Robust Head-shoulder Detection Using a Two-Stage Cascade Framework

Ronghang Hu[†‡], Ruiping Wang[†], Shiguang Shan[†] and Xilin Chen[†]
[†]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
[‡]Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China
{ronghang.hu, ruiping.wang, shiguang.shan, xilin.chen}@vipl.ict.ac.cn

***Abstract*—Head-shoulder detection is widely used in many applications, and robust image descriptors are crucial to the detection performance. In this paper, by exploiting the second-order region covariance descriptor as a complement to widely-used histogram-based descriptors, we propose a new two-stage coarse-to-fine cascade framework to make full use of both types of descriptors for robust head-shoulder detection. Specifically, in the first stage, two histogram-based descriptors, i.e., local Histogram of Oriented Gradients (HOG) and histogram of Local Binary Pattern (LBP), are utilized by a Viola-Jones classifier to rapidly reject most non-head-shoulder candidate windows. In contrast, the second stage further boost the performance via multiple kernel learning on Riemannian manifold formed by Region Covariance Matrix (RCM), a second-order statistic descriptor with stronger discriminative power. Experimental results on a public dataset demonstrate that our method improves detection rate significantly with satisfactory detection speed.**

I. Introduction

Head-shoulder detection is an important research topic in computer vision and is widely used in many applications, including human tracking [7], people counting (especially in crowded scenes) [6], [16], and other tasks in surveillance systems. In many aspects, head-shoulder detection shares some similarities with pedestrian detection. However, the two tasks have their inherent differences mainly in that the head-shoulder part of human is less deformable than the entire human body and is less often occluded even in crowded scenes, where pedestrian detectors often suffer from occlusion problems. On one hand, head-shoulder detection can be seen as a complement to pedestrian detection for human related vision tasks. On the other hand, head-shoulder detectors can also be incorporated into pedestrian detection systems as part detectors.

Recently, much work has been done on head-shoulder detection [6], [7], [16], most of which is based on HOG (Histogram of Oriented Gradients) [2] and LBP (Local Binary Pattern) [14] features following the sliding-window approach. However, although HOG feature is effective in describing the omega-like shape of human head-shoulder and LBP feature as a texture descriptor further improves the performance, the detection rate of HOG and LBP based detectors in previous work is still unsatisfactory for real applications.

It is known that HOG and LBP, as histogram-based descriptors, only describe the distribution of some certain image properties (*e. g.* gradient orientation) in a region. In order to further improve the detection performance, we intend to model not only such certain image properties (or features) individually, but also the correlation between them. A natural approach to modeling the correlation between different features is to calculate their covariance. Then, the covariances of several image features inside a region of interest can be used as a second-order region descriptor. Therefore, we would like to use Region Covariance Matrix (RCM) [11] as a complement to HOG and LBP to obtain better performance. However, RCM descriptor can be computationally expensive compared with histogram-based descriptors, so using it directly in the traditional feature extraction-classification process in a sliding-window approach will largely sacrifice the detection speed. Another difficulty to use RCM is that non-singular covariance matrices reside on non-linear Riemannian manifold, so traditional classification approaches designed for Euclidean space cannot be applied directly with RCM features.

In this paper, we propose a novel two-stage cascade framework for robust head-shoulder detection, combining the merits of different descriptors in a coarse-to-fine structure. Based on HOG and LBP features, the first stage is a Viola-Jones type classifier [13], which is also a multi-level cascade itself and can reject more than 99% of non-head-shoulder patches with high speed. In the second stage, RCMs are extracted from subwindows of different sizes in the detection windows that pass the first stage. Since RCM resides on non-linear Riemannian manifold, we map it onto Euclidean space using Log-Euclidean distance [1], [15]. Then, we build an effective classifier using multiple kernel learning [12], with each kernel defined on one of the subwindows. Since only a very small fraction of detection windows can pass the first stage and need further classification, the second stage also runs quite fast. Experimental results on NLPR-HS [6], a public head-shoulder dataset, demonstrate that our detection method by incorporating different types of descriptors in a two-stage cascade framework improves the detection performance significantly and achieves satisfactory detection speed at the same time.

The rest of the paper is organized as follows. We present a review of related work in Section 2. A detailed description of our two-stage detection framework is given in Section 3. We show experimental results in Section 4 and conclude our work in Section 5.

II. Related Work

In general, previous work on the topic of head-shoulder detection and other objects detection mainly focuses on exploiting one or more of the following aspects.

A. Better Feature

Viola and Jones successfully used Haar feature for fast human face detection [13], but it is shown in [6] that Viola-

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.482

Jones classifier based on Haar feature has poor classification performance in head-shoulder detection. It is also shown in [6] that SIFT feature is not discriminative enough for head-shoulder detection.

HOG (Histogram of Oriented Gradients) feature has been proven to be effective in describing shape and boundary. Dalal and Triggs first used HOG feature for pedestrian detection [2]. Li *et al.* [6] later applied HOG feature in head-shoulder detection.

LBP (Local Binary Pattern) is also histogram-based region descriptor and has been widely used for object detection. Wang *et al.* [14] improved the performance of pedestrian detection by combining HOG feature with uniform LBP feature as the feature set. Zeng and Ma [16] applied multi-level HOG-LBP feature in head-shoulder detection.

RCM (Region Covariance Matrix) is another type of region descriptor different from HOG and LBP. As a second order statistic, RCM is powerful in describing the correlation between different low-level features, and has been shown to be effective in pedestrian detection [11] and texture classification [10]. However, RCMs, as symmetric positive definite matrices, reside on Sym^+ Riemannian manifold rather than Euclidean space, which makes it difficult and also non-trivial to apply traditional classification methods directly.

B. Stronger Classifier

The most common classification method is to first divide the detection window into several subwindows (blocks) and extract local feature vectors from each subwindow [2], [14], [16]. Then the feature vectors from different subwindows are concatenated into one single vector to characterize the detection window and a SVM classifier can be learned by using the detection windows as training samples.

Boosting is also widely used for classification. Viola and Jones [13] used AdaBoost to select local Haar features from a large pool to build a real-time face detector. Later, AdaBoost was also applied in pedestrian detection by Zhu *et al.* to select local HOG features [17]. Based on local RCMs, Tuzel *et al.* used LogitBoost on Sym^+ Riemannian manifold for pedestrian detection and achieved promising performance [11]. However, their approach involves complicated nonlinear mapping onto tangent space of the Sym^+ manifold and is thus computationally expensive in both training and detection.

Multiple kernel learning [12] is proven to be another effective classification approach. It learns an appropriate linear combination of different kernels from training data, and can be viewed as a method of metric learning in Reproducing Kernel Hilbert Space (RKHS) by adjusting the weight of each kernel. Using a nonlinear mapping, multiple kernel learning can be extended to Sym^+ manifold spanned by RCMs [5].

C. Higher Speed

Consisting of coarse-to-fine stages, cascade is a frequently used method to increase detection speed by rejecting most of negative detection windows in earlier stages [13], [17]. Other methods, including integral image [13] and integral histogram [9], are often applied to further speed up the detection by computing local features more efficiently.

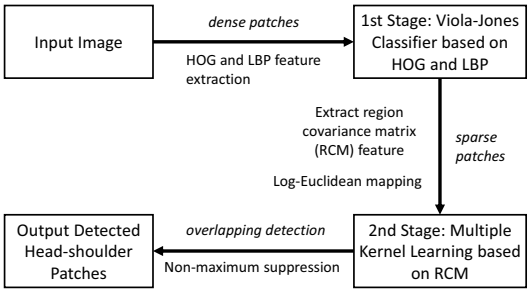


Fig. 1. The flowchart of our head-shoulder detection framework

III. Proposed Method

In our two-stage cascade framework, different types of descriptors and different classification methods are used respectively in each stage. We use a Viola-Jones type classifier based on local HOG and LBP features in the first stage and quickly reject over 99% of non-head-shoulder patches. Then, an RCM based multiple kernel learning classifier is used in the second stage to further classify the detection windows that have passed the first stage. Finally, non-maximum suppression is applied to suppress overlapping detections. Figure 1 illustrates the process of our head-shoulder detection framework. The size of detection windows used in our method is 32×32.

A. HOG-LBP based Viola-Jones Classifier

Viola-Jones classifier [13] is a multi-level cascade classifier and is used in the first stage of our detection framework to achieve a high detection rate with high speed. We sample blocks of various sizes ranging from 6×6 to 30×30 in the 32×32 detection window. Then, a 36-D HOG feature vector and a 59-D LBP feature vector are extracted from each of these blocks. For each level of the cascade classifier, we train linear SVMs as weak classifiers based on local HOG and LBP features, then the strong classifier of each level is trained using AdaBoost.

To compute the HOG feature, each block is further divided into four rectangle cells, and gradient magnitude of all pixels in each cell is voted into a histogram divided evenly into 9 bins according to gradient orientation. Histograms of the four cells are concatenated into a 36-D vector and normalized using L2 normalization. Details of how to extract HOG feature can be seen in [2].

To compute the LBP feature, we use $LBP^2_{8,1}$. The intensity of each pixel is compared with its 8 neighboring pixels with radius 1 (bilinear interpolation is used if the coordinates of a nearby pixel are not integers), and a binary pattern of eight bits is extracted. Then, the binary patterns of all pixels in the block are voted into a 59-D histogram consisting of 58 bins of each uniform pattern plus 1 bin of all non-uniform patterns, and normalized using L2 normalization. Details of how to extract

LBP feature can be seen in [14]. Using integral histogram [9], both HOG and LBP features can be calculated efficiently.

It is worth mentioning that Zhu *et al.* [17] used a similar method to train a Viola-Jones classifier for pedestrian detection, but only HOG feature is used in their implementation. However, it is found in our experiments that LBP feature improves the detection rate significantly for head-shoulder detection.

B. Region Covariance Matrix (RCM)

RCM is used to further classify the detection windows that pass the first stage. As we mentioned previously, RCM, as a second-order region descriptor, is proven to be powerful in describing the distribution of different low-level features and the correlation between them. This makes it discriminative for many tasks including pedestrian detection [11].

To construct RCM, each pixel in a $W \times H$ image patch is mapped onto a 8 dimensional feature vector as

$$[x \ y \ |I_x| \ |I_y| \ \sqrt{I_x^2 + I_y^2} \ |I_{xx}| \ |I_{yy}| \ \arctan \frac{I_x}{I_y}]^T \quad (1)$$

where x and y are pixel coordinates and I_x, I_{xx}, I_y and I_{yy} are intensity derivative of the pixel. Note that this feature mapping can be calculated efficiently as we only need to compute I_{xx} and I_{yy}, because I_x, I_y, $\sqrt{I_x^2 + I_y^2}$ and $\arctan \frac{I_x}{I_y}$ of the entire image have already been computed when extracting HOG feature in the first stage.

In the second stage of our detection framework, we manually select 59 subwindows of 3 different sizes: 32×32, 16×16 and 8×8. Two nearby subwindows of the same size overlap with each other by 50%. Then, the covariance matrix C_r calculated from the mapped features (1) of all pixels in a subwindow is used as the descriptor of this subwindow. Note that although C_r is an 8×8 matrix, it has only 36 degrees of freedom due to symmetry.

To make the covariance matrix C_r (which corresponds to each subwindow) robust to illumination variations, it is further normalized using the covariance matrix of the whole detection window,

$$\hat{C}_r = diag(C_R)^{-1/2} C_r diag(C_R)^{-1/2} \quad (2)$$

where C_R is the covariance matrix of the whole detection window and $diag(\cdot)$ means to keep the diagonal entries and truncate the rest to zero.

For each detection window, we extract its 59 normalized RCMs from all its subwindows. Although RCM can be computed using integral image approach [11], we do not calculate the integral image of the entire image, which is computationally expensive. Instead, we only calculate the integral image of the detection window, since only a small fraction of the image patches can pass the first stage.

C. Multiple Kernel Learning on Riemannian Manifold

An effective classification approach is needed to classify a detection window using its 59 RCMs. It is well known that non-singular covariance matrices reside on Sym^+ Riemannian manifold rather than Euclidean space, and concatenating the columns of an RCM into one vector will ignore the geometry of the manifold and result in inferior performance [11]. Therefore, traditional classification approaches like SVM operating in vector space cannot be applied directly to RCMs.

Fortunately, by exploiting the novel Log-Euclidean distance [1], one can map points from the Sym^+ Riemannian manifold onto Euclidean space while at the same time preserving the geometry of the manifold. The Log-Euclidean distance between two covariance matrices C_1 and C_2 is defined as

$$d_g(C_1, C_2) = ||\log(C_1) - \log(C_2)||_F \quad (3)$$

where $\log(\cdot)$ is the matrix logarithm and $||\cdot||_F$ is the matrix Frobenius norm. The Log-Euclidean distance defines a true geodesic distance on Sym^+ manifold and can be efficiently computed using eigenvalue decomposition of C_1 and C_2 [11].

Based on Eq. 3, the Log-Euclidean mapping from a $d \times d$ covariance matrix C to a $d(d+1)/2$ dimensional vector $\mathbf{c}$ in a Euclidean space is defined as

$$\mathbf{c} = \text{vec}(\log(C)) \quad (4)$$

where the vector operator $\text{vec}(\cdot)$ of a $d \times d$ symmetric matrix X is defined as

$$\text{vec}(X) = [X_{1,1} \ \sqrt{2}X_{1,2} \ \sqrt{2}X_{1,3} \cdots X_{2,2} \ \sqrt{2}X_{2,3} \cdots X_{d,d}]^T \quad (5)$$

which takes the upper-triangle part of X and multiplies its non-diagonal entries by $\sqrt{2}$. It can be easily verified that the Euclidean distance in the vector space after Log-Euclidean mapping equals to the geodesic distance in the Sym^+ manifold, i.e. $||\mathbf{c_1} - \mathbf{c_2}||_2 = d_g(C_1, C_2)$, and thus the geometry of the manifold is preserved after the mapping, while traditional classification approaches can be applied now in the vector space.

Based on the above Log-Euclidean mapping, we construct a classifier using multiple kernel learning [12], [5] in the second stage of our detection framework. Multiple kernel learning is an effective method to learn a new kernel as a linear combination of several existing kernels (weights are non-negative to ensure positive definiteness). It corresponds to the concatenation of the Reproducing Kernel Hilbert Space (RKHS) of each kernel, while the different weights of the kernels scale the metric in their RKHS. Therefore, multiple kernel learning can be regarded as a method to learn an appropriate metric in kernel space.

In our approach, we define 59 Gaussian kernels on the 32×32 detection window, each of which corresponds to one of the 59 subwindows:

Fig. 2. Some typical samples in NLPR-HS dataset

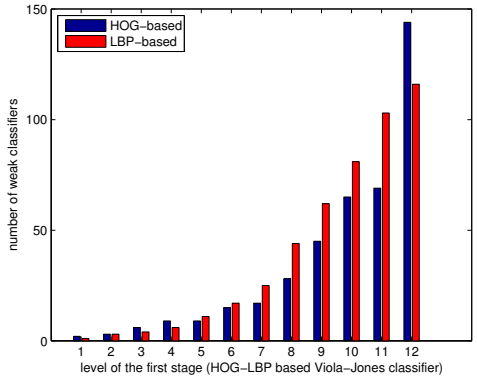


Fig. 3. Numbers of HOG-based and LBP-based weak classifiers in each level in the 1st stage

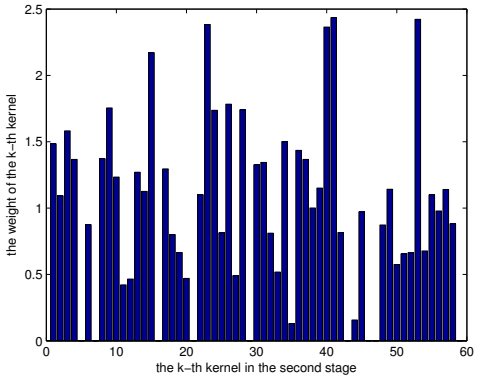


Fig. 4. The weights of 59 kernels in multiple kernel learning in the 2nd stage

$$k_m(W_i, W_j) = \exp(-\gamma||\mathbf{c}_{i,m} - \mathbf{c}_{j,m}||_2^2), m = 1, 2, \cdots, 59 \quad (6)$$

where W_i and W_j are detection windows, and $\mathbf{c}_{i,m}$ and $\mathbf{c}_{j,m}$ are the Log-Euclidean mapping of the normalized RCMs extracted from the mth subwindow of W_i and W_j respectively.

Then, a linear combination of these 59 kernels

$$k(W_i, W_j) = \sum_{m=1}^{59} d_m k_m(W_i, W_j) \quad (7)$$

can be learned from the training set $\{(W_i, y_i)\}$ through the following optimization similar to the standard SVM optimization:

$$\min_{\mathbf{w},\mathbf{d},\boldsymbol{\xi}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\mathbf{1}^T\boldsymbol{\xi} + \sigma\mathbf{1}^T\mathbf{d} \quad (8)$$

$$\text{subject to } y_i(\mathbf{w}^T\phi(W_i) + b) \geq 1 - \xi_i, \boldsymbol{\xi} \geq 0, \mathbf{d} \geq 0 \quad (9)$$

$$\text{where} \quad \phi(W_i)^T\phi(W_j) = k(W_i, W_j) \quad (10)$$

Here $\mathbf{d} = [d_1, d_2, \cdots, d_{59}]^T$ is the weight vector, and C is the training error cost. The L1 regularization on $\mathbf{d}$ can avoid overfitting and enforce sparsity on the weights [12], so that less kernels are needed to be computed, which helps improve the detection speed.

The positive training set in the second stage is the same as that in the first stage, while the negative training set is collected from the false positives of the first stage. Cross-validation is used to determine the values of hyper-parameters, including γ used in Gaussian kernels and σ used in regularization term.

IV. Experiments

A. Dataset

Compared with abundant benchmark datasets for pedestrian detection and face detection, there are few widely used public datasets for human head-shoulder detection, which makes it difficult to compare different detection approaches. We chose the public NLRP-HS dataset [6] to evaluate our head-shoulder detection framework. With head-shoulder patches cropped from Internet images, surveillance videos and two pedestrian datasets (MIT set [8] and INRIA set [2]), the NLRP-HS dataset consists of 3510 positive samples of size 32×32 and 399 head-shoulder-free images for training, and 1812 positive samples and 331 head-shoulder-free images for testing. It is a challenging dataset, since positive samples have low resolution and large variations, and are sometimes blurred. These challenges are often encountered in real applications. Figure 2 shows some typical samples in this dataset.

B. Training

The HOG-LBP based Viola-Jones classifier in the first stage was trained level by level using AdaBoost. We sampled 1519 blocks of various sizes from the 32×32 detection window. Then, for each level, we trained 3038 linear SVMs as weak classifiers (each from one block using HOG or LBP) and used AdaBoost to build a strong classifier from them. We required the minimum detection rate of each level to be 0.997 and the maximum false positive rate to be 0.6. The training process of the first stage was stopped when the weight of a newly added weak classifier is below a certain threshold (10^{-4} in our implementation).

To train the multiple kernel learning classifier in the second stage, false positives of the first stage were used as negative samples and positive samples were directly obtained from the NLPR-HS training set.

After an initial training of our two-stage detector, we scanned on the head-shoulder-free images from the training set and added false positives into the negative training set of the second stage to retrained the multiple kernel learning classifier. This retraining process was repeated twice.

The first stage of our final detector consists of 12 levels and 885 weak classifiers in total (412 HOG-based and 473 LBP-based). Figure 3 shows the numbers of HOG-based and LBP-based weak classifiers in each level respectively. It can be seen that the multiple levels in the first stage also follow a coarse-to-fine structure, and over 90% negative patches can be quickly rejected in the first 4 levels involving only 34 weak classifiers.

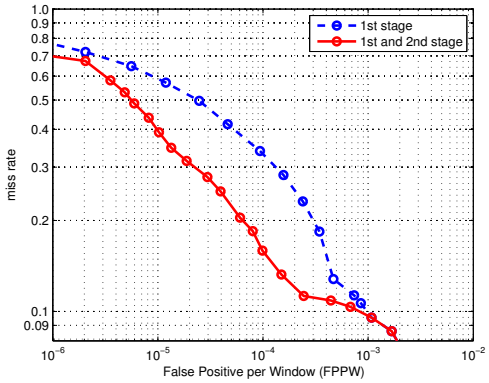


Fig. 5. Performance of different stages in our detection framework on NLPR-HS dataset

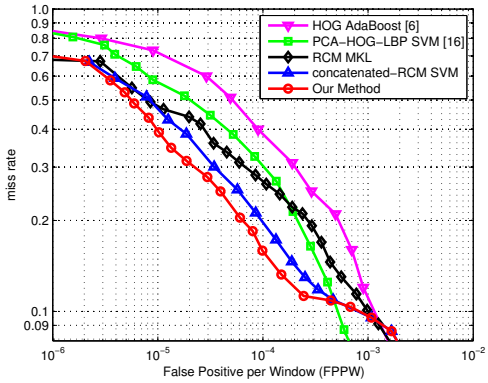


Fig. 6. Comparision of our detection method with other methods on NLPR-HS dataset. See text for details of each method.

The second stage of our final detector involves a linear combination of 59 different kernels and the weights of these kernels are shown in Figure 4. It can be seen that there is large variation between these weights, corresponding to the scaled metric in the RKHS of each kernel. Moreover, the weights of 9 kernels from 59 ones (15%) are zero, so detection time can be saved by skipping these kernels.

C. Evaluation and Comparisons

We used Detection-Error Tradeoff (DET) curves with per-window metric to evaluate the detection performance. On the test set, the first stage has a detection rate of 91% with a false positive rate of 1.7×10^{-3} FPPW, and adding the second stage reaches a detection rate of 85% with a false positive rate of 10^{-4} FPPW. Figure 5 shows the DET curves of our detector. The DET curve of the first stage is obtained by adjusting the threshold of the final level in the Viola-Jones classifier, and the DET curve of the whole detector is generated by adjusting the threshold of the multiple kernel learning classifier. As shown in Figure 5, the detection rate is significantly improved when the second-order RCM detector is introduced, which indicates that it is effective to incorporate RCM feature as a complement to HOG and LBP for head-shoulder detection.

We then compared our head-shoulder detector with other methods, and plot the DET curve of each method in Figure 6.

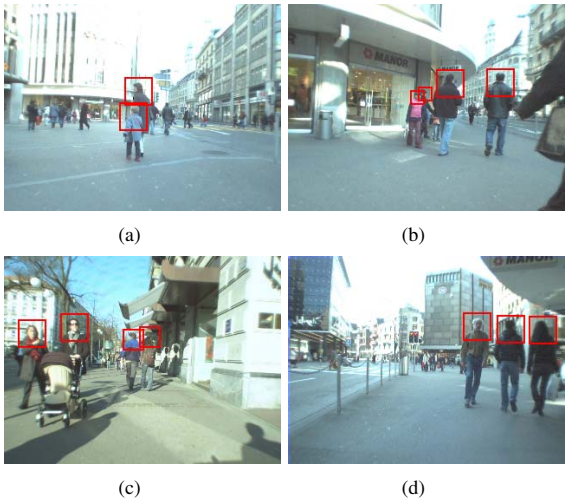
1) HOG AdaBoost: The HOG based AdaBoost detector was trained and tested on NLPR-HS dataset in [6]. It used AdaBoost as classification approach similar to the first stage of our method, but only HOG feature was extracted. The result is directly obtained from the original paper [6]. DET curves show that the detection rate of this method is inferior to the others, which indicates that HOG feature itself is insufficient in discriminative power for head-shoulder detection.

2) PCA-HOG-LBP SVM: Zeng and Ma [16] used HOG-LBP feature calculated from manually selected subwindows of different sizes, and applied PCA to reduce noise and to prevent overfitting. Then, they used linear SVM for classification and reported a detection rate of 89% at 10^{-4} FPPW on their own dataset. Since their dataset is kept private, we have carefully implemented their method and evaluated the method on the public NLPR-HS dataset with accuracy of 72% at 10^{-4} FPPW. However, it is worth noting that the two datasets are largely different in resolution and even in definition of head-shoulder region. Moreover, while 84 subwindows were manually selected in their implementation, we only selected 59 subwindows since the sample size (32×32) of NLPR-HS dataset is smaller than theirs (48×64).

3) RCM MKL: The discriminative power of RCM descriptor itself (i.e. only the second stage of our method) is also evaluated on NLPR-HS dataset. We trained an RCM based multiple kernel learning classifier for head-shoulder detection and tested its performance. As shown in Figure 6, the detection rate of this method is superior to the HOG-LBP based SVM detector [16] at 10^{-4} FPPW. However, the detection speed is too slow (over 31 seconds per frame) for practical use.

4) Concatenated-RCM SVM: In this approach, the first stage is the same Viola-Jones classifier as in our detection framework, but in the second stage, the 36-D covariance vectors (obtained through Log-Euclidean mapping) of the 59 subwindows is concatenated into a 2124-D vector and a Gaussian kernel SVM is trained for classification. Note that by concatenating feature vectors from all subwindows followed by one Gaussian kernel, the resulting kernel space is essentially the tensor product of the Reproducing Kernel Hilbert Space (RKHS) of all Gaussian kernels defined on each subwindow, so its dimensionality should be much higher compared with the linear combination of such kernels learned through multiple kernel learning. In our experiment, however, we evaluated the performance of this approach and found it inferior to the multiple kernel learning approach. This may be explained by the fact that multiple kernel learning not only concatenates the RKHS of the kernels but also scales the metric, and that too high dimensionality often causes overfitting.

Through above comparisons, it is shown that our method using different types of descriptors in a coarse-to-fine two-stage



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

Fig. 7. Some detection results on ETH Pedestrian Dataset

cascade framework is effective for head-shoulder detection.

D. Experiment on ETH Pedestrian Dataset

To evaluate the generalization ability of our method, we also tested our detector on ETH Pedestrian Dataset [4]. Some of the detection results are shown in Figure 7. It can be seen from Figure 7 (a) (c) that our detector can still successfully detect the head-shoulder part of pedestrians when the body is occluded, whereas a pedestrian detector may fail in such cases due to occlusion. Since the performance of most pedestrian detectors is disappointing for partially occluded pedestrians [3], our head-shoulder detector can be used as a complement to pedestrian detectors to improve their performance in occluded scenes, as the head-shoulder part of pedestrians is less likely to be occluded.

E. Speed Issue

The detection speed is often as important as the detection rate in real applications. Although it can be time-consuming to use Log-Euclidean mapping and Gaussian kernels, by building a fast Viola-Jones classifier based on HOG and LBP features extracted using integral histogram, we are able to rapidly reject most of negative detection windows in the first stage so a lot of time is saved. In our CPU implementation using C++, we achieved a detection speed of 6 fps (frame per second) on 320×240 images on a laptop computer (with Intel Core i5 CPU and 4GB memory), which is satisfactory for most applications including head-shoulder tracking and people counting in crowded scenes. For applications requiring a higher speed, one can build a real-time head-shoulder detector on GPU using our method. Here we also note that the detection speed of an HOG and LBP based Viola-Jones classifier (simply removing the second stage in our detection framework) reaches 11 fps but with inferior detection rate, as shown in Figure 5.

V. Conclusion

We have proposed a novel two-stage cascade framework for robust head-shoulder detection in this paper. Using HOG and LBP as histogram-based region descriptors, the first stage of Viola-Jones classifier quickly rejects over 99% of non-head-shoulder detection windows and ensures relatively high detection speed, while the second stage exploits the correlation between different image properties using RCM as a second-order region descriptor and multiple kernel learning as an effective classification method. We combine the merits of different types of region descriptors by incorporating them in the coarse-to-fine two-stage cascade framework. Experimental results on a public head-shoulder dataset have demonstrated that our method improves the detection rate significantly with satisfactory detection speed, and has the potential to be used as a complement to pedestrian detectors in occluded scenes.

Acknowledgment

The work is partially supported by Natural Science Foundation of China under contracts nos.61025010, 61173065, and 61379083.

References

[1] V. Arsigny, P. Fillard, X. Pennec and N. Ayache, Geometric Means in a Novel Vector Space Structure on Symmetric Positive-definite Matrices, in *SIAM Journal on Matrix Analysis and Applications*, 29, 2007.
[2] N. Dalal and B. Triggs, Histograms of Oriented Gradients for Human Detection, in *CVPR*, 2005.
[3] P. Dollar, C. Wojek, B. Schiele, and P. Perona, Pedestrian Detection: An Evaluation of the State of the Art, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 2012.
[4] A. Ess, B. Leibe, K. Schindler and L. van Gool, A Mobile Vision System for Robust Multi-Person Tracking, in *CVPR*, 2008.
[5] S. Jayasumana, R. Hartley, M. Salzmann, H. Li and M. Harandi, Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices, in *CVPR*, 2013.
[6] M Li, Z. Zhang, K. Huang and T. Tan, Estimating the Number of People in Crowded Scenes by MID based Foreground Segmentation and Head-shoulder Detection, in *ICPR*, 2008.
[7] M Li, Z. Zhang, K. Huang and T. Tan, Rapid and Robust Human Detection and Tracking Based on Omega-Shape, in *ICIP*, 2009.
[8] C. Papageorgious and T. Poggio, A trainable system for object detection, in *International Journal of Computer Vision*, 38, 2000.
[9] F. Porikli, Integral histogram: A fast way to extract histograms in cartesian spaces, in *CVPR*, 2005.
[10] O. Tuzel, F. Porikli and P. Meer, Region Covariance: A Fast Descriptor for Detection and Classification, in *ECCV*, 2006.
[11] O. Tuzel, F. Porikli and P. Meer, Pedestrian Detection via Classification on Riemannian Manifolds, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 2008.
[12] M. Varma and D. Ray, Learning the Discriminative Power-Invariance Trade-off, in *ICCV*, 2007.
[13] P. Viola and M. J. Jones, Robust Real-Time Face Detection, in *International Journal of Computer Vision*, 57, 2004.
[14] X. Wang, T. X. Han and S. Yan, An HOG-LBP Human Detector With Partial Occlusion Handling, in *ICCV*, 2009.
[15] R. Wang, H. Guo, L. S. Davis and Q. Dai, Covariance Discriminative Learning: A Natural and Efficient Approach to Image Set Classification, in *CVPR*, 2012.
[16] C. Zeng and H. Ma, Robust Head-Shoulder Detection by PCA-Based Multilevel HOG-LBP Detector for People Counting, in *ICPR*, 2010.
[17] Q. Zhu, S. Avidan, M. C. Yeh, and Kwang-Ting Cheng, Fast Human Detection Using a Cascade of Histograms of Oriented Gradients, in *CVPR*, 2006.