Improve Pedestrian Attribute Classification by Weighted Interactions from Other Attributes

Jianqing Zhu, Shengcai Liao, Zhen Lei, Stan Z. Li
jianqingzhu@foxmail.com, {scliao, zlei, szli}@cbsr.ia.ac.cn

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

Abstract. Recent works have shown that visual attributes are useful in a number of applications, such as object classification, recognition, and retrieval. However, predicting attributes in images with large variations still remains a challenging problem. Several approaches have been proposed for visual attribute classification; however, most of them assume independence among attributes. In fact, to predict one attribute, it is often useful to consider other related attributes. For example, a pedestrian with *long hair* and *skirt* usually imply the *female* attribute. Motivated by this, we propose a novel pedestrian attribute classification method which exploits interactions among different attributes. Firstly, each attribute classifier is trained independently. Secondly, for each attribute, we also use the decision scores of other attribute classifiers to learn the attribute interaction regressor. Finally, prediction of one attribute is achieved by a weighted combination of the independent decision score and the interaction score from other attributes. The proposed method is able to keep the balance of the independent decision score and interaction of other attributes to yield more robust classification results. Experimental results on the Attributed Pedestrian in Surveillance (APiS 1.0) [1] database validate the effectiveness of the proposed approach for pedestrian attribute classification.

1 Introduction

The smart video surveillance technologies [2–4] including object detection [5, 6], object tracking [7] and object classification [8], have attracted more and more attentions in public security field. Pedestrian related technique is one of the hottest topic in this field. Pedestrian detection [9, 6], pedestrian tracking [10], behavior analysis [11] and clothing recognition [12] are all extensively studied in these years.

In this paper, we focus on pedestrian attribute classification [1] to present a more comprehensive description of pedestrian. As shown in Fig. 1, pedestrian attribute classification is to predict the presence or absence of several attributes. Pedestrian attributes used in surveillance application include *gender*, *hair*, *clothing appearance* and *carrying thing*, etc. The pedestrian attribute classification can be used to provide useful information for applications such as pedestrian tracking, re-identification [13] and retrieval, etc. As shown in Fig. 2, attributes such as *male*, *hand carrying* and *back bag* can be effectively applied to assist locating the desired target in the pedestrian retrieval application.

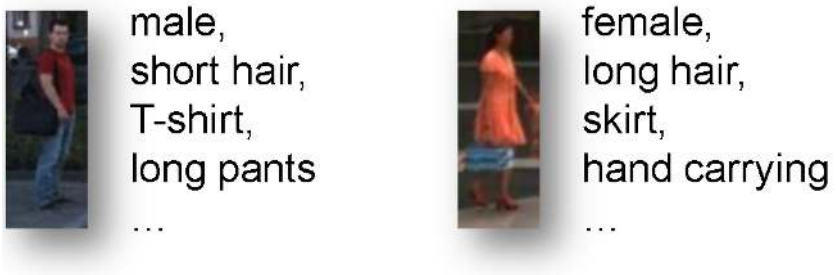

Fig. 1. Pedestrian attribute classification describes pedestrians with a list of visual attributes.

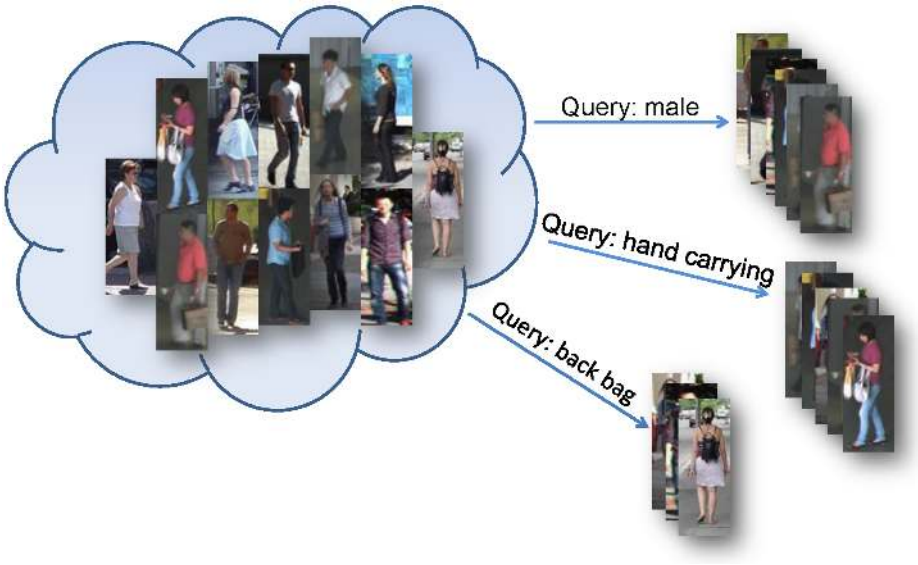

Fig. 2. Pedestrian retrieval based on attributes. Attributes such as *male*, *hand carrying* and *back bag* can be effectively applied to assist locating the desired target.

1.1 Related work

Attributes are powerful to infer high-level semantic knowledge. There are many computer vision applications based on attribute, such as face verification [14], object recognition [15], clothing description [16], image retrieval [17] and scene classification [18], etc. The successes of these applications rely heavily on the accuracy of predicted attribute values (i.e. the decision scores of separated attribute classifiers). Kumar et al. [14] used semantic attributes as mid-level features to aid face verification. In this application, the prediction model of each attribute on an input image is first learned, and the supervised object models on top of those attribute predictions are then built.

The most popular method for attribute classification is to extract low-level features from an image, and then train classifier for each attribute separately. Daniel et al. [19] proposed an attribute-based people searching system in surveillance environments. In this application, people are identified by a series of attribute detectors. Each attribute detector is independently trained from large amounts of training data. Layne et al. [13] utilized scores from 15 attribute classifiers as mid-level representations to aid person re-identification, where each attribute classifier is independently trained by the SVM algorithm. The main drawback of separately training each attribute classifier is that it ignores the interactions between different attributes which are helpful for improving classification performance. In fact, there is the interaction issue among pedestrian attributes. For instance, if the *long hair* and *skirt* of a pedestrian have shown, the attribute *male* is unlikely to appear.

In order to build interaction models among different attributes, Chen et al. [16] explored the mutual dependencies between attributes by applying a Conditional Random Field (CRF) with the SVM margins from the separately trained attribute classifiers. Each attribute function is used as a node and the edge connecting every two attribute nodes reflects the joint probability of these two attributes. Their CRF is fully connected, which means that all attribute nodes are pairwise connected. In the CRF, it is assumed that the observed feature f_i is independent of all other features once the attribute a_i is known. However, this assumption is not always held, because two attributes may appear in the same region.

Bourdev et al. [20] used SVM algorithm to explore interactions between different attributes. Firstly, each attribute classifier is separately trained by SVM algorithm on a set of Poselets [21]. Then, they used the SVM algorithm learning on the scores of all separately trained attribute classifiers to capture interactions between different attributes. In other words, the final decision score of an attribute is constructed by linearly combining all decision scores that come from separately trained attribute classifiers and the linear coefficients are learned by SVM. However, since an attribute is most relevant to itself, the final decision score of an attribute in this interaction model will heavily rely on the decision score of its own attribute classifier, resulting in the role of other attributes is ignorable.

For that, a more effective pedestrian attribute classification is proposed in this paper by exploiting interactions among different attributes. The proposed method linearly combines the independent decision score and the interaction score of an attribute to yield the final decision score of attribute classification. The independent decision score is produced by an independently trained classifier. The interaction score of an attribute is learned by using Lasso regression algorithm on all independent decision scores excluding its own independent decision scores. For each attribute, the proposed approach introduces a weight parameter to control the contribution of interaction score, achieving more robust classification results. Experiment results on the APiS 1.0 database show that the proposed approach can exploit the interactions among different attributes more effectively than the interaction model proposed in [20].

The remaining of this paper is organized as follows. Section 2 introduces the details of the proposed pedestrian attribute classification method. Section 3 shows the experiment results on APiS 1.0 database. Finally, Section 4 concludes the paper.

2 Modeling Attribute Interaction for Pedestrian Attribute Classification

2.1 Feature Extraction

We apply a sliding window strategy for feature extraction. In each sub-window, a joint color histogram, a MB-LBP histogram and a Histogram of Oriented Gradient (HOG) are extracted. The color histogram has 8, 3 and 3 bins in the H, S, and V color channels, respectively. The MB-LBP [22] histogram includes 30 bins, 10 from 3×3 scale descriptor, 10 from 9×9 scale one, and 10 from 21×21 scale one. For the HOG feature extraction, each sub-window is equally divided into 2×2 sub-regions, and in each sub-region a histogram of oriented gradient is extracted with 9 orientation bins. The HOG feature associated with each sub-window is obtained by concatenating the above four histograms into a 36-dimensional vector. The diagram of feature extraction is shown in Fig. 3. The details of feature extraction are described in [1].

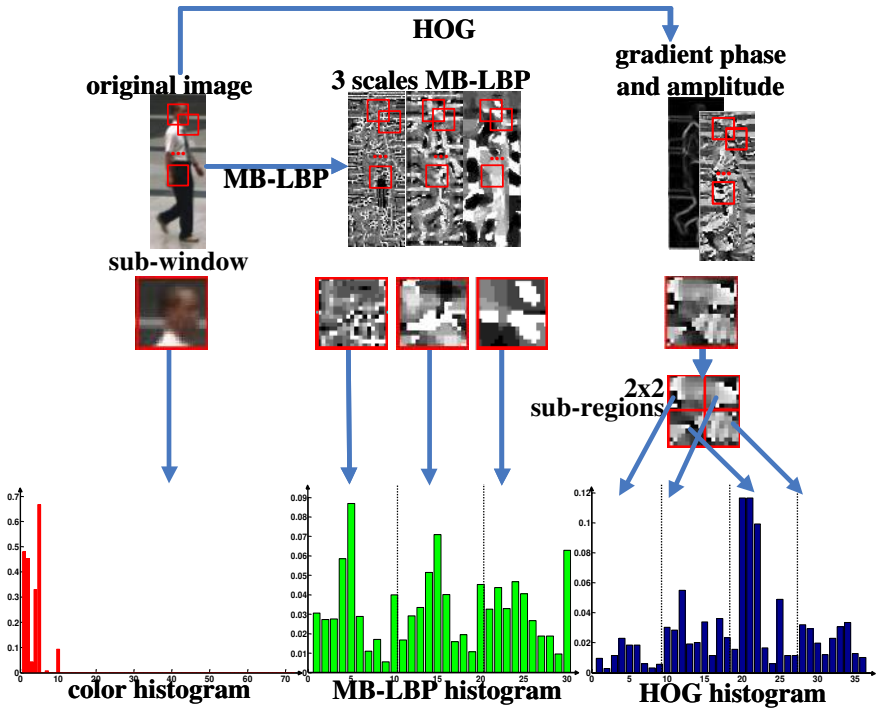

Fig. 3. The diagram of feature extraction.

2.2 Independent Attribute Classifier

In this work, the Gentle AdaBoost [23] algorithm is chosen to independently train each attribute classifier. More specifically, we first concatenate the color, MB-LBP and

HOG features, and then use Gentle AdaBoost to construct classifiers. We select the stump classifier with the minimum square error as the weak classifier in the Gentle AdaBoost algorithm.

2.3 Interaction Model

In order to exploit the interactions among different attributes, an interaction model is required. The interaction model firstly proposed in [20] is shown in Fig. 4. From this figure, we can find that each combined attribute classifier directly connected with all independent attribute classifiers. It means that each combined attribute classifier is constructed by linearly combining all independent attribute classifiers.

Assume that there are m attributes to predict; $\boldsymbol{x}$ is a testing sample; h_i is the i-th ($i \in \{1, 2, ..., m\}$) attribute classifier which is independently trained by using the Gentle AdaBoost algorithm; H_i represents the combined classifier for i-th attribute. H_i is calculated as follows:

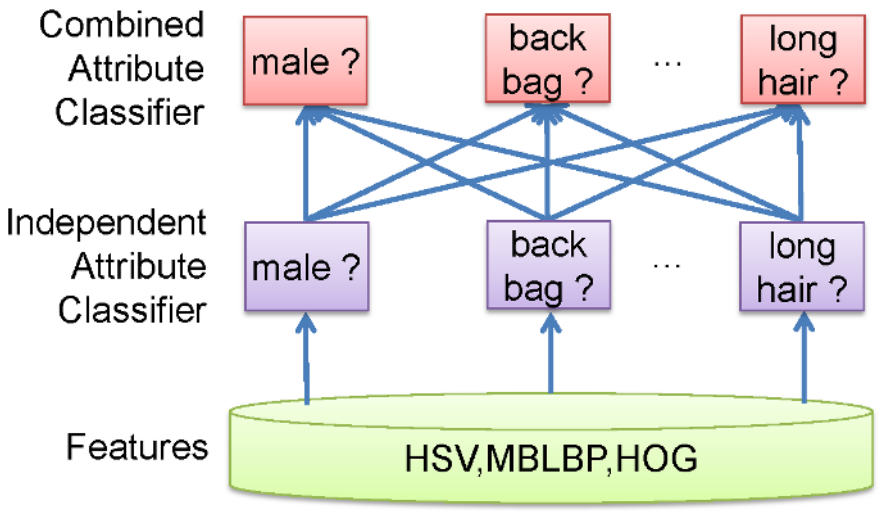

Fig. 4. The interaction model proposed in [20]. A combined attribute classifier is learned on the independent decision scores produced by all separated attribute classifiers.

$$H_i(\boldsymbol{x}) = \sum_{j=1}^{m} w_{ij} h_j(\boldsymbol{x}), \tag{1}$$

where the linear coefficients $\boldsymbol{w}_i = \{w_{ij}, j = 1, 2, ..., m\}$ can be learned by using the SVM algorithm. However, the problem is that an attribute is most relevant to itself, which may bring about such a fact that the combined decision score of an attribute in this method heavily relies on the independent decision score produced by its own

independent attribute classifier, thus ignoring the role of other attributes. This drawback will be validated in our following experiments.

To address this problem, we propose our interaction model, as shown in Fig. 5. Different from that in Fig. 4, an combined attribute classifier consists of an independent classifier and an interaction regressor trained on other independent attribute decision scores. We further introduce parameters to control the weight of the interaction regressors.

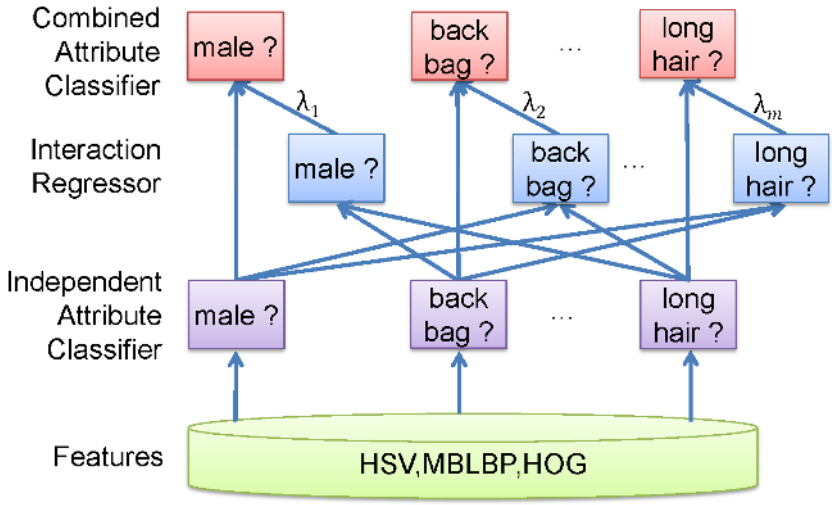

Fig. 5. The proposed interaction model. $\{\lambda_1, \lambda_2, ..., \lambda_m\}$ are used to control the weight of interactions.

In the proposed interaction model, H_i is learned as follows:

$$H_i(\boldsymbol{x}) = h_i(\boldsymbol{x}) + \lambda_i G_i(\boldsymbol{x}), \tag{2}$$

where

$$G_i(\boldsymbol{x}) = \sum_{j=1, j\neq i}^{m} w_{ij} h_j(\boldsymbol{x}). \tag{3}$$

From Eq. (2), we can find that the interaction regressor of the i-th attribute G_i only involves $m-1$ attribute classifiers, excluding the i-th attribute classifier h_i. This strategy can directly avoid the problem of the combined decision score relying on h_i too heavily and capture the interactions among the rest attributes. Meanwhile, the parameter λ_i is used to keep the balance between h_i and G_i. If $\lambda_i = 0$, H_i will degrade into h_i.

Generally speaking, one attribute may only be related to a part of the rest attributes. Therefore, the linear coefficients $\boldsymbol{w}_i = \{w_{ij}, j = 1, 2, ..., m \text{ and } j \neq i\}$ in Eq. (3) should be sparse. With consideration of this potential sparse characteristic, the following objective formulation is designed:

$$\boldsymbol{w}_i = \arg\min_{\boldsymbol{w}} \{\frac{1}{2}\|A_i\boldsymbol{w} - \boldsymbol{y}_i\|_2^2 + \gamma_i\|\boldsymbol{w}\|_1\}, \tag{4}$$

where A_i represents the independent decision scores of $\{h_j, j = 1, ..., m \text{ and } j \neq i\}$ on training set; $\boldsymbol{y}_i$ represents the i-th attribute labels of the training set; γ_i is a parameter used to control the sparseness of $\boldsymbol{w}_i$. The larger γ_i is, the more sparsely $\boldsymbol{w}_i$ will be. Assume that the training set includes n samples, then A_i is organized as a matrix with $n \times (m-1)$ dimensions and $\boldsymbol{y}_i$ is a column vector with $n \times 1$ dimensions.

Eq. (4) is the Lasso [24] problem, which formulates the least-square estimation problem with l_1-norm penalty to approximate the sparse representation solution. If $\gamma_i = 0$, Eq. (4) will degrade into the least-square estimation problem. In our implementation, the Dual Augmented Lagrangian (DAL) algorithm [25] is used to solve Eq. (4). In our experiments, we determinate λ_i and γ_i with cross validation, because cross validation is a simple and general way used to find appropriate parameters.

3 Experiments

To evaluate the performance of the proposed interaction model, it is compared with the baseline [1] and the interaction model proposed in [20] on Attributed Pedestrian in Surveillance (APiS 1.0) [1] database under the same evaluation protocol. The APiS 1.0 database has 3661 images, and each image is labeled with 11 binary attribute annotations. The linear coefficients $\boldsymbol{w}$ in Eq. (1) and Eq. (3) are learned by the Lasso algorithm. Since APiS 1.0 database does not provide validation sub-set, we randomly divide the APiS 1.0 database into 5 equal sized sub-sets (the partition is different with that in [1]) for the selection of λ and γ. Specially, the best pair of parameters is the one whose corresponding result has the largest Area Under ROC curve (AUC). Based on the color, MB-LBP and HOG features extraction described in Section 2.1, each attribute classifier independently trained by the Gentle AdaBoost algorithm includes 3,000 weak classifiers as suggested in [1]. Note that both the interaction model proposed in [20] and our proposed interaction model are built on attribute scores predicted from the same feature representation.

3.1 Average Recall Rate Comparison

Table 1 lists the comparison of average recall rates when the average false positive rates are 0.1. We can find that there are only 3 of 11 attributes achieve higher average recall rates than the baseline method [1] when using the interaction model proposed in [20], and the biggest improvement increased by only 2.78% recall rate for *long pants* attribute. However, the proposed model offers higher average recall rates for 9 of 11 attributes and 6 of them have obvious improvements (1.90% for *M-S pants*, 8.45% for *long pants*, 6.25% for *skirt*, 3.04% for *male*, 3.18% for *long hair* and 4.47% for *S-S bag*).

From Table 1, we can find that the proposed method fails to improve the average recall rates of *long jeans* and *hand carrying* attributes; however, the proposed method has equal performance with the baseline method [1] for *hand carrying* attribute and

Table 1. The comparison of average recall rates when the average false positive rates are 0.1. Where *M-S pants* is the abbreviation of Medium and Short pants, and *S-S bag* is the abbreviation of Single Shoulder bag.

attribute	recall rate(%)		
	baseline [1]	interaction model in [20]	the proposed
long jeans	**89.85**	89.74	89.18
M-S pants	78.65	78.65	**80.55**
long pants	76.68	79.46	**85.13**
skirt	68.23	67.71	**74.48**
male	58.30	57.53	**61.34**
back bag	56.16	56.16	**56.51**
T-shirt	55.22	56.36	**56.47**
long hair	55.15	55.15	**58.33**
shirt	54.62	54.22	**54.82**
hand carrying	**52.14**	**52.14**	**52.14**
S-S bag	38.45	39.64	**42.92**
average improvements(%)	-	0.30	**2.58**

very minor degradations for *long jeans* attribute. These results validate that the strategy of introducing parameters to control the weight of interaction score is robust. Though it may not improve performance for some attributes, it will not cause significant performance degradation. In addition, compared with the baseline method, the average improvements in recall rates obtained by the interaction model proposed in [20] and the proposed model are 0.30% and 2.58%, respectively. We also tried a least square solution of our model, corresponding to $\gamma_i = 0$ in Eq. (4). As a result, the sparse solution has a marginal mean accuracy improvement (0.47%) over the least square solution. Nevertheless, the Lasso model has a better interpretation of finding the most relevant interaction between attributes. To sum up, our method can achieve better recall rate performances compared with the interaction model proposed in [20].

3.2 Average ROC Comparison

Fig. 6 compares the average ROC curves of all attribute defined in APiS 1.0 database. The AUC sum of 11 attributes obtained by the baseline method [1], the interaction model proposed in [20] and the proposed model are 9.5371, 9.5372 and 9.5935, respectively. This shows that the interaction model proposed in [20] almost does not obtain performance improvements with respect to the baseline method [1], while the proposed method achieves performance improvements. It can be seen that the proposed model offers larger AUC for 9 of 11 attributes and 5 attributes (*long pants*, *M-S pants*, *skirt*, *long hair* and *male*) obtain obvious improvements. For *back bag* and *shirt* attributes, the proposed method fails to improve their performances. However, the propose method has equal performance with the baseline method [1] for *back bag* attribute and very minor degradations for *shirt* attribute. These results also validate that the strategy of introducing parameters to control the weight of interaction part is robust.

3.3 Interaction Analysis

We provide visualizations of the coefficients learned by the interaction model proposed in [20] and our method, as shown in Fig. 7 and Fig. 8. From Fig. 7, we can find that each attribute is most relevant to itself, because the corresponding absolute coefficient has the maximum value. This indicates that the combined decision score of a given attribute heavily relies on the independent decision score produced by its corresponding attribute classifier, thus ignoring the role of other attributes.

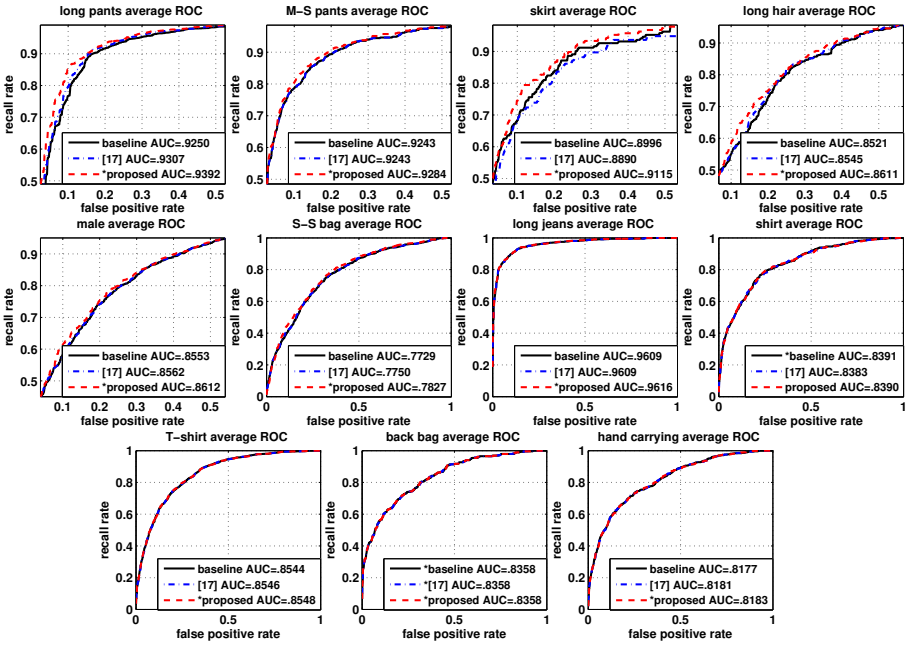

Fig. 6. The comparison of average ROC curves. Where *M-S pants* is the abbreviation of Medium and Short pants, and *S-S bag* is the abbreviation of Single Shoulder bag; '*' indicates the corresponding average ROC curve has maximum AUC value.

As shown in Fig. 8, in our model, the interaction part of an given attribute excludes the corresponding attribute classifier and only involves 10 other attribute classifiers. This strategy effectively avoids the situation that the combined decision score of a given attribute heavily relies on its independent decision score, and then capture the interactions from the rest attributes. Therefore, from Fig. 8 we can find that *long jeans* attribute is highly positive correlated with *long pants* attribute; *long hair* attribute is highly negative correlated with *male* attribute and *skirt* attribute is highly positive correlated with *longhair* attribute.

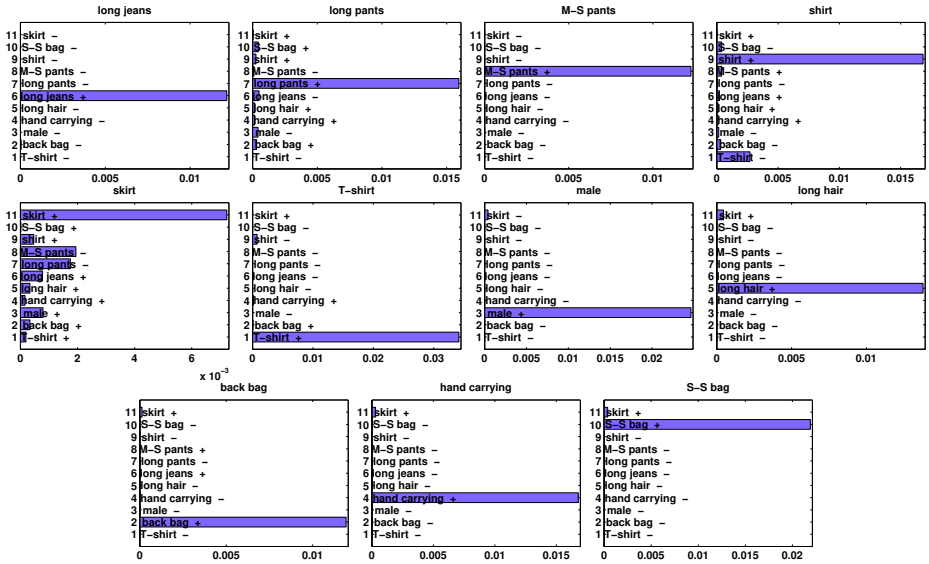

Fig. 7. The absolute coefficients learned by the interdependency model proposed in [20].Where *M-S pants* is the abbreviation of Medium and Short pants, and *S-S bag* is the abbreviation of Single Shoulder bag; '+' and '-' represent the two attributes hold positive correlation and negtive correlation relationship, respectively.

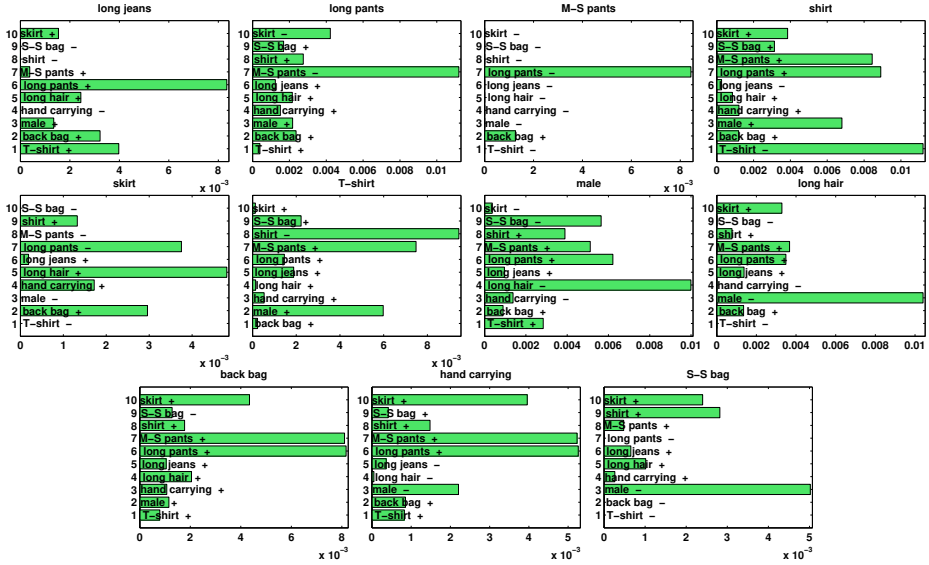

Fig. 8. The absolute coefficients learned by the proposed model. Where *M-S pants* is the abbreviation of Medium and Short pants, and *S-S bag* is the abbreviation of Single Shoulder bag; '+' and '-' represent the two attributes hold positive correlation and negtive correlation relationship, respectively.

4 Conclusion

This paper has proposed a novel method for pedestrian attribute classification which exploits interactions among different attributes. In our method, prediction of one attribute is achieved by a weighted combination of the independent decision score and the interaction score. The independent decision score of an attribute is obtained from a classifier independently trained by using the Gentle AdaBoost algorithm. The interaction score of an attribute is obtained from a regressor trained by using Lasso algorithm on all the rest independent decision scores. The proposed method further introduces weight parameters to keep the balance of the independent decision score and the interaction score. Experimental results on APiS 1.0 database have shown that the interactions among different attributes have effectively improved the attribute classification performance.

Acknowledgement. This work was supported by the Chinese National Natural Science Foundation Projects #61105023, #61103156, #61105037, #61203267, #61375037, #61473291, National Science and Technology Support Program Project #2013BAK02B01, Chinese Academy of Sciences Project No. KGZD-EW-102-2, and AuthenMetric R&D Funds.

References

1. Zhu, J., Liao, S., Lei, Z., Yi, D., Li, S.Z.: Pedestrian attribute classification in surveillance: Database and evaluation. In: In ICCV workshop on Large-Scale Video Search and Mining. (2013)
2. Shu, C.F., Hampapur, A., Lu, M., Brown, L., Connell, J., Senior, A., Tian, Y.: IBM smart surveillance system (s3): a open and extensible framework for event based surveillance. In: IEEE Conference on Advanced Video and Signal Based Surveillance. (2005) 318–323
3. Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H., Pankanti, S.: Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. IEEE Signal Processing Magazine **22** (2005) 38–51
4. Sedky, M.H., Moniri, M., Chibelushi, C.C.: Classification of smart video surveillance systems for commercial applications. In: IEEE Conference on Advanced Video and Signal Based Surveillance. (2005) 638–643
5. Gu, C., Arbeláez, P., Lin, Y., Yu, K., Malik, J.: Multi-component models for object detection. In: European Conference on Computer Vision. (2012) 445–458
6. Yan, J., Lei, Z., Yi, D., Li, S.Z.: Multi-pedestrian detection in crowded scenes: A global view. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 3124–3129
7. Yang, B., Nevatia, R.: Online learned discriminative part-based appearance models for multi-human tracking. In: European Conference on Computer Vision. Springer (2012) 484–498
8. Hariharan, B., Malik, J., Ramanan, D.: Discriminative decorrelation for clustering and classification. In: European Conference on Computer Vision. Springer (2012) 459–472
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition. (2005) 886–893
10. Guan, Y., Chen, X., Wu, Y., Yang, D.: An improved particle filter approach for real-time pedestrian tracking in surveillance video. In: International Conference on Information Science and Technology Applications, Atlantis Press (2013)
11. Jackson, S., Miranda-Moreno, L.F., St-Aubin, P., Saunier, N.: A flexible, mobile video camera system and open source video analysis software for road safety and behavioural analysis. (In: Transportation Research Board 92nd Annual Meeting)

12. Yang, M., Yu, K.: Real-time clothing recognition in surveillance videos. In: IEEE International Conference on Image Processing. (2011) 2937–2940
13. Layne, R., Hospedales, T., Gong, S., Mary, Q.: Person re-identification by attributes. In: British Machine Vision Conference. (2012)
14. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: IEEE Conference on International Conference on Computer Vision. (2009) 365–372
15. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.: Describing objects by their attributes. In: IEEE Conference on Computer Vision and Pattern Recognition. (2009) 1778–1785
16. Chen, H., Gallagher, A., Girod, B.: Describing clothing by semantic attributes. In: European Conference on Computer Vision. (2012) 609–623
17. Yu, F.X., Ji, R., Tsai, M.H., Ye, G., Chang, S.F.: Weak attributes for large-scale image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 2949–2956
18. Li, L.J., Su, H., Lim, Y., Fei-Fei, L.: Objects as attributes for scene classification. In: Trends and Topics in Computer Vision. Springer (2012) 57–69
19. Vaquero, D., Feris, R., Tran, D., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: IEEE Workshop on Applications of Computer Vision. (2009)
20. Bourdev, L., Maji, S., Malik, J.: Describing people: A poselet-based approach to attribute classification. In: IEEE International Conference on Computer Vision. (2011) 1543–1550
21. Bourdev, L., Malik, J.: Poselets: Body part detectors trained using 3d human pose annotations. In: IEEE International Conference on Computer Vision. (2009) 1365–1372
22. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: Advances in Biometrics. Springer (2007) 828–837
23. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The Annals of Statistics **28** (2000) 337–407
24. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288
25. Tomioka, R., Suzuki, T., Sugiyama, M.: Super-linear convergence of dual augmented-lagrangian algorithm for sparsity regularized estimation. arXiv preprint arXiv:0911.4046 (2009)