

Rotation angle recovery for rotation invariant detector in lying pose human body detection

Dao-Xun Xia^{1,2}, Shao-Zi Li^{1,2}

¹School of Information Science and Technology, Xiamen University, Xiamen 361005, People's Republic of China

²School of Mathematics and Computer Science, Guizhou Normal University, Guiyang 550001, People's Republic of China
E-mail: szlig@xmu.edu.cn

Published in *The Journal of Engineering*; Received on 2nd March 2015; Accepted on 4th March 2015

Abstract: A method for rotation-invariant lying-pose human body detection in overlooking images is proposed. The rotation-invariant histogram of oriented gradient using Fourier analysis in polar coordinate is exploited as descriptor for lying-pose human body. And then the authors used the exhaustive sliding window search strategy with multiple scale scan to localise human body. Finally, principal component analysis (PCA) is used to determine the rotation angle of the exhaustive sliding window based on the classifier output scores. Experiments on their built XiaMen University Lying-Pose Dataset (XMULP) show the effectiveness of their proposed method.

1 Introduction

Lying pose of human body detection is a specific object detection with various applications, such as elderly fall detection, victims detection, and search and rescue missions based on unmanned aerial vehicles. It is more challenging than face or pedestrian detection. However, its related research is still in its infancy, and has many unsolved problems because of the variable appearance, wide range of poses and cluttered background.

In most common object detection frameworks, two important intergradients are: (i) finding out a descriptor to encode the object, such as histogram of oriented gradient (HOG), local binary pattern (LBP) or aggregated channel features (ACFs); (ii) choosing a localisation strategy to determine where the objects are, such as sliding window, jumping window or bound and branch search. The Pattern Analysis, Static Modeling and Computational Learning (PASCAL) Visual Object Classes (VOC) results show that sliding window based object detection is the best choice for the tradeoff between detection performance and running time. As for lying-pose human detection, because of the in-plane rotation, one can use the descriptor without rotation invariant and then localise human body with sliding window in the image rotation and scale space, which is very time-consuming. Rotation-invariant descriptors are necessary or preferred in many real applications. If the descriptor is rotation invariant, then the localisation can be performed only in the image scale space, avoiding the time-consuming search in rotation space. The difference of in-plane rotation object detection between sliding window strategy combining with rotation-invariant descriptor and with non-rotation-invariant descriptor.

Currently, there exist many rotation-invariant descriptors. Takacs *et al.* [1] proposed a radial gradient transform method and incorporated it into a rotation-invariant fast feature, obtaining a similar performance with speeded up robust feature (SURF) for image retrieval and matching but 16× faster than SURF. Wang *et al.* [2] proposed a rotation-invariant detector based on polar-HOGs, which transforming rotation object to cycle translation with polar coordinate mapping, and then applying reverse cycle translation to eliminate the rotation effect. Liu and Skibbe *et al.* [3] presents a method to build rotation-invariant HOG descriptors using Fourier analysis in polar coordinates. They further explore the usage of tensor derivatives for fast computation of rotation-invariant descriptions in biomedical images. There also exist many local rotation-invariant descriptors for image matching, such as scale-invariant feature transform (SIFT), rotation-invariant LBP

(ri-LBP). Among these rotation-invariant descriptors, since the Fourier HOG (F-HOG) has the following advantages: (i) F-HOG is a truly rotation-invariant HOG descriptor without neglecting some intrinsic properties of rotations; (ii) F-HOG is higher-level semantic features, including both the local and global information of input image; (iii) the mapping from the raw image to the final descriptors is continuous and smooth, meaning that F-HOG avoids the discretisation artefacts and keeps the substantial information about the gradient patterns. Therefore we choose it as our lying-pose human body descriptors.

However, there is no free lunch for rotation-invariant descriptor. We cannot obtain the object rotation angle during detection. Fortunately, we observe that the shape of the distribution of the confidence score, whose rotation angle is very similar to the orientation of lying-pose human body. Therefore we propose a PCA-based rotation angle recovery method. Experiments on our built dataset show the effectiveness of our proposed method.

Three major contributions in this paper are: (i) the F-HOG is introduced to describe lying-pose human body. (ii) A rotation angle recovery method based on the confidence map with PCA is proposed. (iii) A novel lying-pose human detection dataset is available for download.

2 F-HOG descriptor

As for F-HOG, gradient histogram is regarded as a continuous angular signal, which can be represented by Fourier basis. Let the image gradient computed from one pixel be $d = [dx, dy]^T \in R^2$, and the rotation angle is $\Phi(d)$, Liu *et al.* [3] use a Dirac delta function h to represent the distribution, as is shown in (1)

$$h(\varphi) = \|d\| \delta(\varphi - \Phi(d)) \quad (1)$$

where $\phi = \Phi(\chi) = \text{atan2}(y, x) \in [0, 2\pi)$ and $\chi = [x, y]^T \in R^2$. To expound a continuous and rotation friendly representation, the Fourier basis is used. The Fourier coefficients \hat{f}_m for h are computed as follows

$$\hat{f}_m = \langle h, e^{-im\varphi} \rangle = \|d\| e^{-im\Phi(d)} = \|d\| \overline{\Psi_m(d)} \quad (2)$$

where $\Psi_m(d)$ is a short notation for $\Psi_m(\Phi(d))$ and $m \in \mathbb{Z}$. Both the normalisation and spatial aggregation can be performed by convolution. Let D be the gradient field, $K_1: R^2 \rightarrow R$ be the convolution kernel for the spatial aggregation, $K_2: R^2 \rightarrow R$ be the convolution

kernel for the local normalisation based on gradient energy and $\hat{F}_m: R^2 \rightarrow C$ be the densely computed Fourier representation \hat{F}_m from (2), then the F-HOG field with degree- m component can be represented as follows

$$\tilde{F}_m = \frac{\hat{F}_m * K_1}{\sqrt{\|D\|^2 * K_2}} \quad (3)$$

In the results of F-HOG field $\tilde{F}_m: R^2 \rightarrow C^M$, M is the number of coefficients representing a density function. It has all the advantages of the HOG. Furthermore, it encodes the local structures with the continuous density functions; hence, the gradient orientations are well retained through the spatial smoothing. Actually, we only need a few low-frequency Fourier coefficients to encode the useful information, $|m| \leq 4$ in our experiments.

Since a basis function in the form $P(r)e^{ik\varphi}$ has a nice, simple rotation behaviour, we can easily construct regional descriptors by filtering on the F-HOG field. We only need to choose a proper radial basis $P(r)$ to build a two-dimensional (2D) basis for describing the F-HOG field. A natural choice is to sample on the radius, thus the 2D basis functions for computing regional descriptors are

$$U_{j,k}(r, \varphi) = \delta(r - r_j)e^{ik\varphi} \quad (4)$$

where $j \in N_0$, $k \in Z$, $r = \|\chi\|$ and $\chi = [x, y]^T$ in polar coordinates. Hence, computing the convolution between such a basis function $U_{j,k}$ and a component of the F-HOG field \tilde{F}_m , generates a feature which describes the configuration of HOG features in the region covered by $U_{j,k}$. $U_{j,k} * \tilde{F}_m$ have the rotation order $k - m$. Thus, the final rotation-invariant regional descriptors are built as shown in (5)

$$B = \overline{(U_{j_1,k_1} * \tilde{F}_{m_1})}(U_{j_2,k_2} * \tilde{F}_{m_2}) \quad (5)$$

where $\forall k_1 - m_1 = k_2 - m_2$. It is worth noting that $e^{in\varphi} = \overline{e^{i(-n)\varphi}}$, therefore (5) covers all possible rotation-invariant quantities from coupling two of the filtering results. The descriptors generated in this way are very effective, as the coupling provides the possibility to create many more invariant features, compared with only taking the magnitude of expansion coefficients. Fig. 1 shows part of the channels of the final rotation-invariant features generating from an overlooking image.

3 PCA for the cardinal direction

After obtaining the rotation-invariant feature, support vector machine (SVM) with linear kernel is applied to determine whether the currently scanning window contains a human body. As the sliding window performs in the whole full image, we can obtain a corresponding confidence map, as shown in Fig. 2. It can be found that the shape of the distribution of the confidences around the human body centre is close to the ground truth bounding box, which inspired us to propose a simple but efficient human body rotation angle recovery method based on PCA. Details are as follows.

First, we obtain detection results with non-maximum suppression without knowing the rotation angles. And then we remove the pixels around the detected bounding box centre whose confidences are under a threshold τ . PCA is performed on the un-removed pixels using their coordinates, denoted by a matrix $X_{2 \times n}$, where each row represents coordinates of a pixel, n is the number of pixels. The DC component of is removed by (6)

$$P = X - \Lambda \times \text{mean}(X) \quad (6)$$

where Λ is a column vector with m elements and value of each element is 1. Next, the eigenvalues and eigenvectors of P are

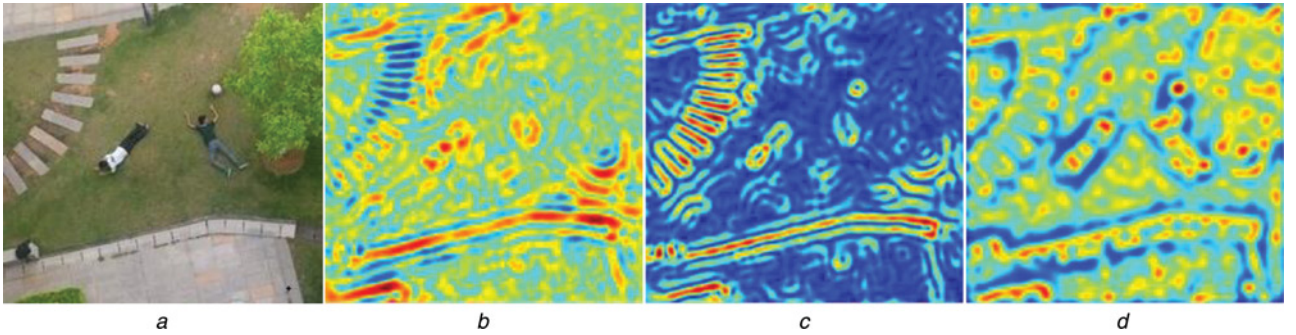


Fig. 1 Final rotation-invariant features are generated from overlooking images (a). The last three features (b), (c) and (d) are only representative

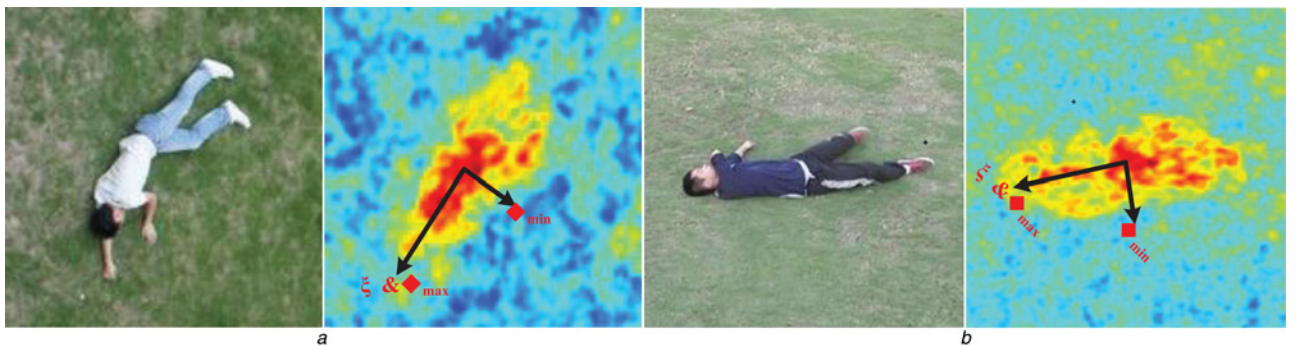


Fig. 2 Distribution of the classifier scores. a (left) denotes the lying-pose region of overlooking images, and a (right) is the detect confidence map. a has a strong resemblance to b

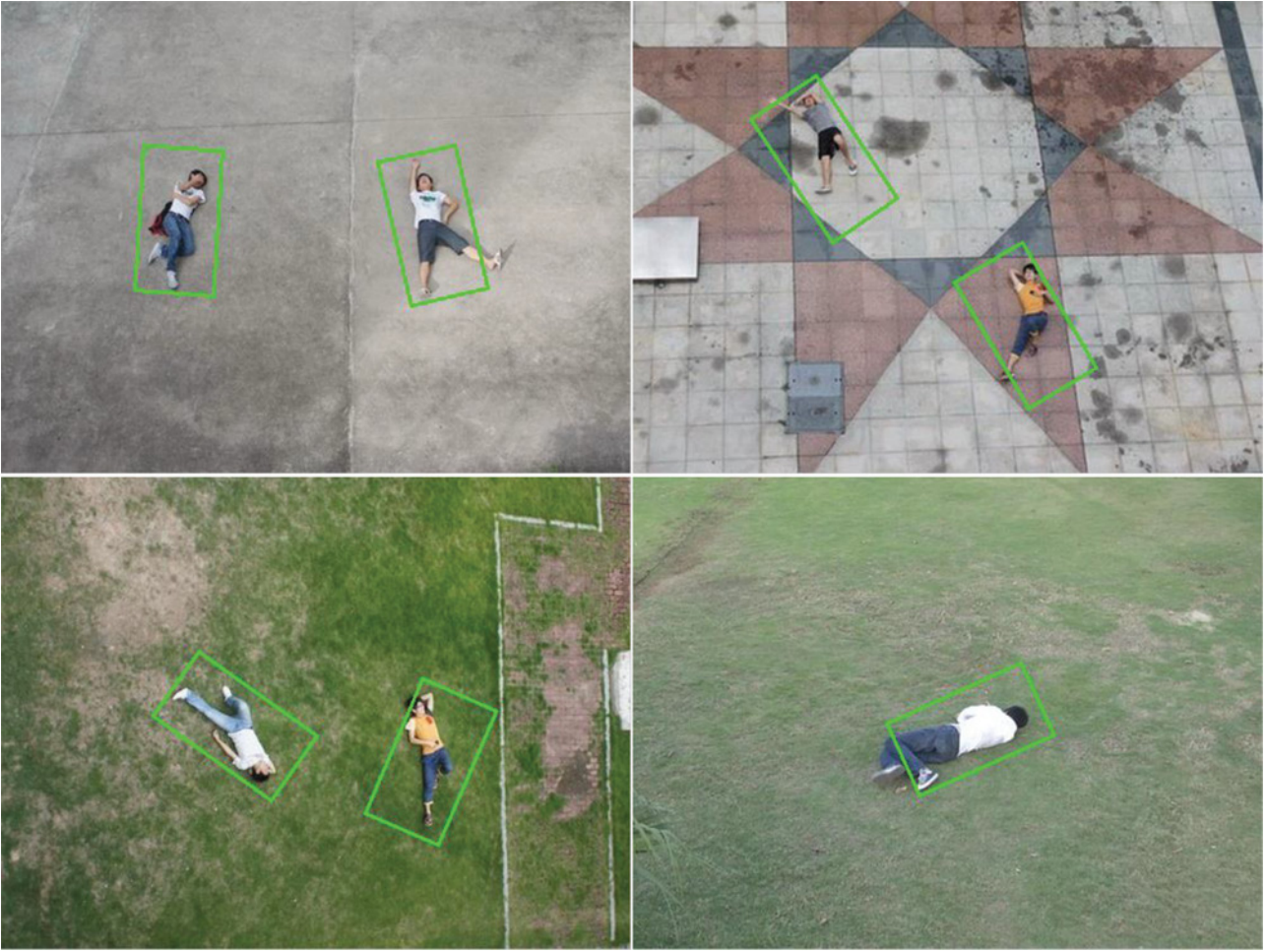


Fig. 3 Partial detection results of our method in overlooking image

computed by (7)

$$[E, D] = \text{Eig}\left(\frac{P \times P^T}{n}\right) \quad (7)$$

where Eig is a function of the singular value decomposition. E is a singular matrix with each column representing a singular vector, whereas D is a diagonal matrix of eigenvalues. After a descending sort for eigenvalues, we can obtain a new sorted matrix \hat{D} and its corresponding eigenmatrix \hat{E} . The rotation angle of human body ξ , can be computed by the principal component of X , as shown in (8)

$$\xi = \frac{180^\circ \times \text{atan}(\hat{E}'(2, 1)/\hat{E}'(1, 1))}{\pi} \quad (8)$$

Finally, the height (σ_{\max}) of the bounding box along principal

Table 1 Bounding box detection AP (%) on XMULP test set

Number	Methods	AP
0	HOG SVM	29.8
1	DPM	38.3
2	ACF	39
3	Our method	50.3

component is the same with that of the original bounding, whereas the width (σ_{\min}) is half of the height. The whole process of searching the principal direction of the lying pose in the distribution of the classifier scores is shown by Fig. 2 and there are two group sketch maps. The left Fig. 2a denotes the lying-pose region of overlooking images, through which the robust and effective detector can quickly detect confidence map, as shown in the right Fig. 2a.

4 Experimental results

Since there are no public datasets available for the evaluation of lying-pose human detection, we built a XMULP (Lying-Pose Dataset of XiaMen University) dataset, containing a total of 1316 images with 1–7 person per image, for a total of 2019 human bodies. Details of the dataset are presented in our previous work [4]. We make a comparison of our method with three state-of-the-art pedestrian detectors performing in the image scale and rotation space. The compared detectors are: HOG + SVM [5], deformable part model (DPM) [6] and ACF [7]. Average precision (AP) is used to gauge detectors' performance here.

In our experiments, we ran the release code of F-HOG in a personal computer with Intel(R) Core (TM) i5 four nuclear, 64 bit operating system. We empirically chose to use the first five degrees, which lead to $\hat{F}_m: m \in \{0, 1, 2, 3, 4\}$. We applied the filters $U_{j,k} = \Delta(r - r_j, \sigma) e^{ik\theta}$ to each \hat{F}_m to obtain the regional descriptor. Only the lower degrees $-4 \leq k \leq 4$ were considered. Finally, a total of 232 real-valued features in the final rotation-invariant

descriptor were obtained. We also mined hard samples and retrained the classifier in order to improve the performance of the detector. As shown in Table 1, our method obtained the highest AP, which is 50.3%, whereas HOG+SVM, DPM and ACF have an AP of 29.8%, 38.3% and 39%, respectively. The performance is promising since it is even better than ACF. Fig. 3 shows part of the detection results of our method. Owing to the rotation-invariant descriptor, our proposed detection method (~ 4.2 s per frame) is the fastest among the compared detectors.

5 Conclusion

This paper introduces a rotation-invariant HOG descriptors algorithm for a novel application, lying-pose detection in overlooking images. We found that F-HOG is very discriminative and performs better than traditional HOG. Moreover, with the rotation-invariant feature, the localising process can avoid the time-consuming scanning in the image rotation and scale space. Finally, we also proposed a simple rotation angle recovery method based on the confidence map via PCA. Experimental results obtained on the three state-of-the-art human-shape detection methods show that our proposed method is effective.

6 Acknowledgments

This work has been supported by the Nature Science Foundation of China (no. 61202143) and the Natural Science Foundation of GuiZhou Province (no. QKH J Zi LKS[2013] 24).

7 References

- [1] Takacs G., Chandrasekhar V., Tsai S.S., Chen D., Grzeszczuk R., Girod B.: 'Fast computation of rotation-invariant image features by an approximate radial gradient transform', *Phys. TIP*, 2013, **22**, (8), pp. 2970–2982
- [2] Wang L., Wu C.D., Chen D.Y., Lu B.H.: 'Rotation invariant human detection scheme based on polar-hogs feature and double scale direction estimation', *Phys. SOPO*, 2011, pp. 1–4
- [3] Liu K., Skibbe H., Schmidt T., Blein T., Palme K., Brox T., *ET AL.*: 'Rotation-invariant HOG descriptors using Fourier analysis in polar and spherical coordinates', *Phys. IJCV*, 2014, **106**, (3), pp. 342–364
- [4] Xia D.X., Su S.Z., Li S.Z., Jodoin P.M.: 'Lying-pose detection with training dataset expansion'. ICIP, October 2014
- [5] Dalal N., Triggs B.: 'Histograms of oriented gradients for human detection'. CVPR, June 2005, vol. **2**, pp. 886–893
- [6] Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D.: 'Object detection with discriminatively trained part-based models', *Phys. PAMI*, 2010, **32**, (9), pp. 1627–1645
- [7] Dollár P., Appel R., Belongie S., Perona P.: 'Fast feature pyramids for object detection', *Phys. PAMI*, 2014, **36**, (8), pp. 1532–1545