

BiCov: a novel image representation for person re-identification and face verification

Bingpeng Ma
bingpeng.ma@unicaen.fr

Yu Su
yu.su@unicaen.fr

Frédéric Jurie
frederic.jurie@unicaen.fr

GREYC — CNRS UMR 6072,
University of Caen Basse-Normandie,
Caen, France

Abstract

This paper proposes a novel image representation which can properly handle both background and illumination variations. It is therefore adapted to the person/face re-identification tasks, avoiding the use of any additional pre-processing steps such as foreground-background separation or face and body part segmentation. This novel representation relies on the combination of Biologically Inspired Features (BIF) and covariance descriptors used to compute the similarity of the BIF features at neighboring scales. Hence, we will refer to it as the BiCov representation. To show the effectiveness of BiCov, this paper conducts experiments on two person re-identification tasks (VIPeR and ETHZ) and one face verification task (LFW), on which it improves the current state-of-the-art performance.

1 Introduction

Person re-identification and face verification tasks are both consisting in recognizing an individual through different images (e.g. images coming from cameras in a distributed network or from the same camera at different time). The key requirement of approaches addressing these tasks is their ability to measure the similarity between two person/face-centered bounding boxes, i.e. to predict if they represent to the same person, despite changes in illumination, pose, viewpoint, background, partial occlusions and low resolution. In this paper, we propose a new image representation allowing to measure effectively the similarity between two persons/faces without requiring any pre-processing step (e.g. background subtraction or body part segmentation). This paper mainly focus on person re-identification which has received less attention than face verification, however we experimentally demonstrate that the proposed representation also works well for face verification.

The proposed method includes two stages. In the first stage, Biologically Inspired Features (BIF) are extracted, through the use of Gabor filters (S1 layer) and MAX operator (C1 layer). In the second stage, the Covariance descriptor [24] is applied to compute the similarity of BIF features at neighboring scales. While the Gabor filters and the Covariance descriptors improve the robustness to the illumination variation, the MAX operator increases the tolerance to scale changes and image shifts. Furthermore, we argue that measuring the

similarity of neighboring scales limits the influence of the background (see Section 3.3 for details). By overcoming illumination, scale and background changes, the performance of person re-identification and face verification is greatly improved.

The proposed method is experimentally validated on two public datasets for person re-identification: VIPeR and ETHZ. They are among the most challenging ones, since all the above-described issues are present: pose changes, viewpoint and lighting variations, occlusions, etc. Regarding face verification, the popular Labeled Faces in the Wild (LFW) dataset is used, allowing comparisons with many recently published approaches.

The remaining of this paper is organized as follows: Section 2 reviews the related works on person re-identification and face verification. Section 3 describes the proposed method in details and analyzes its advantages. Experimental validations are given Section 4. Finally, Section 5 concludes the paper.

2 Related Work

The recent literature abounds with image descriptors for person re-identification and face verification. Person re-identification can be based on different cues such as: (i) color [7] – widely used since the color of clothing constitutes simple but efficient visual signatures – usually encoded by histograms of RGB or HSV values [7], (ii) shape, e.g. using HOG based signatures [17, 20], (iii) texture, often represented by Gabor filters [9, 18, 26], differential filters [9], Harr-like representations [2] or Co-occurrence Matrices [20], (iv) interest points, e.g. SURF [8] and SIFT [13, 27] or (v) image regions [7, 17].

Region-based methods usually split the human body into different parts and extract features for each part. In [17], to recognize humans across aerial images, the authors extracted blobs which are treated as independent regions, and, by assigning a weight to each region, those which are most likely to represent the target’s identity have higher weights. In [4, 7], Maximally Stable Color Regions (MSCR) were extracted and described by their area, centroid, second moment matrix and average color. Finally, the covariance descriptors [1, 3, 26] has also been widely used for representing regions.

Since these elementary features (color, shape, texture, etc.) capture different aspects of the information, they are often combined to give a richer signature. For example, [18] combined 8 color features with 21 texture filters (Gabor and differential filters). [7] proposed the *Symmetry-Driven Accumulation of Local Features* (SDALF), achieving the state-of-the-art performance. SDALF is the combination of weighted Color Histograms (wHSV), MSCR and Recurrent High-Structured Patches (RHSP). However, [4] has observed that RHSP can be removed without losing too much performance. Interestingly, the covariance descriptor can also be used to combine several cues. For example, in [26], Gabor features and Local Binary Patterns (LBP) are combined to form a covariance descriptor which handles efficiently illumination and viewpoint changes, as well as non-rigid body deformations.

Descriptors adapted to face representation are usually different from those for person re-identification. Indeed, face recognition relies more on smaller details of the input image, as intra-class and inter-class variations are smaller than for person re-identification. The most popular approaches rely on LBP or its variants [16], or histogram of SIFT or HOG descriptors [12]. The approach performing best on LFW (in the unsupervised setting) is the Locally Adaptive Regression Kernel (LARK) descriptor [21], which measures the similarity between a pixel and the surrounding pixels. It’s worth noting that contrarily to these previous works that are specialized for either persons or faces, our descriptor is effective on both.

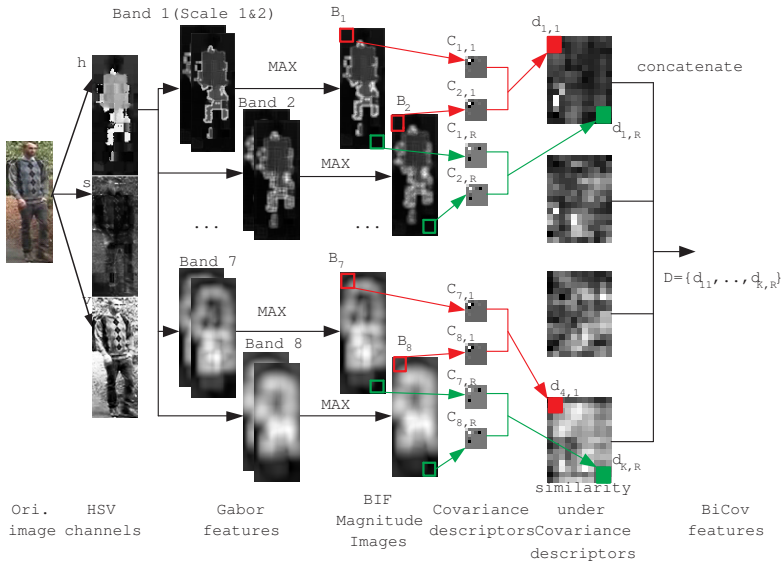


Figure 1: Flowchart of the proposed approach: (i) color images are split into 3 color channels (HSV) (ii) for each channel, Gabor filters are computed at different scales (iii) pairs of neighboring scales are grouped to form one band (iv) magnitude images are produced by applying the MAX operator within the same band (v) magnitude images are divided into small bins and each bin is represented by a covariance descriptor (vi) the difference of covariance descriptors between two consecutive bands is computed for each bin and concatenated to form the image representation.

Finally, metric-learning can be used to further improve the performance by providing a metric adapted to the task (e.g. [10, 18, 27]). While this paper concentrates on image representation, any of these metric-learning approaches can be used on the top of it.

3 Covariance Descriptor based on Bio-inspired Features

The proposed covariance descriptor based on bio-inspired features, BiCov for short, is a two stages representation (see Fig. 1): biologically inspired features are first extracted and then encoded by difference of covariance descriptors at different scales. In the following, the two stages are first introduced, before discussing the advantages of the proposed representation.

3.1 Low-Level Biologically Inspired Features (BIF)

Based on the study of human visual system, bio-inspired features [19] have obtained excellent performances on several computer vision tasks such as object category recognition [22], face recognition [14], age estimation [11] and scene classification [23].

Considering the great success of these BIF, the first step consists in extracting such features to model image low-level properties. For an image $I(x, y)$, we compute its convolution

Table 1: Scales of Gabor filters in different bands.

Band	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8
filter sizes	11x11	15x15	19x19	23x23	27x27	31x31	35x35	39x39
filter sizes	13x13	17x17	21x21	25x25	29x29	33x33	37x37	41x41

with Gabor filters accordingly to the following equations [25]:

$$G(\mu, \nu) = I(x, y) * \Psi_{\mu, \nu}(z) \quad (1)$$

where:

$$\Psi_{\mu, \nu}(z) = \frac{\|k_{\mu, \nu}\|^2}{\sigma^2} e^{\left(\frac{-\|k_{\mu, \nu}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{ik_{\mu, \nu}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (2)$$

$$k_{\mu, \nu} = k_{\nu} e^{i\phi_{\mu}}, k_{\nu} = 2^{-\frac{\nu+2}{2}} \pi, \phi_{\mu} = \mu \frac{\pi}{8} \quad (3)$$

where μ and ν are scale and orientation parameters respectively. In our work, μ is quantized into 16 scales while the ν is quantized into 8 orientations.

In practice, we have observed that for person re-identification, the image representations $G(\mu, \nu)$ for different orientations can be averaged without significant loss of performance. Thus, in this case, we replace $\Psi_{\mu, \nu}(z)$ in Equ. 1 by $\Psi_{\mu}(z)$: $\Psi_{\mu}(z) = \frac{1}{8} \sum_{\nu=1}^8 \Psi_{\mu, \nu}(z)$. This simplification makes the computations of $G(\mu)$ – which is the average of $G(\mu, \nu)$ over all orientations – more efficient.

In all our experiments, the number of scales is fixed to 16 and two neighborhood scales are grouped into one band (we therefore have 8 different bands). The scales of Gabor filters in different bands are shown in Tab. 1. We then apply *MAX pooling* over two consecutive scales (within the same orientation if the orientations are not merged):

$$B_i = \max(G(2i-1), G(2i)) \quad (4)$$

The MAX pooling operation increases the tolerance to small scale changes which often occurs in person and face images since images are only roughly aligned. We refer to B_i $i \in [1, \dots, 8]$ as the *BIF Magnitude Images*. Fig. 2 shows a pair of images of one person and its respective BIF Magnitude Images. The image in the first column is the input image while the ones in the second column are three HSV channels. The images from the 3rd to the 8th column are the BIF Magnitude Images for 6 different bands.

3.2 BiCov Descriptor

In the second stage, BIF Magnitude Images are divided into small overlapping rectangular regions, allowing to preserve the spatial information. Then, each region is represented by a covariance descriptor [24]. Covariance descriptors can capture shape, location and color information, and their performances have been shown to be better than other methods in many situations, as rotations and illuminations changes are absorbed by the covariance matrix [24].

In order to do this, for each pixel of the BIF Magnitude Image B_i a 7-dimensional feature vector is computed to capture the intensity, texture and shape statistics:

$$f_i(x, y) = [x, y, B_i(x, y), B_{ix}(x, y), B_{iy}(x, y), B_{ixx}(x, y), B_{iyy}(x, y)] \quad (5)$$

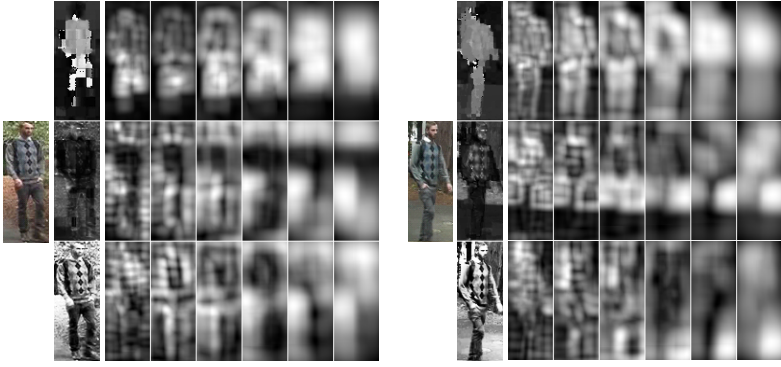


Figure 2: A pair of images and their BIF Magnitude Images. From left to right: the original image, its three HSV channels, six BIF Magnitude Images for different bands.

where x and y are the pixel coordinates, $B_i(x, y)$ is the raw pixel intensity at position (x, y) , $B_{ix}(x, y)$ and $B_{iy}(x, y)$ are the derivatives of image B_i with respect to x and y , while $B_{ixx}(x, y)$ and $B_{iyy}(x, y)$ are the second-order derivatives. Finally, the covariance descriptor is computed for each region of the image:

$$C_{i,r} = \frac{1}{n-1} \sum_{(x,y) \in \text{region } r} (f_i(x,y) - \bar{f}_i)(f_i(x,y) - \bar{f}_i)^T \quad (6)$$

where \bar{f}_i is the mean of $f_i(x, y)$ over the region r and n is the size of region r (in pixels).

Usually, the covariance matrices computed by Eq.6 are considered as the image representation. Differently, in this paper we compute for each region the difference of covariance descriptors between two consecutive bands:

$$d_{i,r} = d(C_{2i-1,r}, C_{2i,r}) = \sqrt{\sum_{p=1}^P \ln^2 \lambda_p(C_{2i-1,r}, C_{2i,r})} \quad (7)$$

where $\lambda_p(C_{2i-1,r}, C_{2i,r})$ is the p -th generalized eigenvalues of $C_{2i-1,r}$ and $C_{2i,r}$, $i = 1, 2, 3, 4$. Finally, the differences are concatenated to form the image representation:

$$D = (d_{1,1}, \dots, d_{1,R}, \dots, d_{K,1}, \dots, d_{K,R}) \quad (8)$$

where R is the number of regions and K is the number of band pairs (4 in our case). The distance between two images I_i and I_j is obtained by computing the Euclidian distance between their representations D_i and D_j :

$$d(I_i, I_j) = \|D_i - D_j\| \quad (9)$$

It is worth pointing out that color images are processed by splitting the image into 3 color channels (HSV), extracting the proposed descriptor on each channel separately and finally concatenating the 3 descriptors into a single signature.

As mentioned in Section 2, it's usually better to combine several image descriptors. In this paper, we combine the BiCov descriptor with two other ones, namely the (a) Weighted



Figure 3: VIPeR dataset: Sample images showing same subjects from different viewpoint.

Color Histogram (wHSV) and (b) the MSCR, such as defined in [7]. For simplicity, we denote this combination as eBiCov (enriched BiCov). The difference between two eBiCov signatures $D_1 = (HA_1, MSCR_1, BiCov_1)$ and $D_2 = (HA_2, MSCR_2, BiCov_2)$ is computed as:

$$d_{eBiCov}(D_1, D_2) = \frac{1}{3}d_{wHSV}(HA_1, HA_2) + \frac{1}{3}d_{MSCR}(MSCR_1, MSCR_2) + \frac{1}{3}d(BiCov_1, BiCov_2) \quad (10)$$

Obviously, further improvements could be obtained by optimizing the weights (*i.e. using a supervised approach*), but as we are looking for an unsupervised method, we let them fixed once for all. Regarding the definition of d_{wHSV} and d_{MSCR} , we use the ones given in [7].

3.3 Analysis of BiCoV

By combing Gabor filters and covariance descriptors – which are both known to be tolerant to illuminations changes [24] – the BiCov representation is robust to illumination variations.

In addition, BiCov is also robust to background variations. Roughly speaking, background regions are not as contrasted as foreground ones, making their Gabor features (and therefore their covariance descriptors) at different neighboring scales very similar. Since the BiCov descriptor is based on the difference of covariance descriptors, background regions are filtered out.

Finally, we want to point out that our approach makes a very different use of the covariance descriptor. In the literature, covariance based similarity is defined by the difference between covariance descriptors computed on two different images. In contrast, BiCov computes the similarity of covariance descriptors *within* the same image, between two consecutive scales. These similarities are then concatenated to obtain the image signature, and the difference of probe and gallery images is obtained by computing the l_2 distance between their signatures. By doing so, we avoid computing the difference of Covariance descriptors of probe image and every gallery image which could be extremely time-consuming when the gallery is large.

4 Experiment

The proposed representation has been experimentally validated on two datasets for person re-identification (VIPeR [10] and ETHZ [6, 20]) and one for face verification (LFW [12]).

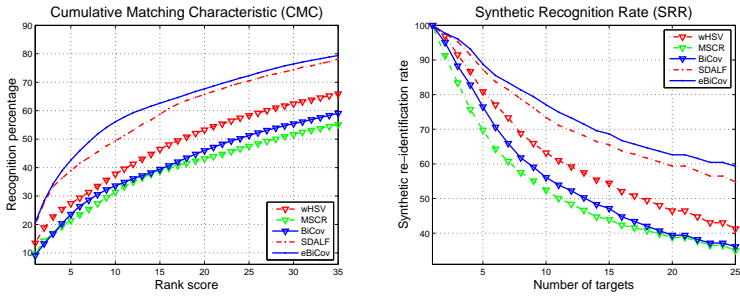


Figure 4: VIPeR dataset: CMC and SRR curves.

4.1 Pedestrian re-identification on VIPeR Dataset

VIPeR is specifically made for viewpoint-invariant pedestrian re-identification. It contains 1,264 images of 632 pedestrians. There are exactly two views per pedestrian, taken from two non overlapping viewpoints. All images are normalized to 128×48 pixels. The VIPeR dataset contains a high degree of viewpoint and illumination variations: most of the examples contain a viewpoint change of 90 degrees, as it can be seen in Fig. 3. This dataset has been widely used and is considered to be one of the benchmarks of reference for pedestrian re-identification. All the experiments on this dataset address the unsupervised setting, i.e. without using training data and therefore not involving any metric leaning.

We use the Cumulative Matching Characteristic (CMC) curve [15] and Synthetic Reacquisition Rate (SRR) curve [10], which are the two standard performance measurements for this task. CMC measures the expectation of the correct match being at rank r while SRR measures the probability that any of the m best matches is correct.

Fig. 4 shows the performance of the eBiCov representation, and gives comparisons with SDALF [7] which is the state of the art approach for this dataset. We follow the same experimental protocol than [7] and report the average performance over 10 different random sets of 316 pedestrians. We can see that eBiCov consistently outperforms SDALF: the rank 1 matching rate for eBiCov is 20.66% while that of SDALF is 19.84%. The rank 10 matching rate for eBiCov is 56.18 while that of SDALF is 49.37. This improvement can be explained in two ways: on one hand, most of the false positives are due to severe lighting changes, changes that the combination of Gabor filters and covariance descriptors can handle efficiently. On the other hand, since many people tend to dress in very similar ways, it is important to capture as fine image details as possible. This is what BIF does. In addition, it's worth noting that for these experiments the orientation of Gabor filters is not used (see section 3), allowing to reduce the computational cost. We have indeed experimentally observed that the performance is almost as good as that with orientations. Finally, Fig. 4 also reports the performance of the 3 components of the eBiCov components (i.e. BiCov, wHSV and MSCR, such as defined in Section 3) when used alone.

4.2 Pedestrian re-identification on ETHZ dataset

The ETHZ dataset contains three video sequences of crowded street scenes captured by two moving cameras mounted on a chariot. SEQ. #1 includes 4,857 images of 83 pedestrians, SEQ. #21, 961 images of 35 pedestrians, and SEQ. #3 1,762 images of 28 pedestrians. The

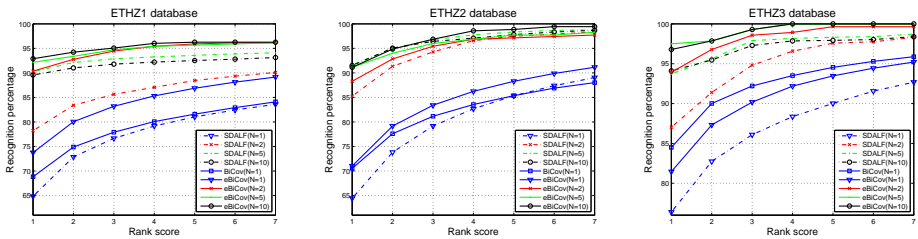


Figure 5: The CMC curves on the ETHZ dataset.

most challenging aspects of ETHZ are illumination changes and occlusions. We follow the evaluation framework proposed by [7] to perform the experiments. Fig. 5 shows the CMC curves for the three different sequences, for both single ($N = 1$) and multiple shots ($N = 2, 5, 10$) cases.

In the single shot case, we can see that the performance of BiCov alone is already much better than that of SDALF, on all of the three sequences. The performance of eBiCov (remind that eBiCov is the combination of BiCov, MSCR and wHSV) is greatly improved on sequence 1 and 2. In particular, on sequence 1, eBiCov is 7% better than SDALF for ranks between 1 and 7. In sequence 2, rank 1 matching rate is around 71% for eBiCov and 64% for SDALF. Compared with the improvements observed on VIPeR, improvements on ETHZ are more obvious. As images come from a few video sequences, they are rather similar and the performance is more heavily depending on the quality of the descriptor.

Besides the single-shot setting, we also tested our method in the multi-shot one. As in [7], N is set to 2, 5, 10. The results are given in Fig. 5. It can be seen that on SEQ. 1 and 3, the proposed eBiCov give much better results than SDALF. It is even more obvious on SEQ. 3 for which our method’s CMC is equal to 100% for $N = 5, 10$, which experimentally validates our descriptor for person re-identification.

4.3 Face verification in uncontrolled environments

Besides person re-identification, we have also experimented our method for face verification on the LFW dataset [12]. LFW consists in 13,233 images of 5,749 people which are originally gathered from news articles on the web. Face recognition on LFW is a challenging problem due to the variations in facial poses, illumination or expressions.

We tested the proposed method on the View 2 of the LFW dataset, following the protocol described in [12]. In View 2, the dataset is partitioned into 10 disjoint folds. A single fold has 600 pairs: 300 positive pairs (i.e. two images of the same person) and 300 negative (images of different persons). The task is to verify if a pair of face images represents the same individual or not. The verification performance is reported as the mean recognition rate and standard deviation over 10-fold cross-validation. The training and testing splits are provided through its website,¹ from which we also obtain the aligned version of the face images with the size of 80×150 .

Face verification is different from person re-identification in several ways. First, more care that should be taken to the preservation of fine details. Here, we compute one BIF image per orientation. As we have 8 different orientations, the size of the descriptor is 8 times bigger than the descriptor used for person re-identification. Second, the color histograms as

¹<http://vis-www.cs.umass.edu/lfw/index.html>

Table 2: Mean classification accuracy (%) and standard deviation on the LFW dataset.

Method	Mean classification accuracy (%) and standard
SD-MATCHES, 125x12512, aligned [5]	64.10±0.62
GJD-BC-100, 122x22512, aligned [5]	68.47±0.65
H-XS-40, 81x15012, aligned [5]	69.45±0.48
LARK unsupervised20, aligned [21]	72.23±0.49
BiCov, aligned	74.03±0.32

well as the MSCR features are less discriminative and we don't use them here.

Tab. 2 reports the performance of our method, as well the performance of state of the art methods. These results are taken from the official LFW website which keeps track of any published results on this dataset. By obtaining a mean classification accuracy and standard deviation of 74.03%, the proposed BiCov descriptor gives a significant improvement over the best performance reported so far.

5 Conclusion

This paper proposed a novel image representation called BiCov which combines Biologically Inspired Features (BIF) and covariance descriptor. BiCov is robust to illumination, scale and background variations, which makes it suitable for both person re-identification and face verification tasks. Experiments on two pedestrian datasets (VIPeR and ETHZ) and one face dataset (LFW) have shown that, in the unsupervised setting, the proposed representation achieves state-of-the-art performance. There are several aspects to be further studied in the future, especially the learning of task-specific metrics.

Acknowledgment

This work was partly realized as part of the Quaero Program funded by OSEO, French State agency for innovation and by the ANR, grant reference ANR-08-SECU-008-01/SCARFACE. The first author is partially supported by National Natural Science Foundation of China under contract No. 61003103.

References

- [1] W. Ayedi, H. Snoussi, and M. Abid. A fast multi-scale covariance descriptor for object re-identification. *Pattern Recognition Letters (In Press)*, 2011.
- [2] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using haar-based and DCD-based signature. *Proc. International Workshop on Activity Monitoring by Multi-camera Surveillance Systems*, 2010.
- [3] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Multiple-shot human re-identification by mean Riemannian covariance grid. *Proc. International Conference on Advanced Video and Signal-Based Surveillance*, 2011.

-
- [4] D. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. *Proc. British Machine Vision Conference*, 2011.
- [5] J. Ruiz del Solar, R. Verschae, and M. Correa. Recognition of faces in unconstrained environments: a comparative study. *EURASIP Journal on Advances in Signal Processing (Recent Advances in Biometric Systems: A Signal Processing Perspective)*, 2009.
- [6] A. Ess, B. Leibe, K. Schindler, and L. van Gool. A mobile vision system for robust multi-person tracking. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [8] N. Gheissari, T. Sebastian, P. Tu, J. Rittscher, and R. Hartley. Person reidentification using spatiotemporal appearance. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2:1528–1535, 2006.
- [9] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. *Proc. European Conference on Computer Vision*, 2008.
- [10] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2007.
- [11] G. Guo, G. Mu, Y. Fu, and T.S. Huang. Human age estimation using bio-inspired features. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [12] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, October 2007.
- [13] J. Kai, C. Bodensteiner, and M. Arens. Person re-identification in multi-camera networks. *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2011.
- [14] E. Meyers and L. Wolf. Using biologically inspired features for face processing. *International Journal of Computer Vision*, 76(1):93–104, 2008.
- [15] H. Moon and P.J. Phillips. Computational and performance aspects of PCA-based face-recognition algorithms. *Perception*, 30(3):303–21, 2001.
- [16] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [17] O. Oreifej, R. Mehran, and M. Shah. Human identity recognition in aerial images. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [18] B. Prosser, W. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. *Proc. British Machine Vision Conference*, 2010.

- [19] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, 1999.
- [20] W. Schwartz and L. Davis. Learning discriminative appearance based models using partial least squares. *Proc. Brazilian Symposium on Computer Graphics and Image Processing*, 2009.
- [21] H.J. Seo and P. Milanfar. Face verification using the LARK face representation. *Proc. IEEE Transactions on Information Forensics and Security*, 6(4):1275–1286, 2011.
- [22] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [23] D. Song and D. Tao. Biologically inspired feature manifold for scene classification. *IEEE Transactions on Image Processing*, 19:174–184, 2010.
- [24] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1713–1727, 2008.
- [25] L. Wiskott, J.M. Fellous, N. Krüger, and C.V.D. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
- [26] Y. Zhang and S. Li. Gabor-LBP based region covariance descriptor for person re-identification. *Proc. Int. Conference on Image and Graphics*, 2011.
- [27] W. Zheng, S. Gong, and T. Xiang. Associating groups of people. *Proc. British Machine Vision Conference*, 2009.