IEIE Transactions on Smart Processing and Computing, vol. 6, no. 4, August 2017
https://doi.org/10.5573/IEIESPC.2017.6.4.262

IEIE Transactions on Smart Processing and Computing

Person Re-identification using Sparse Representation with a Saliency-weighted Dictionary

Miri Kim, Jinbeum Jang, and Joonki Paik*

Graduate School of Advanced Image Science, Multimedia, and Film, Chung-Ang University / Seoul 06974, Korea

* Corresponding Author: Joonki Paik, paikj@cau.ac.kr

Received June 16, 2017; Revised July 8, 2017; Accepted July 14, 2017; Published August 30, 2017

** Short Paper*

***Abstract*:** Intelligent video surveillance systems have been developed to monitor global areas and find specific target objects using a large-scale database. However, person re-identification presents some challenges, such as pose change and occlusions. To solve the problems, this paper presents an improved person re-identification method using sparse representation and saliency-based dictionary construction. The proposed method consists of three parts: i) feature description based on salient colors and textures for dictionary elements, ii) orthogonal atom selection using cosine similarity to deal with pose and viewpoint change, and iii) measurement of reconstruction error to rank the gallery corresponding a probe object. The proposed method provides good performance, since robust descriptors used as a dictionary atom are generated by weighting some salient features, and dictionary atoms are selected by reducing excessive redundancy causing low accuracy. Therefore, the proposed method can be applied in a large scale–database surveillance system to search for a specific object.

***Keywords*:** Re-identification, Sparse representation, Saliency, Occlusion, Surveillance

1. Introduction

Surveillance systems for person identification have developed during the last few years from simple monitoring into intelligent video analysis with massive databases. From the original monitoring depending on decisions by a manager about important information, surveillance systems have developed into single camera–based analysis systems that detect and track objects using computer vision techniques. They identify space and time-constrained objects with low accuracy from a video captured by each camera. The camera network–based approach solves the space limitation problem using transmission of some messages between connected cameras. However, it is difficult to decide if something is the same object, when using two or more neighboring cameras, if the cameras do not share the same ground plane, and it is difficult to resolve time-independent object identification.

Person re-identification techniques have been studied to overcome spatio-temporal limitations on the basis of massive databases with objects detected by multiple cameras. They recognize the same objects although they appear in two or more cameras that have respectively different spaces and different time frames. For this reason, it is possible to monitor a global area for an overall period of time. However, re-identification presents many issues when it comes to high performance, as shown in Fig. 1. All cameras have different illuminance and resolution, and an object may be captured in multiple cameras, but the shape and pose are detected and displayed differently. In addition, occlusion sometimes causes poor identification accuracy.

There are two approaches to solving these problems: 1) feature description and 2) metric learning. A feature-based

(a) (b) (c) (d)

Fig. 1. Challenging problems of person re-identification (a) illumination, (b) low resolution, (c) pose change, (d) occlusion.

approach makes a descriptor using discriminative features from other objects [1-5]. Hamdoun *et al*. proposed speeded-up robust features (SURF)-based re-identification using appearance information [6]. Farenzena *et al*. proposed a feature extraction method to fuse three complementary aspects: chromatic information, structural information, and recurrent informative patches [7]. This method achieves robust results for pose changes, different viewpoints, and illumination changes. Zhao *et al*. re-identified a person using unsupervised learning to extract saliency about the person [8]. This method calculates the saliency of the patch using salience learning, and extracts discriminative and reliable features after performing patch matching with constraints in neighboring areas to solve the misalignment caused by changes of viewpoint and posture. Karanam *et al*. proposed dictionary learning that discriminatively and sparsely encodes features representing people. This method is suited to background clutter and occlusion [9]. Matsukawa *el al*. presented a descriptor combining color and texture by representing local features of an image in a hierarchical Gaussian distribution [10].

The metric learning–based approach obtains a proper measurement method using machine learning to calculate the distance between images obtained from other cameras [11-15]. Weinberger *el al*. presented a method that learns the Mahalanobis distance metric based on semidefinite program [16]. It clusters training data belonging to the same class. Sugiyama proposed a dimension-reduction method that preserves the local structure of data in the class and maximizes the separation between classes in order to embed multi-modal data well [17].

In this paper, we propose a sparse representation-based person re-identification method using a weighted-dictionary-by-saliency map. After generating a saliency map of an object representing discriminative features, the proposed method extracts descriptors based on a color histogram and a local binary pattern (LBP) in each object image. And then, some elements of each descriptor are weighted in the saliency map. The weighted descriptors are used by a dictionary of sparse representation. Finally, a sparse coefficient is estimated using a least absolute shrinkage and selection operator (LASSO) [18], determining identity with minimal residuals.

This paper is organized as follows. Section 2 introduces sparse representation theory and its application to person re-identification. In Section 3, we propose sparse representation with a saliency-weighted dictionary. Section 4 shows experimental results, and Section 5 concludes the paper.

2. Theoretical Background

In this paper, we solve the matching problem of person re-identification using sparse representation. Sparse representation is used to recover or reconstruct signals using a few atoms [19]. Generally, the signal reconstruction problem using sparsity is defined by linear combination as

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{\alpha}$ represents the weight assigned as an $N \times 1$ column vector, $\mathbf{D}$, a pre-defined $M \times N$ dictionary with N samples that have M-dimensional features, and $\mathbf{y}$, reconstructed signals the same size as $\boldsymbol{\alpha}$. To satisfy the sparsity, atoms of $\boldsymbol{\alpha}$ should have a lower non-zero value. In other words, $\mathbf{y}$ is generated by combining only a few dictionary elements selected by non-zero atoms of $\boldsymbol{\alpha}$. For this reason, given dictionary $\mathbf{D}$, the problem in (1) is modified to minimize the number of the non-zero atoms as follows:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_0 \tag{2}$$

where $\hat{\boldsymbol{\alpha}}$ denotes the modified sparse vector, λ is a regularization factor, $\|\cdot\|_2^2$ represents the l_2-norm operator, and $\|\cdot\|_0$ the l_0-norm operator. In (2), $\hat{\boldsymbol{\alpha}}$ ensures the optimal reconstruction as the first term, since the error between input signal $\mathbf{y}$ and recovered version $\mathbf{D}\boldsymbol{\alpha}$ is minimized. In addition, $\hat{\boldsymbol{\alpha}}$ satisfies the sparsity condition as the second term because of the computation of l_0-norm. In terms of person re-identification, sparse representation is used to measure the reconstruction error as a matching rate.

Sparse representation-based person re-identification methods have been researched for matching the error between target and reconstructed object. Khedher *et al*. presented a re-identification method using SURF matching-based sparse representation [20]. It represents target SURF by linearly combining a dictionary consisting of existing images without learning. Liu *et al*. used semi-supervised coupled dictionary learning to bridge the variations in object appearance between cameras [21]. This method learned a dictionary that minimizes the energy function of reconstruction error and locality penalty. Lisanti *et al*. proposed iterative re-weighted sparse ranking [22]. This method applied a novel soft- and hard-reweighting method to sparse representation. This aims to remove the weight of a coefficient that contributes little to reconstruct a given target. And the weight is redistributed to other coefficients. In addition, in order to find the best target, already ranked individuals are excluded in each iteration. Compared with other re-identification–based simple feature matching, sparse representation produces good results.

3. Re-identification using a Weighted Dictionary

The proposed re-identification consists of three parts: i) feature description based on salient colors and textures for dictionary elements, ii) orthogonal atom selection using cosine similarity to deal with pose and viewpoint change, and iii) measurement of reconstruction error to rank the

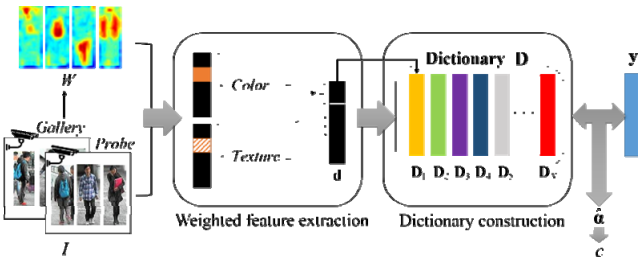

Fig. 2. Block diagram of the proposed method.

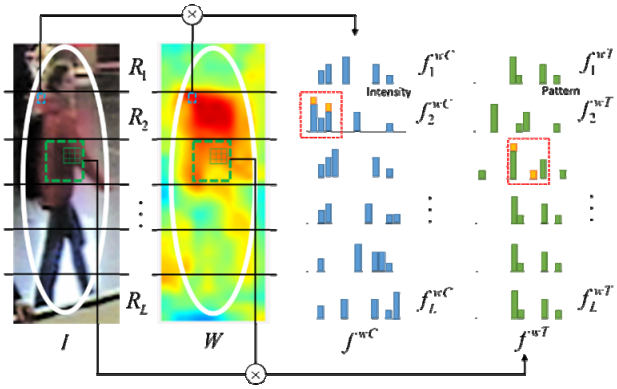

Fig. 3. Flow map of weighted feature description in the proposed method.

gallery corresponding to a probe object. Each part is described in the following subsections.

3.1 Weighted Feature Descriptor Generation using Saliency

Given some detected object images, the proposed re-identification method generates feature descriptors to construct a dictionary for sparse representation. Many feature description methods have been proposed to match the same objects. However, the size of an object is very small in an image, so fewer features are extracted, although each object has one or more discriminative features. To solve this problem, we extract a saliency map to make the weighted descriptor. Saliency means a relatively noticeable visual-perceptual characteristic of an object. Therefore, it provides a more accurate matching result between a probe and an object in a gallery set.

In this paper, we obtain the saliency map based on unsupervised learning proposed by Zhao *et al.* [8]. This method learns noticeable features of an object from all elements of the gallery and probe using k-nearest neighbor and one-class support vector machines. Since it uses dense correspondence using a L*a*b* color–based histogram and SIFT descriptor, the generated saliency map considers illumination and viewpoint change. For that reason, it represents the weight of each pixel.

Using the saliency map, the proposed method generates a weighted feature descriptor, as shown in Fig. 3. We first divide input image I and saliency map W into L local blocks after Gaussian windowing to exclude the background region. To match a probe with a gallery, a color histogram in the hue saturation value (HSV) space is first obtained for color descriptions from all blocks. Each histogram consists of some representative colors, but is not distinguished from others. To obtain the weighted color descriptor, we first make the hue channel histogram, f_s^C, from each stripe region, and compute the weight with the same size of the histogram as follows:

$$W_s^C(b) = \begin{cases} W_s^C(b) + W(\mathrm{u},\mathrm{v}), & \text{if } I_s^H(u,v) \in b \\ W_s^C(b), & otherwise \end{cases} \tag{3}$$

where I_s^H represents the hue channel of the s-th stripe, W_s^C is the weight vector, b is the bin of the histogram, and (u,v) the coordinates in each stripe. Therefore, each histogram bin is modified by multiplying it by the corresponding weight as follows:

$$f_s^{wC}(b) = W_s^C(b) f_s^C(b) \tag{4}$$

where f_s^{wC} is a weighted color histogram of the s-th stripe satisfying $f_s^{wC} \in \mathbb{R}^{1\times M_C}$ with the M_C histogram bins. Thus, each value of f_s^{wC} is weighted by the saliency map, as shown in the third column of Fig. 3.

Next, the proposed method generates a weighted texture descriptor using local binary pattern (LBP) [23]. Texture provides the object's internal pattern information to distinguish it from objects having similar color descriptors. For this reason, the weighted texture descriptor is computed in each stripe region in the same manner as the weighted color histogram, as follows:

$$f_s^{wT}\left(I_s^T(u,v)\right) = W_s^T\left(I_s^T(u,v)\right) f_s^T\left(I_s^T(u,v)\right) \tag{5}$$

where f_s^{wT} denotes the weighted texture descriptor of the s-th stripe satisfying $f_s^{wT} \in \mathbb{R}^{1\times M_T}$ with the M_T bins, I_s^T is the texture pattern computed by LBP at (u,v), f_s^T is the LBP histogram from each stripe region, and W_s^T is the weight vector at the same size as f_s^T obtained in the same manner as (3). Finally, total descriptor $\mathbf{d} \in \mathbb{R}^M$ of an object is defined by aligning all stripe regions as

$$\mathbf{d} = \left[f_1^{wC} \quad f_1^{wT} \quad f_2^{wC} \quad f_2^{wT} \quad \cdots \quad f_L^{wC} \quad f_L^{wT} \right]^{\mathrm{T}}. \tag{6}$$

3.2 Dictionary Construction

Given the weighted descriptors of all object images, the proposed method constructs a dictionary to deal with pose and occlusion. Existing re-identification methods make the dictionary using a descriptor of all images or random images or some consecutive images. But it requires extensive computation time because of large dimensions from the number of images. Also, performance is not constant for random images. In addition, the descriptor-based saliency [8] does not consider the object's pose with

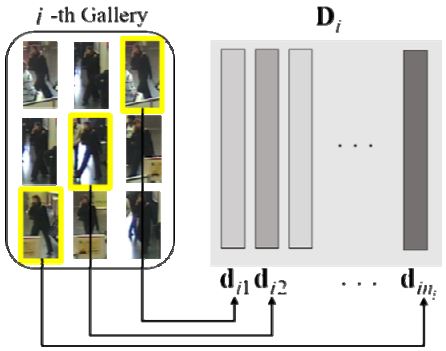

Fig. 4. Dictionary construction of the proposed method.

occlusion.

To construct the dictionary efficiently, the proposed method selects some images with different characteristics compiling multiple images of each object. The similarity function between images is defined as

$$CosSimilarity = \frac{\mathbf{d}_{ip} \cdot \mathbf{d}_{iq}}{\|\mathbf{d}_{ip}\|\|\mathbf{d}_{iq}\|}, \tag{7}$$

where $\mathbf{d}_{ij}$ represents the weighted feature descriptor of the i-th person's j-th image. In this method, a descriptor is picked up for a dictionary element that has the lowest similarity value on the basis of the first image randomly selected. This process is repeated until the dictionary elements for the number of predefined multi-shot images for all objects are picked up. Fig. 4 shows the dictionary construction result.

Consequently, i-th gallery $\mathbf{D}_i \in \mathbb{R}^{M\times n}$ is defined as

$$\mathbf{D}_i = [\mathbf{d}_{i1} \quad \mathbf{d}_{i2} \quad \cdots \quad \mathbf{d}_{in}], \tag{8}$$

where n represents the number of predefined multi-shot images. Then, dictionary $\mathbf{D} \in \mathbb{R}^{M\times K}$ is generated as

$$\mathbf{D} = [\mathbf{D}_1 \quad \mathbf{D}_2 \quad \cdots \quad \mathbf{D}_N], \tag{9}$$

where N represents the number of galleries, and $K = \sum_{i=1}^{N} n_i$ is the number of descriptors. Finally, the generated dictionary consists of various poses and occlusion images with salient features.

3.3 Coefficient Estimation and Identity Decision

To re-identify a probe from other galleries, the proposed method obtains a sparse vector and computes its reconstruction error. As mentioned above, sparse representation recovers or reconstructs a signal using fewer non-zero values. In the re-identification problem, it is used to measure the matching error for a gallery decision corresponding to the probe. To measure the reconstruction error, (1) is modified as

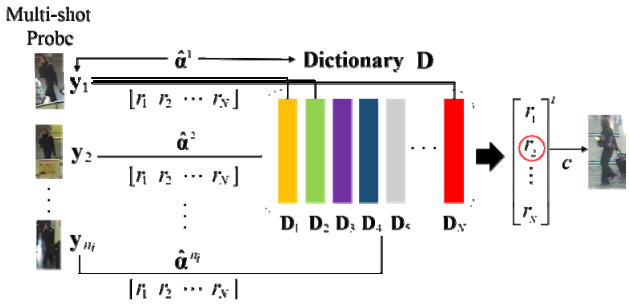

Fig. 5. Process of identity decision with the proposed method.

$$\hat{\mathbf{y}} \approx \mathbf{D}\hat{\boldsymbol{\alpha}}, \tag{10}$$

where $\hat{\mathbf{y}}$ represents the reconstructed probe satisfying

$$\hat{\mathbf{y}} \approx \mathbf{D}_1\hat{\boldsymbol{\alpha}}_1 + \mathbf{D}_2\hat{\boldsymbol{\alpha}}_2 + \cdots + \mathbf{D}_N\hat{\boldsymbol{\alpha}}_N, \tag{11}$$

and $\hat{\boldsymbol{\alpha}} = [\hat{\boldsymbol{\alpha}}_1 \quad \hat{\boldsymbol{\alpha}}_2 \quad \cdots \quad \hat{\boldsymbol{\alpha}}_N] \in \mathbb{R}^K$ is the estimated sparse vector of a probe. In (2), we modify l_0-norm to l_1-norm to solve the NP-hard problem. Therefore, (2) is also modified as follows:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}}(\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1) \tag{12}$$

In this paper, we resolve sparse representation using LASSO to find the optimal solution with efficient computation. Finally, the proposed method searches the optimal gallery using a reconstructed probe from the estimated coefficient and $\mathbf{D}$ from N gallery images. To obtain the reconstruction error of the gallery, we compute the differences between the reconstructed probe and the galleries, respectively. Fig. 5 shows the process of optimal gallery selection for a probe.

Since the proposed method also uses multi-shot images of a probe, one is selected as

$$r_i = \min_k \|\mathbf{y}_k - \mathbf{D}_i\hat{\boldsymbol{\alpha}}_i^k\|, \tag{13}$$

where r_i is the minimum reconstruction error of the i-th gallery for a probe, and $\mathbf{y}_k$ is the feature descriptor of the k-th image in a probe.

After that, an index of the gallery corresponding to the probe is determined as

$$c = \arg\min_i \{r_i\}, i = 1, 2, ..., N, \tag{14}$$

where c represents the selected identity of the probe to re-identify.

Table 1. Comparison of various feature descriptions.

Dataset	iLIDS-VID			
Rank	1	5	10	20
HSV	10%	24.6%	31.3%	41.3%
LBP	5.3%	17.3%	26%	38.6%
HSV+LBP	16.6%	34.6%	43.3%	50%
Weighted HSV	14%	28.6%	37.3%	48.6%
Weighted LBP	8.3%	21.3%	28.6%	40.6%
Proposed descriptor	**28.6%**	**47.3%**	**58.6%**	**66.6%**

4. Experimental Results

4.1 Datasets

In the experiment, we evaluated the proposed method with a public dataset: iLIDS-VID [24]. In the dataset, continuous images are captured from non-overlapping cameras. The iLIDS-VID dataset contains image sequences of 300 pedestrian pairs taken in a crowded airport arrival hall. Specifically, it has some challenges, such as background clutter, occlusion, plus viewpoint and illumination changes.

4.2 Evaluating the Weighted Feature Descriptor

To verify the performance of the weighted feature descriptor, we measured the reconstruction error using all galleries and the rank of the corresponding gallery of the probe image using iLIDS-VID. We compared it with various features, such as HSV, LBP, combined HSV with LBP, weighted HSV, and weighted LBP. As shown in Table 1, the proposed saliency-weighted feature description improved by more than 12% in rank 1. Moreover, the standard rank number is lower, and the rate of the indexed rank is higher than others.

4.3 Evaluating Dictionary Construction

In the second experiment, we evaluated the impact of the proposed dictionary construction method. In order to compare performance under the same conditions as the proposed method, we constructed dictionaries using some sequential images and atom selection at a specific interval. As shown in Table 2, the proposed method provided better performance than the others. This implies that a dictionary composed of images with the most different characteristics influences the estimation of sparse coefficients.

4.4 State-of-the-art Comparison

We compare the proposed method with existing person re-identification methods. As shown in Table 3, the proposed method outperforms the other methods. Although the sparse re-id (SRID) method has better performance than our method only in rank 20, the proposed method shows superior performance in ranks 1, 5, and 10.

Table 2. Comparison with dictionary construction of atom selection methods.

Dataset	iLIDS-VID			
Rank	1	5	10	20
Sequential	14.6%	34%	43.3%	60%
Stride selection	20%	42.6%	53.3%	65.3%
Proposed dictionary construction	28.6%	47.3%	58.6%	66.6%

Table 3. State-of-the-art comparison.

Dataset	iLIDS-VID			
Rank	1	5	10	20
ISR [22]	10%	24.6%	31.3%	41.3%
eSDC [8]	5.3%	17.3%	26%	38.6%
LC-KSVD [25]	24.3%	38.5%	42.3%	47.3%
DVR [24]	23.3%	42.4%	55.3%	68.4%
SRID [9]	24.9%	44.5%	54.1%	**68.8%**
Proposed method	**28.6%**	**47.3%**	**58.6%**	66.6%

5. Conclusion

This paper presents personal re-identification using saliency map-based weighted feature description and sparse representation based on orthogonal dictionary atom selection. Since the proposed method describes the features extracted from an object using the saliency map of the object, robust atoms are added to the dictionary. An additional contribution of this work is to form the dictionary by selecting gallery images with different characteristics using cosine similarity. The experimental results demonstrate the proposed method provides better performance than existing feature descriptors and constructions of a dictionary. As a result, the proposed method can accurately re-identify a person under challenging environments, such as pose variations and occlusion, when compared against state-of-the-art methods. Therefore, it can be applied to object retrieval across cameras and video summarization for the same object.

Acknowledgement

This work was partly supported by the ICT R&D program of MSIP/IITP [2017-0-00250, Intelligent Defense Boundary Surveillance Technology Using Collaborative Reinforced Learning of Embedded Edge Camera and Image Analysis] and by the Commercialization Promotion Agency for R&D Outcomes (COMPA) funded by the Ministry of Science, ICT and Future Planning (MSIP).

References

[1] N. Gheissari, T. B. Sebastian, and R. Hartley, “Person re-identification using spatiotemporal appearance,” *Proc. IEEE Conf. Computer Vision and*

Pattern Recognition, pp. 1528–1535, June 2006. Article (CrossRef Link)

[2] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," *Proc. European Conference on Computer Vision (ECCV)*, pp. 262-275, June 2008. Article (CrossRef Link)

[3] B. Ma, Y. Su, and F. Jurie, "Covariance descriptor based on bio-inspired features for person re-identification and face verification," *Image and Vision Computing*, vol. 32, no. 6, pp. 379–390, April 2014. Article (CrossRef Link)

[4] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," *Proc. European Conference on Computer Vision Workshops and Demonstrations*, pp. 413-422, 2012. Article (CrossRef Link)

[5] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 144-151, June 2014. Article (CrossRef Link)

[6] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," *Proc. IEEE Conf. Distributed Smart Cameras*, pp. 1–6, September 2008. Article (CrossRef Link)

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person Re-Identification by Symmetry-Driven Accumulation of Local Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2360-2367, June 2010. Article (CrossRef Link)

[8] R. Zhao, W. Ouyang, and X. Wang. "Unsupervised salience learning for person re-identification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3586-3593, June 2013. Article (CrossRef Link)

[9] S. Karanam, Y. Li, and RJ. Radke, "Sparse Re-Id: Block Sparsity for Person Re-Identification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 33-40, June 2015. Article (CrossRef Link)

[10] T. Matsukaw, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian Descriptor for Person Re-Identification", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1363-1372, June 2016. Article (CrossRef Link)

[11] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," *In Computer Vision–ACCV, Springer*, pp. 501–512, 2010, 2011. Article (CrossRef Link)

[12] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," *In Proc. Conf. Machine learning ACM*, pp. 209–216, June 2007. Article (CrossRef Link)

[13] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2288-2295, June 2012. Article (CrossRef Link)

[14] A. Mignon and F. Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2666-2672, June 2012. Article (CrossRef Link)

[15] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2197-2206, June 2015. Article (CrossRef Link)

[16] KQ. Weinberger, J. Blitzer, and LK. Saul, "Distance metric learning for large margin nearest neighbor classification," *Advances in neural information processing systems*, pp. 1473-1480, 2006. Article (CrossRef Link)

[17] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," *Proc. Conf. Machine learning ACM,* pp. 905-912, June 2006. Article (CrossRef Link)

[18] MJ. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using l1-constrained quadratic programming (Lasso)," *IEEE Trans. Information Theory*, vol. 55, no. 5, pp. 2183-2202, May 2009. Article (CrossRef Link)

[19] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, TS. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031-1044, June 2010. Article (CrossRef Link)

[20] MI. Khedher, MA. El. Yacoubi, and B. Dorizzi, "Multi-shot surf-based person re-identification via sparse representation," *IEEE Conf. In Advanced Video and Signal Based Surveillance (AVSS)*, pp. 159-164, August 2013. Article (CrossRef Link)

[21] X. Liu, M. Song, X. Zhou, C. Chen, and J. Bu "Semi-Supervised Coupled Dictionary Learning for Person Re-identification," pp. 3550-3557, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3550-3557, June 2014. Article (CrossRef Link)

[22] G. Lisanti, I. Masi, AD. Bagdanov, and A. Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1629-1642, November 2015. Article (CrossRef Link)

[23] T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, December 2006. Article (CrossRef Link)

[24] T. Wang, S. Gong, X. Zhu and S. Wang. "Person Re-Identification by Video Ranking," *Proc. European Conference on Computer Vision (ECCV)*, pp. 688-703, September 2014. Article (CrossRef Link)

[25] Z. Jiang, Z. Lin, LS. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651-2664, November 2013. Article (CrossRef Link)

Miri Kim was born in Busan, Korea, in 1992. She received a BSc in integrative engineering from Chung-Ang University, Korea, in 2016. Currently, she is pursuing an MSc in Image Engineering at Chung-Ang University. Her research interests include object tracking, person re-identification, and object detection.

Jinbeum Jang was born in Suwon, Korea, in 1989. He received a BSc in digital media from Sang-Myung University, Korea, in 2014. Also, he graduated with an MSc in image science from Chung-Ang University, Korea, in 2016. Currently, he is pursuing a PhD in Image Science at Chung-Ang University. His research interests include image enhancement, auto focusing, and depth generation.

Joonki Paik was born in Seoul, Korea, in 1960. He received a BSc in control and instrumentation engineering from Seoul National University in 1984. He received an MSc and a PhD in electrical engineering and computer science from Northwestern University in 1987 and 1990, respectively. From 1990 to 1993, he worked at Samsung Electronics, where he designed image stabilization chip sets for consumer camcorders. Since 1993, he has been on the faculty at Chung-Ang University, Seoul, Korea, where he is currently a professor in the Graduate School of Advanced Imaging Science, Multimedia, and Film. From 1999 to 2002, he was a visiting professor in the Department of Electrical and Computer Engineering at the University of Tennessee, Knoxville. Dr. Paik was a recipient of the Chester Sall Award from the IEEE Consumer Electronics Society, the Academic Award from the Institute of Electronic Engineers of Korea, and the Best Research Professor Award from Chung-Ang University. He has served the IEEE Consumer Electronics Society as a member of the editorial board. Since 2005, he has been the head of the National Research Laboratory in the field of image processing and intelligent systems. In 2008, he worked as a full-time technical consultant for the System LSI Division at Samsung Electronics, where he developed various computational photographic techniques, including an extended depth-of-field (EDoF) system. From 2005 to 2007, he served as Dean of the Graduate School of Advanced Imaging Science, Multimedia, and Film. From 2005 to 2007, he was Director of the Seoul Future Contents Convergence (SFCC) Cluster established by the Seoul Research and Business Development (R&BD) Program. Dr. Paik is currently serving as a member of the Presidential Advisory Board for Scientific/Technical Policy for the Korean government and is a technical consultant for the Korean Supreme Prosecutor's Office for computational forensics.