Regularized Transfer Boosting for Face Detection Across Spectrum

Zhiwei Zhang, Dong Yi, Zhen Lei, and Stan Z. Li, *Fellow, IEEE*

***Abstract*—This letter addresses the problem of face detection in multispectral illuminations. Face detection in visible images has been well addressed based on the large scale training samples. For the recently emerging multispectral face biometrics, however, the face data is scarce and expensive to collect, and it is usually short of face samples to train an accurate face detector. In this letter, we propose to tackle the issue of multispectral face detection by combining existing large scale visible face images and a few multispectral face images. We cast the problem of face detection across spectrum into the transfer learning framework and try to learn the robust multispectral face detector by exploring relevant knowledge from visible data domain. Specifically, a novel Regularized Transfer Boosting algorithm named R-TrBoost is proposed, with features of weighted loss objective and manifold regularization. Experiments are performed with face images of two spectrums, 850 nm and 365 nm, and the results show significant improvement on multispectral face detection using the proposed algorithm.**

***Index Terms*—Face detection, multispectral, transfer boosting.**

I. Introduction

RECENTLY, multispectral face biometrics has been proposed as a novel insight into the face biometrics. By providing active illuminations, the face recognition process no longer suffers uncertain environmental illumination risks. In [1] Li *et al.* proposed a highly accurate near-infrared (NIR at 850 nm) based face recognition system, and has been successfully utilized in many real world applications. Chang *et al.* [2] showed than much more facial information can be obtained if more than one spectrums are used. Moverover, multispectral imaging is inborn robust to spoofing attacks, and in [3], [4] promising results are given. Based on the above advantages, multispectral face biometrics is surely a promising research area in future.

However, the first and fundamental step, the face detection, has rarely been studied in the field of multispectral face biometrics. Some works [1], [7] utilized the same methodology as in visible face detection, while many other works [5], [6] simply omitted it. Currently, Boosting [20] based Viola-Jones framework [11] dominates the field of face detection, and numerous variants in either framework or feature types, such as Nesting [8], Vector Boosting [9], MBLBP [10], have been proposed. Despite the success, however, they all require a huge number of training data collection, which is not only time consuming but also cost expensive. For instance, researches [1], [11] collected thousands, or even tens of thousands of images for the training, which requires huge manual work on data collection and sample labeling. For multispectral face biometrics, unfortunately the progress is still initial and only scarce data are accessible, for instance, in [2] only dozens of people are collected, which prohibits the face detector training. Considering the visible face detection is quite mature with many accessible face databases(such as FRGC, FERET), it is reasonable to ask could we borrow the strength from these visible face data for the usage in multispectral case so that the expensive data collection could be reduced?

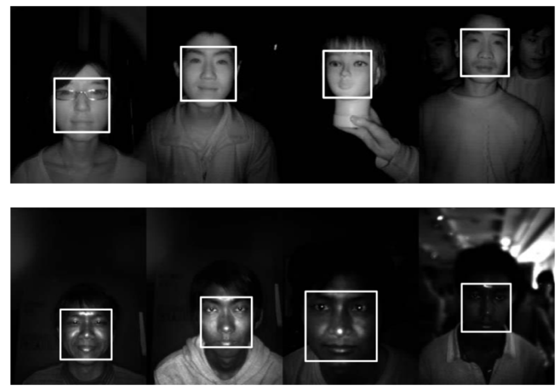

Fig. 1. Some detection examples. The first row is the 850 nm detection results and the second row is the 365 nm detection results.

The answer is yes. In this letter we propose to combine the abundant and available visible face data with the few multispectral face data. Specifically we design a novel regularized transfer boosting algorithm (named R-TrBoost), which features two aspects: 1) Data of different domains are weighted differently, so that in the optimization, more attention is paid on the target data domain (in our case, multispectral data); 2) manifold regularization is imposed to achieve label/score smoothness among target data, which prevents an overfitting problem trained on the few target data. Experiments on two spectrum, 850 nm and 365 nm, clearly prove the effectiveness of the proposed R-TrBoost algorithm. These two spectrums are at two ends of the visible spectrum (400 nm–750 nm), and compared with visible faces, very distinctive facial appearances can be observed (see Figs. 1 and 2). Although [1] reported excellent detection results on 850 nm, up to 178 000 samples were collected for training. In our experiments we show that with only few accessible data, our algorithm can greatly improve the performance, both on 850 nm and 365 nm.

Manuscript received August 20, 2011; revised September 23, 2011; accepted October 01, 2011. Date of publication October 18, 2011; date of current version January 16, 2012. This work was supported by the Chinese National Natural Science Foundation under Projects 61070146, 61105023, 61103156, 61105037, and by the TABULA RASA project (http://www.tabularasa-euproject.org) under the Seventh Framework Programme for research and technological development (FP7) of the European Union (EU) under Grant 257289. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiao-Ping Zhang.

The authors are with the Center for Biometrics and Security Research & National Laboratory of Pattern Recognition, CASIA, Beijing, China (e-mail: zwzhang@cbsr.ia.ac.cn; dyi@cbsr.ia.ac.cn; zlei@cbsr.ia.ac.cn; szli@cbsr.ia.ac.cn).

Digital Object Identifier 10.1109/LSP.2011.2171949

Fig. 2. Some face images used for training, from top to bottom are: visible face images, 850 nm images and 365 nm images. Distinctive appearances can be easily observed, especially from 365 nm.

The issue we tackle is essentially a transfer learning problem [12], which focuses on the situation when training and testing data are drawn from different distributions. Methods such as [13], [14] combine data of different data domains for knowledge transfer. Some other works [15], [16] directly adapted a pre-trained classifier in target data domain, known as domain adaptation. For the Boosting based transfer algorithms, however, we are only aware of Dai's work [17], which updates the sample weights in different ways for each distribution while remaining the objective as the ordinary Boosting algorithms. [18] further extended Dai's work by using multiple data sources. Compared with Dai's work, our formulation is more biased to multispectral data, and therefore achieves better results which is proved by both 850 nm and 365 nm experiments.

The remaining of the letter is organized as follows: in Section II, we give the detailed procedures of proposed regularized transfer boosting (R-TrBoost) algorithm; In Section III we give experiments on the visible-850 nm and visible-365 nm cases, which prove the effectiveness of our algorithm. In Section IV, we conclude the letter.

II. Regularized Transfer Boosting

Currently almost all of the Boosting algorithms share the same optimization problem which minimizes the following Exponential Criteria loss function

$$Loss(x,y,F) = E[e^{-y(x)F(x)}] \tag{1}$$

where x is the training data, $y(x) \in \{+1,-1\}$ is data label, and F is the final strong classifier. At each iteration t, F_t is updated by adding a new weak learner f to the current F as

$$F_{t+1} \leftarrow F_t + f \qquad s.t. \quad f = \arg\min E[e^{-y(F_t+f)}]. \tag{2}$$

As mentioned above, (1) treats all data equally due to the assumption that all data are drawn from the same distribution. However, the assumption doesn't hold in many real world applications as well as the issue we concern. Consequently the loss function becomes inappropriate and so does the derived strong classifier $F(x)$.

We now generalize the Boosting into the regularized transfer learning version. Suppose we have the large amount of visible face data, $x_i^v \in D^v, i = 1:N$, and small amount of multispectral face data $x_i^m \in D^m, i = 1:M$. As the detection accuracy on multispectral data is our concern, it is reasonable to impose a higher penalty if a multispectral data is misclassified. Furthermore, as there are only few multispectral face data, the overfitting problem should be considered. Therefore we impose a manifold regularization item, which forces the label/score smoothness among sample neighborhood.

Based on the above principles, the loss function of R-TrBoost is formulated as below:

$$\begin{aligned} Loss(x,y,F) &= E[b(x)e^{-y(x)F(x)}] \\ &\quad + \lambda \sum_{x_i^m, x_j^m} S_{x_i^m,x_j^m}[F(x_i^m) - F(x_j^m)]^2 \end{aligned} \tag{3}$$

where $b(x)$ is the loss weight over different data and S are the similarity scores between sample pairs of multispectral data.

The first term reflects the misclassification loss on the combination of x^v and x^m. By proper setting $b(\cdot)$, the optimization is more biased towards x^m. The second term is the manifold regularization penalty, which gives a high penalty if two similar multispectral data are assigned two dissimilar scores (thus the label). By such a penalty, the label/score smoothness can be achieved among the sample neighborhood. λ controls the balance between the two terms.

The above optimization, however, is difficult to solve as the second penalty term cannot be optimized on x directly. Therefore a relaxed form of (3) is proposed for the sake of optimization as follows:

$$F = \arg\min E[b(x)e^{-y(x)F(x)}] \tag{4}$$

where

$$\begin{aligned} &F_{t+1} \leftarrow F_t + f + \lambda g \\ &f = \arg\min E[b(x)e^{-y(x)(F_t(x)+f(x))}] \\ &g = \\ &\arg\min \sum_{x_i^m,x_j^m} S_{x_i^m,x_j^m}[(F(x_i^m)+g(x_i^m))-(F(x_j^m)+g(x_j^m))]^2. \end{aligned} \tag{5}$$

The original formulation is parsed into two separate subproblem, and the final F is the combination of f and g. Although it is not the global optimum for the original (3), numerically we can still get a suboptimal solution.

A. Misclassification Term for Weighted Loss

For the first misclassification term, we adopt adaptive Newton method for solution. The first and second derivatives of (4) is:

$$\begin{aligned} \left.\frac{\partial J(F(x)+f(x))}{\partial f(x)}\right|_{f(x)=0} &= -E(ybe^{-yF}) \\ \left.\frac{\partial^2 J(F(x)+f(x))}{\partial (f(x))^2}\right|_{f(x)=0} &= E(be^{-yF}). \end{aligned} \tag{6}$$

Thus the Newton update is

$$f = -\frac{F'(x)}{F''(x)} = \frac{E(ybe^{-yF}|x)}{E(be^{-yF}|x)}. \tag{7}$$

In our specific concern of face detection across spectrum, it is sufficient that $b(x)$ has only two possible values for our two distributions, and without loss of generality, $b(x)$ is constructed as

$$b(x) = \begin{cases} 1, & \text{if } x \in D^v \\ \beta, & \text{if } x \in D^m \\ 1, & \text{if } x \text{ is nonface} \end{cases} \tag{8}$$

where generally $\beta > 1$. Obviously if $\beta = 0$ then only visible face data are trained; if $\beta \to \infty$ then only multispectral face data are trained. As nonface data is generated by randomly sampling patches in images with no face, it is equivalent for both visible and multispectral data. Therefore we specify the $b(x) = 1$ if x is nonface data.

B. Score Inconsistency Term for Manifold Regularization

For similar samples, it is reasonable to argue that they should share similar classification scores, and such neighborhood preserving property is known as manifold regularization [19]. We firstly define the similarity matrix S in (4) as

$$S(x_i, x_j) = \begin{cases} 1, & \text{if } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $N_k(x)$ indicates the k nearest neighbors of x. In this letter we empirically set $k = 2$ and use Euclidean distance as similarity measure.

Given F_t, the update g has

$$\begin{aligned}
&\sum_{x_i, x_j} S(x_i, x_j) \times [(F_t(x_i) + g(x_i)) - (F_t(x_j) + g(x_j))]^2 \\
&\quad = \sum S(x_i, x_j) \\
&\quad\quad \times \Bigg[(F_t(x_i) - F_t(x_j))^2 + 2(F_t(x_i) - F_t(x_j)) \\
&\quad\quad\quad \times (g(x_i) - g(x_j)) + (g(x_i) - g(x_j))^2 \Bigg].
\end{aligned} \tag{10}$$

Calculate the derivative of the above equation towards $(g(x_i) - g(x_j))$ (Notice: not on g!) and set it zero, and we have

$$g(x_i) - g(x_j) = -(F_t(x_i) - F_t(x_j)). \tag{11}$$

The equation has a very clear meaning: given strong classifier $F_t(x)$, the optimum g at this iteration should compensate the score differences assigned by $F_t(x)$. Therefore, one possible g can be formulated as follows:

$$g(x) = E((F(x_n) - F(x)) | x_n \in N_k(x)). \tag{12}$$

Based on the above analysis, the final update at each iteration can be obtained using (5), with a balance parameter λ. Although this is not the exact optimal solution for the original formulation (3), our experiments show that it can still produce an excellent result as will be shown in the experiment section.

C. Other Implementation Details for Face Detection

In this letter, we adopt the well known MultiBlock Local Binary Pattern (MBLBP) Operator as the feature in face detection [10]. In [10] MBLBP has been shown better performance than traditional Haar and LBP features, and has now been integrated into the OpenCV2.3. For weak classifier, as the MBLBP feature is non-numerical, traditional weak classifiers based on numerical values are inappropriate. We adopt the Lookup Table (LUT) for it. Cascade is also adopted to achieve both accuracy and efficiency. Readers are referred to [11] for more details about cascade, and we simply omit it here.

III. EXPERIMENTS

In this section, we give two main experiments covering 850 nm and 365 nm spectrum, to prove the effectiveness of the algorithm proposed in this letter: 850 nm and 365 nm are at two ends of the visible spectrum (400 nm–750 nm), at which skin has very different albedos [4]. Illumination at 850 nm is already a mature spectrum for face recognition, and our data mainly come from existing database. For wavelength at 365 nm, however, this is a brand new spectrum for face biometrics with fewer data available. This in turn proves the necessity of our algorithm to reduce the manual labor in data collection. See Fig. 2 for examples.

TABLE I
PARAMETER SETTING IN R-TRBOOST

β \ λ	0	0.005	0.01	0.02	0.04	0.08
1	0.81	0.81	0.84	0.84	0.82	0.81
3	0.85	0.88	0.88	0.88	0.88	0.85
5	0.87	0.91	0.9	0.89	0.91	0.89
10	0.89	0.89	0.90	0.89	0.91	0.92
15	0.90	0.91	0.92	0.92	0.91	0.91

TABLE II
THE EFFECT OF k

β/λ \ k	2	4	6	8
5/0.04	0.91	0.89	0.86	0.88
10/0.03	0.91	0.90	0.91	0.89
15/0.02	0.92	0.91	0.92	0.91

For all the following experiments, we compare the following methods.

1) R-TrBoost with visible and multispectral face images for training (R-TrBoost).
2) Traditional method [10] with visible and multispectral face images for training. This is a special case of R-TrBoost by setting $\beta = 1$ and $\lambda = 0.(\mathrm{VIS} + 850\ \mathrm{nm\ or\ VIS} + 365\ \mathrm{nm})$.
3) With mere visible face images for training (VIS) [10].
4) With mere multispectral face images for training (850 nm or 365 nm) [10].
5) Dai's Boosting [17] with visible and multispectral faces for training (Dai's).

A. VIS to 850 nm

We design two subexperiments for VIS to 850 experiment. The first subexperiment is about parameter setting. We simplify face detection into a binary classification problem on cropped face images, to choose proper parameters for later detection. Seven thousand visible and 400 850 nm face images are used for training, and all images are cropped by eye positions into the size of $20 * 20$. About 40 000 nonface data are randomly sampled at each stage during cascade training from images with no faces.

For the first subexperiment, the classification accuracies on another 1000 cropped 850 nm faces at various parameters are tested and shown in Table I, from which we can see that the proposed weighted loss and manifold regularization can both improve the classification accuracy. Notice when $\beta = 1, \lambda = 0$, it is the traditional method [10] treating visible and multispectral data equally.

We further test the effect of neighbor number k in the manifold regularization under several β and λ choices, as listed in Table II. We observe that most of the time k doesn't effect too much on the classification result, and for the sake of computation, our choice of $k = 2$ is reasonable.

The second subexperiment for 850 nm is about the detection in practical 850 nm images. Another 858 images without face localization or cropping, are used to test the detection performance. For simplicity, we just use several parameters from

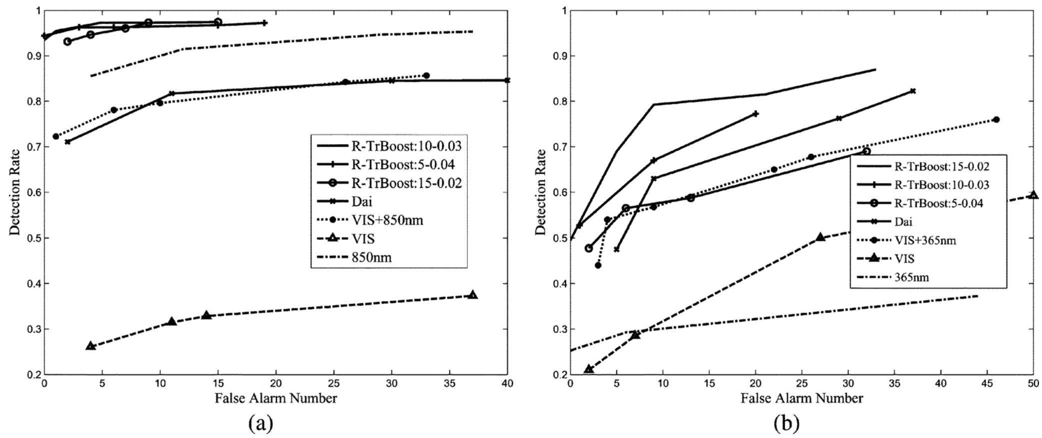

Fig. 3. (a) ROC curves for VIS to 850 nm experiment; (b) ROC curves for VIS to 365 nm experiment.

Table I. In Fig. 3 we give the ROC curves of different methods, which obviously proves the effectiveness of our method over other methods.

Notice that we deliberately use only 400 training images to simulate the situation where the multispectral face images are currently rare and expensive to collect. Meanwhile, we use much more images for test (up to thousands of 850 nm face images). The large amount of test data is a strong proof to the effectiveness of our R-TrBoost algorithm. Furthermore, in the next experiment we have only hundreds of 365 nm face images for experiment, indicating that it is necessary to give a test on a relatively large amount of 850 nm data if possible.

B. VIS to 365 nm

As the 365 nm imaging is still at the beginning of development, we have collected only 705 365 nm images for experiment. We use the same 7000 visible face images as above, and 305 face images of 365 nm for training. The detection is conducted on the other 400 face images of 365 nm. As there is no extra 365 nm face images for the parameter selection, we simply use the same parameters as in VIS to 850 nm detection experiment. The ROC curves can be seen in Fig. 3, which clearly proves the superiority of our algorithms. One noticeable point is that as face tend to exhibit quite distinctive appearance in 365 nm illumination, the accuracy for face detection is lower than that of 365 nm faces, which are more similar with visible face images.

From the above two experiments, we can see that neither large amount of visible data nor few multispectral data alone are sufficient for good detection. By combination, much higher detection rate can be achieved. Furthermore, by carefully designing the R-TrBoost algorithm and selecting proper parameters, our method achieves better result than other boosting algorithms.

IV. Conclusion

In this letter, we propose and tackle the issue of face detection across spectrum. The motivation is derived from the emerging research on multispectral face biometrics, which suffers the lack of sufficient face data at the beginning of research. We formulate a novel regularized transfer Boosting (R-TrBoost) algorithm by (1) adding weighted loss function into the optimization equation, and (2) adding manifold regularization to enforce label/score smoothness. The proposed R-TrBoost algorithm is proved to be effective in both VIS to 850 nm and VIS to 365 nm experiments, which covers two ends of the visible spectrum. Our future work will be the generalization of R-TrBoost into other related fields.

References

[1] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Patt. Anal. Mch. Intell.*, 2007.

[2] H. Chang, "Multispectral Imaging for Face Recognition Over Varying Illumination," Ph.D. dissertation, Univ. Tennessee, Knoxville, 2008.

[3] Y. Kim, J. Na, S. Yoon, and J. Yi, "Masked fake face detection using radiance measurements," *J. Opt. Soc. Amer.*, 2009.

[4] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face lIveness Detection by Learning Multispectral Reflectance Distributions," *AFGR*, 2011.

[5] R. Singh, M. Vatsa, and A. Noore, "Hierarchical fusion of multi spectral face images for improved recognition performance," *Inf. Fusion*, 2008.

[6] T. Bourlai, N. Kalka, A. Ross, B. Cukic, and L. Hornak, "Cross-spectral face verification in the short wave infrared (SWIR) band," in *ICPR*, 2010.

[7] D. Socolinsky, "Multispectral Face Recognition," in *Handbook of Biometrics*, A. K. Jain, P. Flynn, and A. A. Ross, Eds. New York: Springer, 2008.

[8] C. Huang, H. Ai, B. Wu, and S. Lao, "Boosting nested cascade detector for multi-view face detection," in *ICPR*, 2004.

[9] C. Huang, H. Ai, Y. Li, and S. Lao, "Vector boosting for rotation invariant multi-view face detection," in *ICCV*, 2005.

[10] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block lbp representation," in *ICB*, 2007.

[11] P. Viola and M. J. Jones, "Robust real-time face detection," in *IJCV*, 2004.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, 2009.

[13] H. Daumé, III and D. Marcu, "Domain adaptation for statistical classifiers," *J. Artif. Intell. Res.*, 2006.

[14] G. Qi, C. Aggarwal, and T. Huang, "Towards cross-domain knowledge propagation from text corpus to web images," presented at the WWW 2011 – Session: Multimedia, Hyderabad, India, 2011.

[15] J. Yang, R. Yan, and A. Hauptmann, "Cross-domain video concept detection using adaptive SVMs," *ACM Multimedia*, 2007.

[16] C. Zhang, R. Hamidz, and Z. Zhang, "Taylor expansion based classifier adaptation: Application to person detection," in *CVPR*, 2008.

[17] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Machine Learning,*. Corvallis, OR: , 2007.

[18] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in *CVPR*, 2010.

[19] K. Chen and S. Wang, "Semi-supervised learning via regularized boosting working on multiple semi-supervised assumptions," *IEEE Trans. Patt. Anal. Mach. Intell.*, 2011.

[20] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Statist.*, 2000.