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

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 the new BiCov 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).

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 [3] 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. By overcoming illumination, scale and background changes, the performance of person re-identification and face verification is greatly improved.

Stage 1. Considering the great success of BIF, the first step of BiCov consists in extracting such features to model image low-level properties. For an image $I(x,y)$, we compute its convolution with Gabor filters: $G(\mu,\nu) = I(x,y) * \psi_{\mu,\nu}(z)$ where μ and ν are scale and orientation parameters respectively. In BiCov, μ is quantized into 16 while the ν is quantized into 8.

In practice, we have observed that for person re-identification task, image representations $G(\mu,\nu)$ of different orientations can be averaged without significant loss of performance. Thus, we replace $\psi_{\mu,\nu}(z)$ in Eq. by $\psi_\mu(z) = \frac{1}{8}\sum_{\nu=1}^{8}\psi_{\mu,\nu}(z)$. This simplification makes the computations much more efficient.

In BiCov, two neighborhing scales are grouped into one band (we therefore have 8 different bands) by applying "MAX" pooling over two consecutive scales: $B_i = \max(G(2i-1), G(2i))$ "MAX" pooling operation increases the tolerance to small scale changes which often appear in person and face images since they are only roughly aligned. We call B_i $i \in [1,\cdots,8]$ as *BIF Magnitude Images*.

Stage 2. For each pixel on the BIF Magnitude Image B_i, a 7-dimensional vector is computed to capture the intensity, texture and shape statistics:

$$f_i(x,y) = [x, y, B_i(x,y), B_{i_x}(x,y), B_{i_y}(x,y), B_{i_{xx}}(x,y), B_{i_{yy}}(x,y)] \qquad (1)$$

where x and y are the pixel coordinates, $B_i(x,y)$ is the raw pixel intensity at position (x,y), $B_{i_x}(x,y)$ and $B_{i_y}(x,y)$ are the derivatives of image B_i, while $B_{i_{xx}}(x,y)$ and $B_{i_{yy}}(x,y)$ are the second-order derivatives.

We divide the BIF Magnitude Image B_i into small regions with equal size and overlap. In this way, the spatial information of the images can be kept. Then, each region is represented by a covariance descriptor [3]:

$$C_{i,r} = \frac{1}{n-1}\sum_{(x,y)\in region\ r}(f_i(x,y)-\bar{f}_i)(f_i(x,y)-\bar{f}_i)^T \qquad (2)$$

where $\bar{f}_i$ is the mean of $f_i(x,y)$ over region r and n is the size of region r. 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 [3].

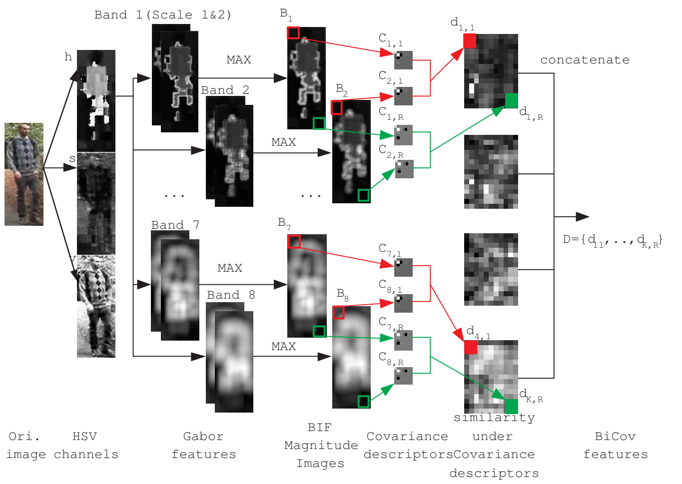

Figure 1: The flowchart of BiCov.

In the traditional covariance-based methods, covariance matrices computed by Eq.2 are considered as image representation. Differently, in this paper, we compute for each region the difference of covariance descriptors between two consecutive bands

$$d_{i,r} = d(C_{2i-1,r}, C_{2i,r}) = \sqrt{\sum_{p=1}^{P}\ln^2\lambda_p(C_{2i-1,r}, C_{2i,r})} \qquad (3)$$

where $\lambda_p(C_{2i-1,r}, C_{2i,r})$ is the pth generalized eigenvalues of $C_{2i-1,r}$ and $C_{2i,r}$, $i = 1,2,3,4$. BiCov avoid computing the difference of covariance descriptors of probe image and every gallery image which could be extremely time-consuming when the gallery is huge.

Finally, the differences are then concatenated to form the image representation: $D = (d_{1,1},\cdots,d_{1,R},\cdots,d_{K,1},\cdots,d_{K,R})$, 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 these representations D_i and D_j.

Better person re-identification performance is usually obtained by combining different type of image descriptors. In this paper, we follow the same methodology and combine BiCov descriptor with other two: (a) Weighted Color Histograms (wHSV) and (b) Maximally Stable Color Regions (MSCR), as defined in [1].

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] 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.

[2] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, 1999.

[3] 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.