarity between feature descriptors extracted from images of objects for person re-identification. In order to compute the similarity between two person images, feature descriptors are computed from the images, and then the similarity between feature descriptors is calculated using an appropriate distance metric. While the feature descriptor-based approaches focus on constructing novel feature descriptors that are useful for person re-identification, the distance learning-based approaches focus on learning to make distance metric more robust to environmental changes between non-overlapping camera views.

The assumption that a pixel being more far away from symmetric axes of a person has a higher chance of belonging to background is used in [5]. First, by comparing pixels in foreground, two horizontal asymmetric axes (dividing a body of a person into head, upper body and lower body) are extracted. Then, two vertical symmetric axes of upper body and lower body are estimated. Next, a pixel that is closer to symmetric axes has a higher weight value when features are extracted from an image. For the feature descriptor, weighted HSV color histogram [5], maximally stable color regions (MSCR) [17] and recurrent high-structured patches (RSHP) [5] are used. Body parts are detected using training data in [6]. Then, appearances are compared based on the body parts. Besides the color histogram that is the most widely used feature, various features are used for person re-identification. For example, Gabor filter is used in [7], and Haar-like feature and dominant color descriptor are used in [8]. The patch match method is used in [4]. The patch match method divides a pedestrian image into partially overlapped patches and matches feature of patches. Salience of each patch is additionally estimated and patch match results are weighted based on saliences of patches in [4]. Salience represents how useful a patch is for person re-identification, and salience of each patch is learned using the k-nearest neighbor (KNN) or one-class support vector machine (OCSVM). We use the patch match method without the salience learning and additionally use angular coordinates of patches for the invariance to the viewpoint change.

Distance metric learning approaches train distance metric to make intra-class distance lower than inter-class distance instead of applying equal metric to all values of a feature descriptor. Existing metric learning methods include large margin nearest neighbor with rejection (LMNN-R) [9], probabilistic relative distance comparison (PRDC) [10], RankSVM [11].

Discussion: Unlike the existing methods, for ensuring viewpoint invariance property, the proposed method in this work compares local feature descriptors considering the position of local feature descriptors. To do it, persons in images are represented by a cylindrical model, and cylindrical coordinates of local features are then used in computing similarity between local features.

3. Proposed Viewpoint Invariant Person Re-Identification

3.1 Overview of the Proposed Method

Appearance-based person re-identification system receives a pedestrian image to be identified and outputs a ranked list by computing similarity between the input image and an image list consisting of images of known persons. The person image needing to be identified is called the probe. The image list consisting of images of known persons is called the gallery. To do person re-identification, foreground is segmented from a pedestrian image. Next, the foreground is

divided into a set of patches which consists of partially overlapped patches. Then, a feature descriptor for each patch is made and an angle of each patch is estimated. The feature descriptors and angles of patches are used in computing the similarity between pedestrian images. Similarity between feature descriptors of different pedestrian images and difference between angles of different pedestrian images are computed. Then, patches are matched considering the similarity and difference at the same time. Patches are matched to patches with high similarity and low difference. The similarity between images is obtained by averaging the similarities of matched patches. Finally, the ranked list is made by sorting similarities in the descending order. The overall procedure is depicted in **Fig. 1**.

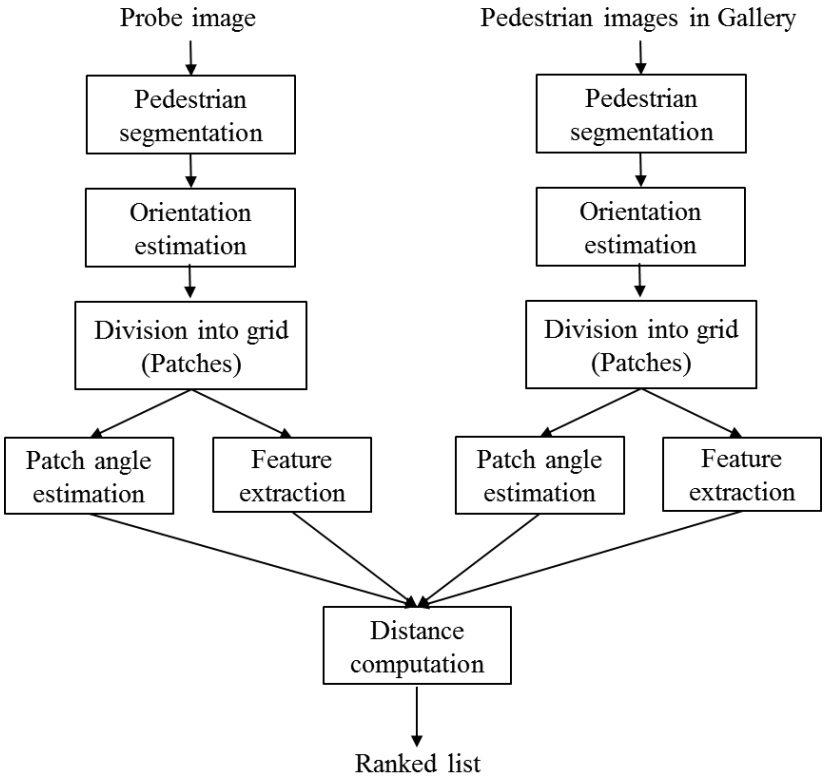

Fig. 1. Overall procedure of the proposed method

3.2 Pedestrian Segmentation

Generally, pedestrian images for person re-identification are rectangles and include background. For exact matching, it is required to segment pixels that belong to a pedestrian. Deep decompositional network (DDN) [18] can be used for this purpose. DDN parses a pedestrian in a still image using deep learning and it includes a phase distinguishing a person and a background. Despite its promising performance, we did not adopt this due to its high time complexity in computation. In a video, a pedestrian can be distinguished using background subtraction and object detection. For this purpose, we use the mean-shift based background subtraction method [10] due to its efficiency in terms of time. **Fig. 2** shows an example of the background subtraction result on the video.

(a) Original image (b) Modeled background (c) Background subtraction result

Fig. 2. Background subtraction result

3.3 Orientation Estimation

The proposed method uses an orientation of a pedestrian image. The orientation of the pedestrian is classified into 8 classes including 0, 45, 90, 135, 180, 225, 270 and 315 degrees. In a video, we can get the orientation using two points on a trajectory of the pedestrian. The direction of movement of the pedestrian is the orientation in the image under the assumption that the person goes forward. In a still image, the orientation can be obtained by an existing method. A pedestrian image is classified into 8 orientations using histogram of gradient (HoG) in [19]. The assumption throughout this work is that orientations of pedestrians are given for still images.

3.4 Division into Grid

A pedestrian image is divided into a set of grid cells. Each grid cell is called a patch, and grid cells are partially overlapped. In this work, the grid step is 4 and the size of the patch is 10×10. Then, an angle from an orientation vector of a pedestrian is estimated and a feature descriptor is constructed for each local patch.

3.5 Patch Angle Estimation

An angular coordinate is an angle from the predefined reference direction. In this work, a reference direction is an orientation of a pedestrian in an image. An angular coordinate of each local patch is estimated as follows. We know a coordinate of each pixel on an image. Also, we can compute the horizontal center of a segmented pedestrian along y-axis and the horizontal distance between the center and the pixel. Thus, we can compute an angular coordinate of each pixel by substituting a distance from the center into a circle equation given as

$$x^2 + y^2 = 1. \tag{1}$$

Then, we convert Cartesian coordinate to polar coordinate using

$$\theta = \begin{cases} arctan\dfrac{y}{x} & \text{if } x > 0 \\ arctan\dfrac{y}{x} + \pi & \text{if } x < 0 \text{ and } y \geq 0 \\ arctan\dfrac{y}{x} - \pi & \text{if } x < 0 \text{ and } y < 0 \\ \dfrac{\pi}{2} & \text{if } x = 0 \text{ and } y > 0 \\ -\dfrac{\pi}{2} & \text{if } x = 0 \text{ and } y < 0. \end{cases} \quad (2)$$

Fig. 3 shows the examples of pedestrian segmentation, division into grid, and patch angle estimation. The angular coordinate of each patch is presented using color. More details of the patch angle estimation are depicted in **Fig. 4**.

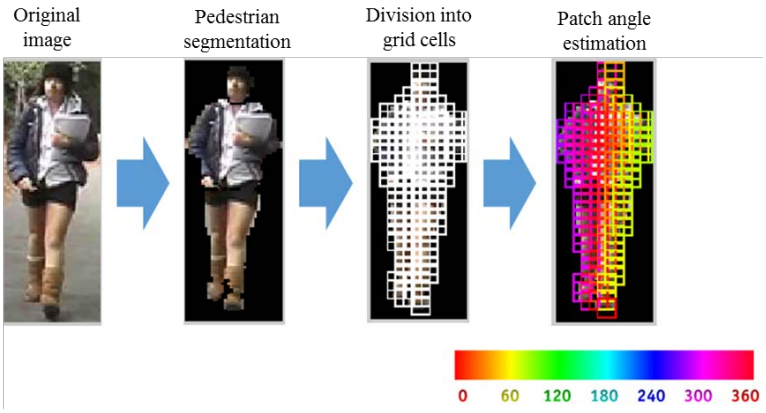

Fig. 3. Examples of pedestrian segmentation, division into grid cells and patch angle estimation

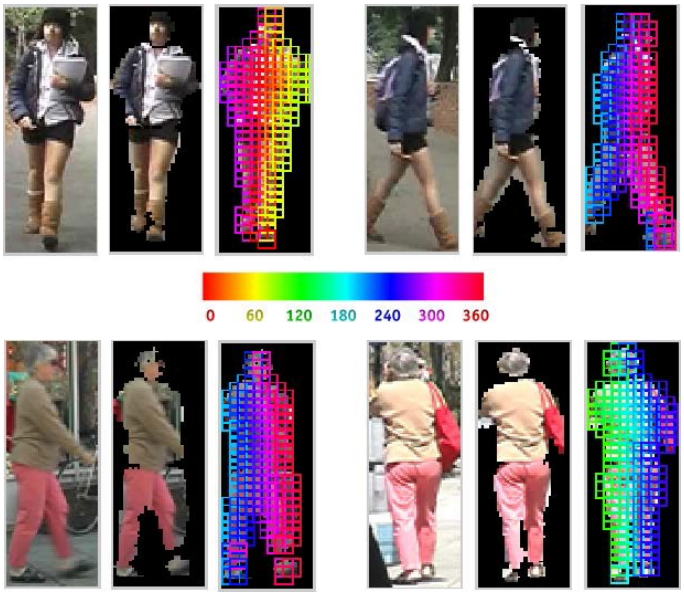

Fig. 4. Examples of detailed patch angle estimation

3.6 Feature Extraction

For each patch, color features are extracted as follows. We use both HSV and LAB color spaces. The RGB image is converted to HSV and LAB images. Next, for each color channel, histograms are constructed, normalized and concatenated to build one feature vector. In addition, SIFT is also used as a local feature due to its scale invariance property. However, SIFT feature is not added to the feature vector consisting of color histograms. Because SIFT does not match the proposed method using the orientation according to results of our experiments. But when SIFT was combined with the proposed method using the orientation by the way explained below, it improved the accuracy. SIFT is used for the method that are the same as the proposed method but don't use an orientation of a person. Thus, the result of the person re-identification by SIFT are obtained separately. Then, two ranked lists that are results by color histograms and SIFT are combined.

3.7 Distance Computation

Distances between feature descriptors made from local patches are calculated to compute a similarity between two pedestrian images. Distance between features is computed by

$$dist_{feat} = \sqrt{1 - \sum_{x \in X} \sqrt{feat(x)\, feat'(x)}}. \tag{3}$$

Distance between two cylindrical coordinates is computed by

$$dist_{cylind} = \sqrt{\rho^2 + \rho'^2 - 2\rho\rho' \cos(\varphi - \varphi') + (z - z')^2}. \tag{4}$$

Based on this, difference between two angular coordinates is computed by

$$diff_{angle} = \sqrt{2 - 2\cos(angle - angle')} / 2, \tag{5}$$

where $angle'$ and $angle'$ are angular coordinates of the two patches.

Local patches from different pedestrian images are matched to compute a similarity between pedestrian images [4]. A local patch in an image is basically matched to another patch that has the smallest distance. In the matching for the distance calculation, not all pair of local patches are used. The local patches that are located vertically adjacent are matched in which there are no horizontal constraints. This is because of an assumption that the pose variation of a person is significant with respect to the horizontal direction but not significant with respect to the vertical direction. In addition, difference between angular coordinates is considered for viewpoint invariant matching. Local patches that are horizontally closer on the person body have lower distance by considering angular coordinates. Thus, it is able to match local patches based on their positions on the body. Distance between image I and image is computed by averaging distance between patches of I and I' from the results of the patch matching. When images are divided into $M \times N$ grid cells, distance between the patch on (m, n) and image I' is computed by

$$dist_{patch}\left(p_{m,n}, I'\right) = \arg\min_{dist_{feat}} \left(\frac{diff_{angle}\left(p_{m,n}, p_{i,j}\right) + dist_{feat}\left(p_{m,n}, p_{i,j}\right)}{2} \right),$$

$$i = \max(1, m-h), \ldots, \min(M, m+), j = 1, \ldots, N, \tag{6}$$

where $p_{i,j}$ is the patch on the image I' and h is the range of the vertical constraint.

The similarities from patch matching are averaged and the resulting value is used to compute the similarity between images. Finally, the similarity between two images is computed as

$$sim_{img}(I, I') = e^{-\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} dist_{patch}(p_{m,n}, I')\right)}. \tag{7}$$

The similarities between the probe pedestrian image and the pedestrian images in the gallery are computed and then the gallery is sorted based on the similarities. The results are called the ranked list being an output of person re-identification.

4. Global Multi-Object Tracking using Person Re-Identification

4.1 Overall Procedure

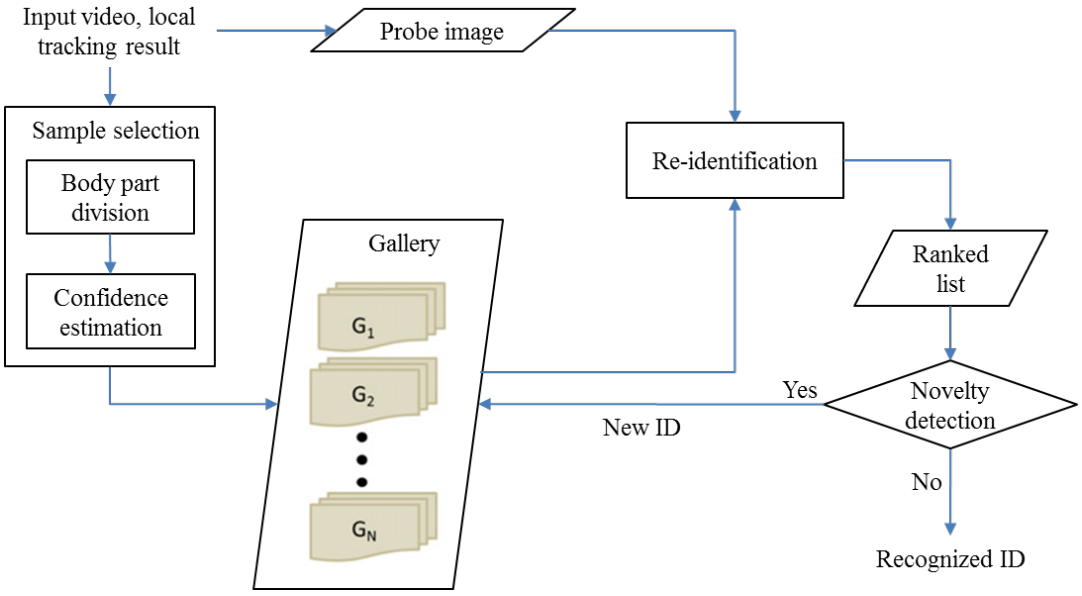

Fig. 5. Overall procedure of global multi-object tracking

Global multi-object tracking is possible by matching the outputs of local object trackers using person re-identification. Local object trackers track multiple objects on the single camera. Person re-identification extends the range of single camera object tracking into the camera network by matching outputs of object trackers. When object trackers detect a new incomming object in the single camera, that object is determined as a novel object (that has not been seen yet) for the camera network or a known object by person re-identification. If an object appear in the camera network at the first time, the object is enrolled in the gallery and receives the new identification (ID) number. If an object has appeared before, the object receives an existing ID

number in the gallery. Also, sample selection is running for each frame to gather suitable samples that are good for re-identification. The confidence levels for all bounding boxes from local object tracking are computed and pedestrian images with high confidence levels are saved. The overall procedure is depicted in **Fig. 5**. Any multi-object tracker that can give bounding boxes of pedestrians can be used as the local object tracker because the proposed person re-identification uses only pedestrian images. In this work, we use the ground truth for the experiments.

4.2 Novelty Detection

For novelty detection, we need a training set including the ground truth having ID numbers of persons. Two Gaussian distributions of similarities from correct matching and wrong matching are modeled using the training set as used in [5]. An object detected by a local object tracker can be determined whether it is a novel object based on the similarity of their matching results. If the probability of wrong matching is higher than the probability of correct matching, it is regarded as the novel object. Conversely, if the probability of correct matching is higher, it is re-identified by receiving ID number enrolled in the gallery. **Fig. 6** is the example of two Gaussian distribution constructed using a training dataset.

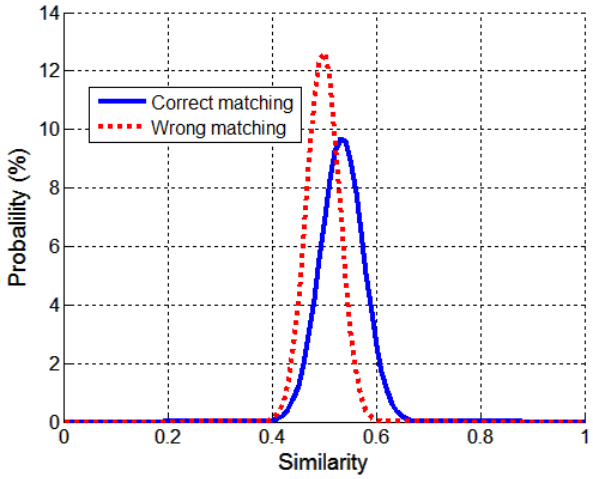

Fig. 6. Two Gaussian distributions of similarity

4.3 Sample Selection

For sample selection, we estimates the confidence level of a pedestrian image based on a ratio of body parts. To divide the pedestrian body parts, we use a silhouette partition method proposed in [5]. In [5], the chromatic bilateral operator and spatial covering operator are defined for silhouette partition. The chromatic bilateral operator is defined as

$$C(i,\delta) \propto \sum_{B_{[i-\delta,i+\delta]}} \left(d^2(p_i, \hat{p}_i) \right), \tag{8}$$

where δ is a vertical range, and p_i and $\hat{p}_i$ are pixels located symmetrically to height i. This means the summation of differences of pixel intensities that are vertically symmetric. The spatial covering operator is defined as

$$S(i,\delta)=\frac{1}{J\delta}|A(B_{[i-\delta,i]})-A(B_{[i,i+\delta]})|, \tag{9}$$

where J is the width and A is foreground area. This means difference of foreground areas. A vertical axis dividing a pedestrian image into the head and the body is estimated by

$$i_{HT}=\arg\min_{i}(-S(i,\delta)). \tag{10}$$

A vertical axis dividing the body into the upper body and the lower body is estimated by

$$i_{TL}=\arg\min_{i}((1-C(i,\delta))+S(i,\delta)). \tag{11}$$

Then, the pedestrian image is divided into three body parts including head, upper body and lower body by these two axes and the confidence level is estimated using the ratio of the three body parts. If the body is well partitioned, the image has high confidence. Finally, well-segmented pedestrian image can be guaranteed through the sample selection. **Fig. 7** represents an example of the results of the body partition.

(**a**) (**b**)

Fig. 7. An example of the sample selection: (a) has a higher confidence level than (b)

5. Experimental Results

5.1 Viewpoint Invariant Person Re-Identification

To evaluate the performance of the proposed viewpoint invariant person re-identification method, we used cumulative matching characteristic (CMC) [5]. CMC consists of accuracy at each rank. Accuracy at any rank is estimated by counting the number of correct matched images existing in any rank in the ranked list. Thus, CMC is an increasing function with respect to the rank. Additionally, the normalized area under the curve (nAUC) [5] is computed by normalizing the area under the CMC curve which represents overall performance with respect to the rank.

We use Viewpoint Invariant Pedestrian Recognition (VIPeR) dataset [20] to evaluate the accuracy of the proposed method. VIPeR dataset includes 632 pedestrian images, where for

each pedestrian two images captured from different cameras are included. The size of images is 128 x 48. It has significant illumination, pose and viewpoint changes. Some examples of VIPeR dataset are presented in **Fig. 8**. Among 632 image pairs, randomly selected 316 image pairs are used to compute CMC, and this is iterated 10 times. One of each image pair is included in the probe set and the other is included in the gallery.

Fig. 8. Examples of VIPeR dataset

In this work, for comparative study, we compared the proposed method with the five other methods. One is the method having the same procedure of the proposed method but without orientation. This is called Patch Match (PM) in this work. The other is Symmetry-Driven Accumulation of Local Features (SDALF) [5]. SDALF uses two color features and one pattern feature that are extracted or weighted based on symmetric body axes. Other three methods introduced in [4] were also used for comparison and they are called Sal_PM, Sal_KNN and Sal_OCSVM in this work: 1) Sal_PM is patch match method not using salience learning, 2) Sal_KNN is a method using KNN for salience learning and 3) Sal_OCSVM is a method using OCSVM for salience learning. Six methods are iteratively tested for 10 times by using the same randomly picked subsets of VIPeR dataset.

The performance was depicted in **Table 1** (where the boldface in columns indicates it achieved the best performance) and **Fig. 9**. From the results, we can see that the proposed method achieved Rank-1 accuracy of 24.49% and nAUC of 88.95% in **Table 1**. Compared to the other methods, the proposed method achieved the best performance: 1) compared to SDALF it achieved increased accuracy of 9.14% on the rank 1 and increased nAUC of 2.67%, 2) compared to Sal_PM, it achieved increased accuracy of 6.69% on the rank 1 and increased

nAUC of 2.52%, 3) compared to Sal_KNN, it achieved increased accuracy of 2.53% on the rank 1 and increased nAUC of 0.69%, 4) compared to Sal_OCSVM, it achieved increased accuracy of 2.12% on the rank 1 and increased nAUC of 0.67%, and 5) compared to PM, it achieved increased accuracy of 0.95% on the rank 1 and increased nAUC of 0.97%.

Table 1. Performance in terms of accuracy (%) and nAUC (%)

Method	Rank 1	Rank 5	Rank 10	Rank 20	Rank 50	Rank 100	nAUC
SDALF	15.35	35.09	44.46	57.37	73.89	85.66	86.28
Sal_PM	18.80	37.59	48.80	60.92	75.66	84.62	86.43
Sal_KNN	21.96	40.85	51.77	62.22	76.77	86.46	88.26
Sal_OCSVM	22.37	42.91	52.97	64.11	77.53	86.14	88.28
PM	23.54	43.67	54.62	64.87	**79.33**	85.03	87.98
Proposed Method	**24.49**	**45.44**	**57.06**	**66.80**	79.02	**86.27**	**88.95**

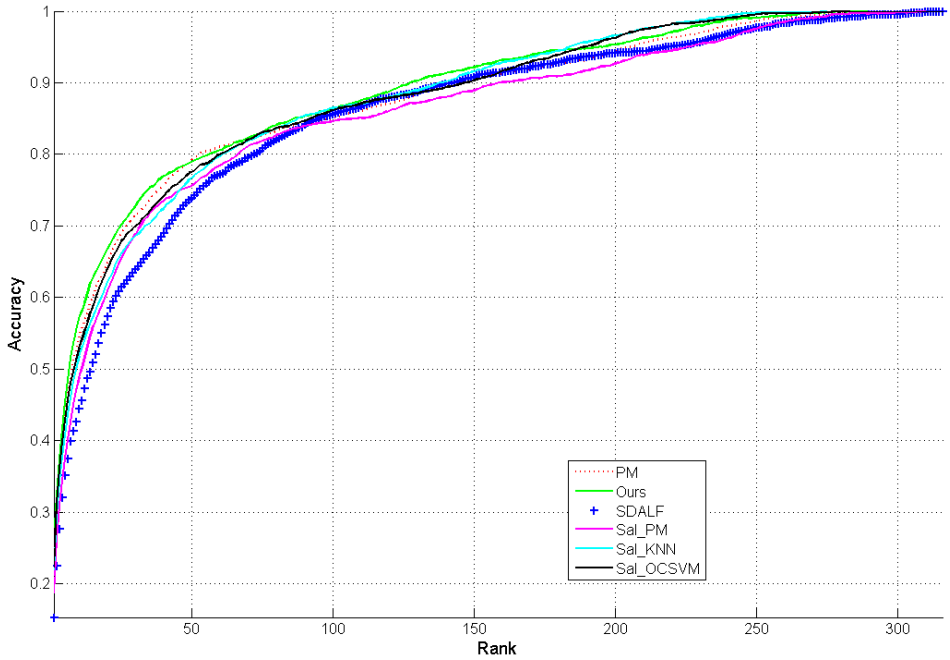

Fig. 9. Performance in terms of nAUC.

Table 2 shows the performance according to the gallery size indicating the number of pedestrians to be compared. The gallery size gives the significant effect to the performance because as the gallery size increases the complexity for similarity comparisons also increases.

Experiments for various gallery size were implemented using the random subset of VIPeR dataset. Also, each experiment is iterated 10 times. As we can see from **Table 2**, the performance is still not good although the gallery size is small. This is because VIPeR dataset includes image pairs with very serious illumination variations. Thus, for the case, the appearance-based model has some limitations as the other general methods.

Table 2. Performance (%) according to the different gallery sizes.

	10	20	30	50	100
Proposed Method	66.47	56.70	51.50	44.83	35.94

5.2 Global Multi-Object Tracking using Person Re-Identification

NLPR MCT dataset [21] consists of videos from non-overlapping camera views. This dataset consists of four sets of videos from different camera networks. Datasets 1 and 2 were from the same camera network consisting of three non-overlapping camera views, and datasets 3, 4 were from different camera networks. Datasets 1 and 2 were obtained from three camera views, dataset 3 was obtained from four camera views and dataset 4 was recorded from five cameras. Each dataset has different illumination, pose and viewpoint variations. All videos of each dataset were synchronized. Thus, all videos begin and end at the same time. Also, ground truth including positions, sizes for every frame and identification number of each pedestrian that appears in the videos is provided. It can be used for evaluation of multi-camera multi-object object tracking.

We used datasets 1 and 2 to evaluate the proposed method. The layout and the field of view of datasets 1 and 2 are depicted in **Fig. 10** and **Fig. 11**, and specification is described in **Table 3**. Dataset 1 was used to learn a threshold for the novelty detection and dataset 2 was used as the test dataset.

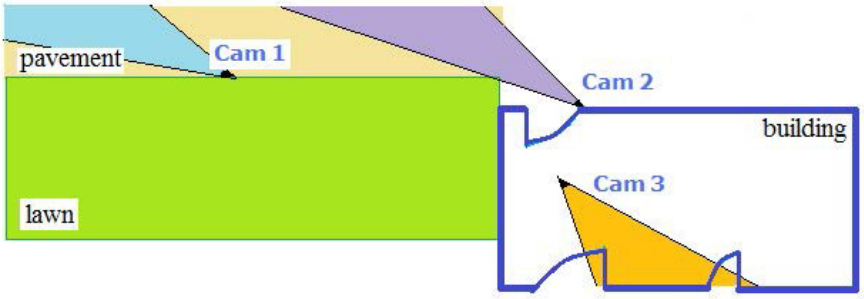

Fig. 10. The layout of three cameras [21]

Fig. 11. The field of view of each camera [21]

The experimental results are shown in **Table 4**. The performance of global multi-object tracking is evaluated in two cases. Novelty detection is to recognize an object appearing in the camera network for the first time. Re-identification is to match an object in the gallery that is not a novel object. In **Table 4**, we compute the novelty detection rate and re-identification rate, respectively. Also, we noted overall performance meaning the performance with the two cases

of novelty detection and re-identification. We compared the proposed method with SDALF. As we can see from **Table 4**, by increasing both the performance of novelty detection (up to 24.4%) and re-identification (up to 10.46%), the overall performance is increased to 19.10%.

Table 3. Specification of the NKPR MCT dataset

	Dataset 1	Dataset 2
Duration	20 min	20 min
Resolution	320×240	320×240
Frame rate	20 fps	20 fps
Number of persons	235	255
Format	avi	avi
Codec	MPEG-4(Xvid)	MPEG-4(Xvid)

Table 4. Performance in terms of accuracy (%) on the NLPR MCT dataset

Method	Novelty Detection	Re-Identification	Overall
SDALF	58.00	33.33	48.64
Proposed Method	82.40	43.79	67.74

6. Conclusions and Future Work

In this work, we proposed the viewpoint invariant person re-identification method. The proposed method uses angular coordinates of local patches in a pedestrian image to be matched with local patches in the other pedestrian image. Concerning difference between angular coordinates of patches matched in the images, the similarity of patches is compared based on the position on the body. In order to this, pedestrian segmentation is done first and pedestrian orientation is estimated. Then, the proposed method divides pedestrian images into locally overlapping patches and the angular coordinate is estimated for each patch. Finally, patch matching is carried out by considering difference between angular coordinates to compute the similarity between different pedestrian images. We also applied the proposed method to a practical application of global multi-object tracking across non-overlapping cameras. Especially, for realizing a global multi-object tracking system the two main difficulties, sample selection and the novelty detection, are properly handled. Samples are selected when they have high confidence that is computed using the ratio of body parts. The novelty detection is implemented based on two Gaussian distributions with respect to the similarity between pedestrian images. Two Gaussian distribution are constructed using similarities of correct matching and wrong matching on the training set. We compared the proposed method with some existing methods to verify the significance of this work. The experiments performed on the VIPeR dataset (for person re-identification) and NLPR MCT dataset (for global multi-object tracking). From the experimental study, we verified that proposed method could make significant improvements compared to the existing methods in terms of the accuracy. In this work, for person re-identification, LAB color histogram, HSV color histogram and SIFT features are used. We finally acknowledge that the impetus of this work is to verify the effectiveness of adoption of the estimation of the orientation of each pedestrian to spatially localize each extracted feature but not to compare our proposal with

state-of-the-art approaches such as [26, 28, 30], where the spatial location of a cylindrical model is exploited during the feature matching and distance computation.

As the future work, more various and sophisticated features with distinctive properties can be incorporated to the features to increase the accuracy. Also, dimension reduction algorithm can be used for better time complexity by reducing dimension of features. Since the proposed method is based on the patch-based approach, other patch-based methods (e.g., distance metric learning) can be combined with the proposed method. For multi-camera multi-object tracking, there is the issues of how to obtain the training set to estimate a threshold for the novelty detection. Thus, the methods to compute a more generalized threshold and to gather robust training set automatically can be researched for improving the performance of the novelty detection. Finally, exhaustive comparative study with other state-of-the-art approaches is essential by adopting such traits in our proposed framework.

References

[1] S Bak, E Corvee, F Brémond, M. Thonnat, "Multiple-shot human re-identification by mean riemannian covariance grid," in *Proc. of IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 179–184, 2011. Article (CrossRef Link)

[2] S Bak, S Zaidenberg, B Boulay, F Bremond, "Improving person re-identification by viewpoint cues," in *Proc. of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 175–180, 2014. Article (CrossRef Link)

[3] Z Wu, Y Li, RJ Radke, "Viewpoint invariant human re-identification in camera networks using pose priors and subject- discriminative features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1095–1108, 2015. Article (CrossRef Link)

[4] R Zhao, W Ouyang, X Wang, "Unsupervised salience learning for person re-identification," in *Proc. of IEEE Conference on.Computer Vision and Pattern Recognition (CVPR)*, pp. 3586– 3593, 2013. Article (CrossRef Link)

[5] L Bazzani, M Cristani, V Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, pp. 130–144, 2013. Article (CrossRef Link)

[6] DS Cheng, M Cristani, M Stoppa, L Bazzani, V Murino, "Custom Pictorial Structures for Re-identification," in *Proc. of British Machine Vision Conference (BMVC)*, 2011. Article (CrossRef Link)

[7] B Ma, Y Su, FB Jurie, "A novel image representation for person re-identification and face verification," in *Proc. of British Machine Vision Conference (BMVC)*, 2012. Article (CrossRef Link)

[8] S Bak, E Corvee, F Brémond, M Thonnat, "Person re-identification using haar-based and dcd-based signature," in *Proc. of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–8, 2010. Article (CrossRef Link)

[9] M Dikmen, E Akbas, TS Huang, N Ahuja, "Pedestrian recognition with a learned metric," in *Proc. of Asian Conference on Computer Vision (ACCV)*, pp. 501–512, 2011. Article (CrossRef Link)

[10] WS Zheng, S Gong, T Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 653–668, 2013. Article (CrossRef Link)

[11] B Prosser, WS Zheng, S Gong, T Xiang, Q Mary, "Person Re-Identification by Support Vector Ranking," in *Proc. of British Machine Vision Conference (BMVC)*, 2010. Article (CrossRef Link)

[12] O Javed, K Shafique, Z Rasheed, M Shah, "Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding*, vol. 109, pp. 146–162, 2008. Article (CrossRef Link)

[13] D Makris, T Ellis, J Black, "Bridging the gaps between cameras," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. II–205, 2004. Article (CrossRef Link)
[14] CC Loy, T Xiang, S Gong, "Multi-camera activity correlation analysis," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1988–1995, 2009. Article (CrossRef Link)
[15] D Makris, T Ellis, "Automatic Learning of an Activity- Based Semantic Scene Model," in *Proc. of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2003. Article (CrossRef Link)
[16] D Makris, T Ellis, "Path detection in video surveillance," *Image and Vision Computing*, vol. 20, pp. 895–903, 2002. Article (CrossRef Link)
[17] PE Forssén, "Maximally stable colour regions for recognition and matching," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2007. Article (CrossRef Link)
[18] P Luo, X Wang, X Tang, "Pedestrian parsing via deep decompositional network," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 2648–2655, 2013. Article (CrossRef Link)
[19] D Baltieri, R Vezzani, R Cucchiara, "People orientation recognition by mixtures of wrapped distributions on random trees," in *Proc. of European Conference on Computer Vision (ECCV)*, pp. 270–283, 2012. Article (CrossRef Link)
[20] D Gray, S Brennan, H Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. of IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007.
[21] Multi-Camera Object Tracking Challenge. http://mct.idealtest.org/datasets.html (August 2014).
[22] T Gandhi, MM Trivedi, "Panoramic Appearance Map (PAM) for Multi-camera Based Person Re-identification," in *Proc. of IEEE International Conference on Video and Signal Based Surveillance*, pp. 78–82, 2006. Article (CrossRef Link)
[23] D Baltieri, R Vezzani, R Cucchiara, "Mapping Appearance Descriptors on 3D Body Models for People Re-identification," *International Journal of Computer Vision*, vol. 111, no 3, pp. 345–364, 2015. Article (CrossRef Link)
[24] R Vezzani, D Baltieri, R Cucchiara, "People reidentification in surveillance and forensics: a survey," *ACM Computing Surveys*, vol. 46, no. 2, pp. 29:1–29:37, 2013. Article (CrossRef Link)
[25] G Doretto , T Sebastian , P Tu, J Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp 127–151, June 2011. Article (CrossRef Link)
[26] S Liao, Y Hu, X Zhu, and SZ Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, June 7–12, Boston, Massachusetts, USA, 2015. Article (CrossRef Link)
[27] L Bazzani, M Cristani, A Perina, M Farenzena, and V Murino, "Multiple-shot person re-identification by HPE signature," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, 2010. Article (CrossRef Link)
[28] S Bak, R Kumar, and F Bremond, "Brownian descriptor: a rich meta-feature for appearance matching," in *Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 363–370, 2014. Article (CrossRef Link)
[29] T Wang, S Gong, X Zhu, and S Wang, "Person re-identification by video ranking, " in *Proc. of European Conference on Computer Vision (ECCV)*, 2014. Article (CrossRef Link)
[30] M Zeng, Z Wu, C Tian, L Zhang, and L Hu, "Efficient person re-identification by hybrid spatiogram and covariance descriptor," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, 2015. Article (CrossRef Link)
[31] M Hirzer, C Beleznai, PM Roth, and H Bischof, "Person re-identification by descriptive and discriminative classification," *Image Analysis*, 2011. Article (CrossRef Link)

Jeonghwan Gwak received Ph.D. degree from Gwangju Institute of Science and Technology (GIST), Gwangju, Korea in 2014. From 2002 to 2007, he had worked for several companies and research institutes as a researcher as well as a chief technician. From Sept. 2014 to Sept. 2016, he worked as a Postdoctoral researcher in GIST. Since Oct. 2016, he is working as a research assistant professor at School of Electrical Engineering and Computer Science (EECS) in GIST. His current research interests include deep learning, pattern recognition, computer vision, image and video processing/understanding, evolutionary computation and optimization, and relevant applications of surveillance systems and medical image analysis.

Geunpyo Park received B.S. degree in computer science from the University of Seoul, Korea and M.S. degree at School of Electrical Engineering and Computer Science (EECS) in Gwangju Institute of Science and Technology (GIST), Korea in 2014 and 2016, respectively. He is now working as a researcher in Mechatro, Inc., Seoul, Korea. His major research interests include multiple object tracking, computer vision and pattern recognition.

Moongu Jeon received the B.S. degree in architectural engineering from the Korea University, Seoul, Korea, in 1988 and the M.S. and Ph.D. degrees in computer science and scientific computation from the University of Minnesota, Minneapolis, MN, USA, in 1999 and 2001, respectively. As a Postgraduate Researcher, he worked on optimal control problems at the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2001--2003, and then moved to the National Research Council of Canada, where he worked on the sparse representation of high-dimensional data and the level set methods for image processing until July 2005. In 2005, he joined the Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a full Professor in the School of Information and Communications. His current research interests are in machine learning, computer vision, and intelligent transportation systems.