Symmetry and Asymmetry Features for Human Detection

Xinchuan Fu(✉) and Shihai Shao

University of Electronic Science and Technology of China, Chengdu 611731, China
xinchuan.fu@foxmail.com, ssh@uestc.edu.cn

Abstract. Edge is a very important type of feature for human detection and Histogram of Oriented Gradient (HOG) is the most popular method to encode edge information since proposed. Because HOG detects edges based on intensity gradients, it is not invariant with respect to image illumination. In this paper, we propose three new types of features based on local phase: Local Phase based symmetry (LPS), Local Phase based Asymmetry (LPA), and Histogram of Oriented Asymmetry (HOA) for human detection. The LPA and HOA are similar with gradient magnitude and HOG features, but from different perspective. The key idea is the intensity around an edge point in an image is always asymmetry. Thus we can detect edges by measuring the asymmetry of the local structure at every point in the image. This is achieved by analyzing the phase of its constituent frequency components. This asymmetry measurement is invariant with respect to image contrast. After the asymmetry is computed, this value could be distributed to different orientation bins according to gradient orientation. We also measure symmetry around each point which yields LPS. This is useful to detect torso and limbs. These local phase induced features are combined with the classical Aggregated Channel Features (ACF) and are fed into the boosted decision tree (BDT) framework. Experiment shows that the proposed features are complementary to the ACF features and will increase the detection accuracy.

Keywords: Human detection · Symmetry · Asymmetry

1 Introduction

Human detection is an active research topic in recent years. The application includes Advanced Driver Assistance Systems (ADASs), visual surveillance and human-robot interaction, etc. Recently, deep learning methods have achieved the state-of-the-art accuracy [7,16], but these methods rely on high-end GPU device because of the high computation cost. On the other hand, hand-craft features together with boosted decision tree (BDT) methods are also competitive for its light-weight CPU implementation [3,10,19]. Among all the handcraft features,

Supported by the National Natural Science Foundation of China 61771107.

B. Li et al. (Eds.): IoTaaS 2018, LNICST 271, pp. 303–311, 2019.
https://doi.org/10.1007/978-3-030-14657-3_30

Histogram of Oriented Gradient (HOG) [4] is the most popular one for human detection.

The basic idea of HOG is to compute gradients in every point in the image. Then distribute the gradients magnitude to several orientation bins and form histogram in different cells. Currently, the HOG features are usually used together with gradient magnitude and CIE-LUV color. The resulting features are called Aggregated Channel Features (ACF) [5]. Many newly devised features are based on ACF, like LDCF [12], InformedHaar [17], Checkerboard [18] and MRFC [2], etc.

Though very effective for human and general object detection, the HOG features have an intrinsic limitation rooted in the gradient computation. In the implementation of HOG, the gradients are computed by the intensity difference, which is sensitive to local contrast. The computed edge of a person in the image will have different strength because of different illumination conditions, or just because the person wear clothes of different color. This is contrary to the goal of invariant representation in feature design. The existence of an edge should determined by the image structure, not the intensity difference. Though HOG compensates for the contrast variance by local normalization, this operation is coarse. For example, the strength of an edge with low contrast will be suppressed by an edge with high contrast in the local region.

Is there an alternative way to capture edge information? The answer is positive. By inspecting the edges in an image, we could see that the edge points usually locate about the midpoint of a transition ramp and the intensity around this point exhibit an approximate asymmetry. Based on this insight, we could detect edges by measuring the local asymmetry. Peter Kovesi [11] proposed a method measuring the symmetry and asymmetry based on local phase. For a point to be an symmetry point, all frequency components must have phase 0 or π. Conversely, for a point to be an asymmetry point, all frequency components must have phase $\pi/2$ or $3\pi/2$. Thus we could measure the extent of asymmetry by phase deviation from $\pi/2$ or $3\pi/2$. This type of features are called Local Phase based Asymmetry (LPA). We could also measure the extent of symmetry by phase deviation from 0 or π. This type of features are called Local Phase based Symmetry (LPS). The reason we compute LPS is that symmetry is a useful cue for human detection [1,3]. Following the idea of HOG, we distribute the estimated asymmetry level to different orientation bins. The resulting features are called Histogram of Oriented Asymmetry (HOA). The symmetry and asymmetry measurement only considers the structure of the local signal hence is invariant to local contrast. We add all these symmetry and asymmetry induced feature channels to ACF channels and put them into BDT to train a human classifier. The experiment result shows that the combined channels have superior performance than ACF.

2 Method

In this section, we first introduce the symmetry and symmetry measure for one dimensional signal, then we extend this method to color image, finally we introduce the HOA and show how these features are used for human detection.

2.1 Symmetry and Asymmetry Measure for One Dimensional Signal

As we stated in Sect. 1, to measure the symmetry and asymmetry around a point of a signal, we could compute the local phase deviation of its constituent frequency components of the signal. Suppose $I(x)$ is a one dimensional signal and $\phi_n(x)$ is the phase of the nth frequency component at point x. According to [11], the symmetry $S(x)$ and asymmetry $R(x)$ could be estimated as

$$\begin{aligned} S(x) &= \sum_n \frac{A_n(x)\lfloor |cos(\phi_n(x))| - |sin(\phi_n(x))| \rfloor}{A_n(x) + \varepsilon} \\ &= \sum_n \frac{\lfloor |e_n(x)| - |o_n(x)| \rfloor}{A_n(x) + \varepsilon} \end{aligned} \tag{1}$$

$$\begin{aligned} R(x) &= \sum_n \frac{A_n(x)\lfloor |sin(\phi_n(x))| - |cos(\phi_n(x))| \rfloor}{A_n(x) + \varepsilon} \\ &= \sum_n \frac{\lfloor |o_n(x)| - |e_n(x)| \rfloor}{A_n(x) + \varepsilon}, \end{aligned} \tag{2}$$

where $\lfloor \cdot \rfloor = max(\cdot, 0)$, $A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$. ε is a small value to avoid zero denominator. In our experiment, ε is set to 0.001. $e_n(x)$ and $o_n(x)$ is the even and odd parts of the bandpass filtered signal, which are computed as

$$e_n(x) = I(x) * g_n^e(x) \tag{3}$$

$$o_n(x) = I(x) * g_n^o(x), \tag{4}$$

where $g_n^e(x)$ is a bandpass filter and $g_n^o(x)$ is its Hilbert transform. Log-Gabor filter [9] is usually chosen as the bandpass filter. The advantage of log-Gabor filter is that it has zero DC component for arbitrary large bandwidth. Note that in the original implementation [11] there is another term to compensate for the noise influence. But in our paper, the features extracted in this stage will be fed into the BDT and the decision tree will automatically choose a threshold for a feature. Thus the noise compensation is not necessary here.

2.2 Symmetry and Asymmetry for Color Images

To detect edges in an image which may exhibit various orientations, we can perform the previous operations with bandpass filters of different orientations

and combine the result. But the more efficient way is to use the monogenic signal [8]. The monogenic signal is a generalization of analytic signal to 2D case with the Hilbert transform in the analytic signal replaced by Riesz transform. For 2D signal, the symmetry and asymmetry estimation is similar with Eqs. 1 and 2, except now there are two odd parts

$$o1_n(\mathbf{x}) = I(\mathbf{x}) * g_n^{o1}(\mathbf{x}) \tag{5}$$

$$o2_n(\mathbf{x}) = I(\mathbf{x}) * g_n^{o2}(\mathbf{x}), \tag{6}$$

where the frequency representation of $g_n^{o1}(\mathbf{x})$ and $g_n^{o2}(\mathbf{x})$ is given by the Riesz transform of $G_n(\boldsymbol{\omega})$

$$G_n^R(\boldsymbol{\omega}) = \frac{i\boldsymbol{\omega}}{|\boldsymbol{\omega}|} G_n(\boldsymbol{\omega}) = [G_n^{o1}(\boldsymbol{\omega}), G_n^{o2}(\boldsymbol{\omega})]. \tag{7}$$

The $o_n(x)$ in Eqs. 1 and 2 is then replaced by

$$o_n(\mathbf{x}) = \sqrt{o1_n(\mathbf{x})^2 + o2_n(\mathbf{x})^2} \tag{8}$$

For color image, there are three channels. We first convert the image from RGB space to LUV space, and then compute $S(\mathbf{x})$ and $R(\mathbf{x})$ for three channels separately. Finally we take the maximum value of the three computed channels as the final feature.

2.3 Histogram of Orientated Asymmetry

For human detection, it is beneficial not only detect edge, but also know the orientation of the edge. In the HOG implementation, the gradient magnitude is distributed to several predefined orientation bins, which encodes orientations of edges. In the same way, we could also distribute the asymmetry value to different orientations to indicate which axis the local structure is asymmetry to and the resulting features are HOA. As we stated in Sect. 1, the asymmetry could be taken as a similarity measure to an edge structure. Thus the distributed asymmetry value will represent the edge orientations. The orientation in each point could be estimated by the two odd filtered signals

$$ori(\mathbf{x}) = atan2(o1_n(\mathbf{x}), -o2_n(\mathbf{x})). \tag{9}$$

However, because in our experiment we will use the ACF features together with our features, the orientation has already been computed by the gradient based method. To save computation, we just use this gradient based orientation to compute HOA. The number of orientation is set to six, the same with that in ACF features. Together with the symmetry and asymmetry feature channels, we have $1 + 1 + 6 = 8$ new feature channels. As for ACF features, these channels are further aggregated with a shrinkage factor. In our experiment, we set the shrinkage factor to four.

3 Experiments

3.1 Experiments for One Dimensional Signal

We first show the symmetry and asymmetry features for a one dimensional signal from real data. We convert a human image (see Fig. 2(a)) from RGB color space to LUV color space and show the U channel in Fig. 1(a). A row from this channel is selected as our 1D signal (denoted by the red dash line). The first row of Fig. 1(b) is the selected 1D signal. Compared with the left image, we could see that the visual edge points are about the middle point of the transition ramp. The second row and the third row of Fig. 1(b) show the computed LPS and LPA. The peaks of the LPS correspond to the peaks or troughs of the original signal. The peaks of the LPA correspond to the transition ramp of the original signal, which correspond the edge of the image. Note that the value of the LPA range from 0 to 1 and a high value does not correspond to a high intensity difference. This is what we want because we need a descriptor to capture the edge structure which is invariant of the local contrast.

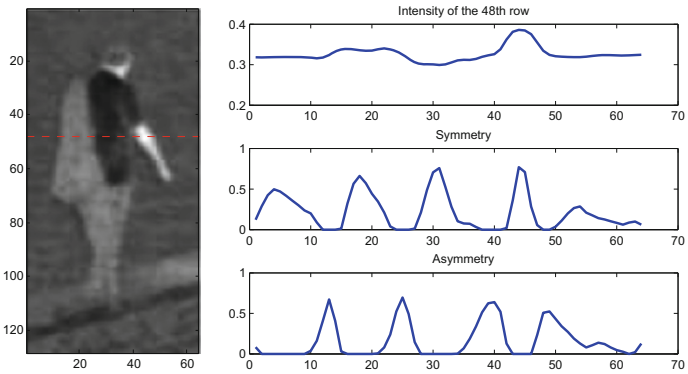

Fig. 1. Illustration of LPS and LPA for one dimensional signal. (a) The U channel of an human image. (b) From the top to bottom: the selected 1D signal, the LPS and the LPA. (Color figure online)

3.2 Experiment for a Color Image

In this section, we show the symmetry and asymmetry feature maps for a color image. The original color image is show in Fig. 2(a). The LPS feature map is show in Fig. 2(b). From the figure, we can see that the LPS feature map have high response at head, torso and limbs. These features are discriminative for human detection. Figure 2(c) shows the LPA feature map, which effectively capture the edge information of the image. Again, we find the asymmetry response is irrelevant with the local contrast, only correlated with the local structure.

(a) (b) (c)

Fig. 2. Illustration of the LPS and LPA feature maps for a color image. (a) The original color image. (b) The LPS feature map. (c) The LPA feature map.

3.3 Evaluation on INRIA Dataset

In this section, we evaluate our proposed method on the INRIA person dataset. Miss Rate (MR) vs. False-Positive-Per-Image (FPPI) curve and log average miss rate are used as evaluation metric [6]. The log average miss rate is computed by averaging miss rate at 9 FPPI points that are evenly spaced in the log-space ranging from 10^{-2} to 10^{0}.

We incorporate the proposed features into the BDT framework to train a human detector. The model size is set as 128×64. Five frequency components are used to compute symmetry and asymmetry features whose center frequencies are $2\pi/\lambda_k$, where $\lambda_k = 10 * 1.5^{k-1}, k = 1, 2, ..., 5$. The training process includes three hard negative mining stages. The final classifier consists of 4096 level-3 decision trees. To show the effectiveness of our new features, two models are trained. The first model only use the 10-channel ACF features to train. For the second model, the first 2048 weak classifiers still use the ACF features to train. For the latter 2048 weak classifiers we replace the seven gradient based channels with our local phase based channels (but keep the color channels). Both models are evaluated on the test set, thus we can see how the detection accuracy evolves with the weak classifier number.

The test result is shown in Fig. 3. As shown in the figure, with increasing number of decision trees, the log average miss rate (the lower the better) gradually decreases (though with fluctuation). For both models, the first 2048 trees use the same features, hence the log average miss rate is always the same. However, as the gradient based channels are replaced by our local phase based channels for the later 2048 trees, the second model achieves higher accuracy than the first one. After 4096 decision trees the ACF features only achieve 13.25% log average miss rate, while our method achieves 11.39%. This improvement shows the effectiveness of our new features.

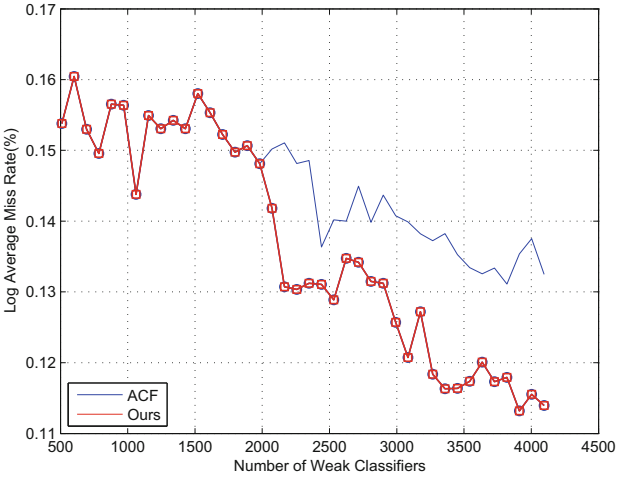

Fig. 3. Weak classifier number versus the log average miss rate on the INRIA test dataset.

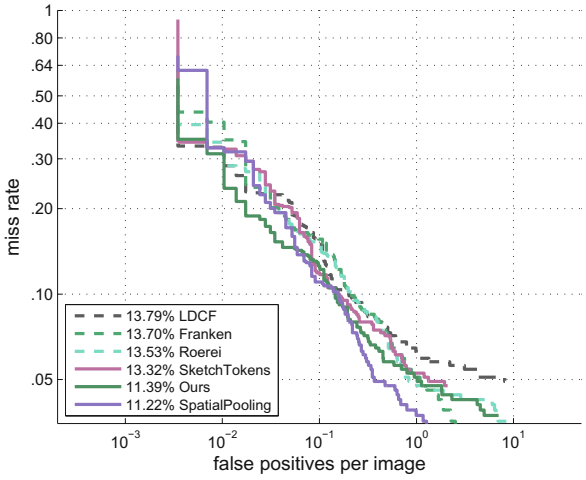

Fig. 4. Comparison with state-of-the-art non-CNN methods on the INRIA test dataset.

Next, we compare our detector with other non deep learning detectors using MR vs. FPPI curves in Fig. 4. For a fixed FPPI point, we prefer a lower MR, thus the curves close to the left-bottom corner are better. Log average miss rate is given at the legend to rank different method. From the figure, we can see that our detector outperforms other detectors, except a slightly worse than the SpatialPooling [14]. Note that we only use 18 feature channels in total, while SpatialPooling has used 259 feature channels. Except for ACF features, it also used other feature types like covariance [15] and LBP [13].

4 Conclusion

In this paper, we propose three types of local phase based features, LPS, LPA, and HOA. The LPA and HOA are alternatives for gradient magnitude and HOG. Compared with gradient magnitude and HOG, the advantage of LPA and HOA is its invariance with respect to local contrast. LPS is used for capture the symmetry structure of a human. Experiment for one dimensional signal and color image shows the visual reasonability of this method. Experiment on INRIA person data set shows that these features will boost the performance of the ACF features.

References

1. Cao, J., Pang, Y., Li, X.: Pedestrian detection inspired by appearance constancy and shape symmetry. IEEE Trans. Image Process. **25**(12), 5538–5551 (2016). https://doi.org/10.1109/TIP.2016.2609807
2. Costea, A.D., Nedevschi, S.: Semantic channels for fast pedestrian detection. In: 2016 IEEE CVPR, pp. 2360–2368 (2016). https://doi.org/10.1109/CVPR.2016.259
3. Costea, A.D., Varga, R., Nedevschi, S.: Fast boosting based detection using scale invariant multimodal multiresolution filtered features. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 21–26 July 2017, Honolulu, HI, USA, pp. 993–1002 (2017). https://doi.org/10.1109/CVPR.2017.112
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 20–26 June 2005, San Diego, CA, USA, pp. 886–893 (2005). https://doi.org/10.1109/CVPR.2005.177
5. Dollár, P., Appel, R., Belongie, S.J., Perona, P.: Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1532–1545 (2014). https://doi.org/10.1109/TPAMI.2014.2300479
6. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 743–761 (2012). https://doi.org/10.1109/TPAMI.2011.155
7. Du, X., El-Khamy, M., Lee, J., Davis, L.S.: Fused DNN: a deep neural network fusion approach to fast and robust pedestrian detection. In: 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017, 24–31 March, Santa Rosa, CA, USA, pp. 953–961 (2017). https://doi.org/10.1109/WACV.2017.111
8. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Trans. Sig. Process. **49**(12), 3136–3144 (2001). https://doi.org/10.1109/78.969520
9. Field, D.J.: Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A-Opt. Image Sci. Vis. **4**(12), 2379–2394 (1987)
10. Kim, H.K., Kim, D.: Robust pedestrian detection under deformation using simple boosted features. Image Vis. Comput. **61**, 1–11 (2017). https://doi.org/10.1016/j.imavis.2017.02.007
11. Kovesi, P.: Symmetry and asymmetry from local phase. In: Tenth Australian Joint Conference on Artificial Intelligence, pp. 2–4 (1997)

12. Nam, W., Dollár, P., Han, J.H.: Local decorrelation for improved pedestrian detection. In: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, 8–13 December 2014, Montreal, Quebec, Canada, pp. 424–432 (2014). http://papers.nips.cc/paper/5419-local-decorrelation-for-improved-pedestrian-detection
13. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002). https://doi.org/10.1109/TPAMI.2002.1017623
14. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Strengthening the effectiveness of pedestrian detection with spatially pooled features. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 546–561. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_36
15. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on Riemannian manifolds. IEEE Trans. Pattern Anal. Mach. Intell. **30**(10), 1713–1727 (2008). https://doi.org/10.1109/TPAMI.2008.75
16. Zhang, L., Lin, L., Liang, X., He, K.: Is faster R-CNN doing well for pedestrian detection? In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 443–457. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_28
17. Zhang, S., Bauckhage, C., Cremers, A.B.: Informed haar-like features improve pedestrian detection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, 23–28 June 2014, Columbus, OH, USA, pp. 947–954 (2014). https://doi.org/10.1109/CVPR.2014.126
18. Zhang, S., Benenson, R., Schiele, B.: Filtered channel features for pedestrian detection. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, 7–12 June 2015, Boston, MA, USA, pp. 1751–1760 (2015). https://doi.org/10.1109/CVPR.2015.7298784
19. Zhao, Y., Yuan, Z., Chen, D., Lyu, J., Liu, T.: Fast pedestrian detection via random projection features with shape prior. In: 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017, 24–31 March 2017, Santa Rosa, CA, USA, pp. 962–970 (2017). https://doi.org/10.1109/WACV.2017.112