FACIAL AGE ESTIMATION USING BIO-INSPIRED FEATURES AND COST-SENSITIVE ORDINAL HYPERPLANE RANK

Xiaohu Sun[1], Pingping Wu[2], Hong Liu[3]

[1]School of Information Science and Technology, Northwest University, Xi'an 710127, China
[2,3]Engineering Lab on Intelligent Perception for Internet of Things (ELIP), and Key Laboratory for Machine Perception, Peking University, Beijing 100087, China
xiaohusun430@sina.com, pingpingwu@pku.edu.cn(Corresponding Author), hongliu@pku.edu.cn

Abstract: Automatic age estimation relying on human facial images is a key technology of many real-world applications, which is still a challenging task in the computer vision field. There are three cascade modules for facial age estimation: facial aging feature extraction, dimension reduction (or feature selection) and estimation method. Many existing literatures focus on the first or last module while for an age estimation system, it's also important to construct a reasonable framework. Our work focuses on creating an effective framework by selecting methods for these modules reasonably. Firstly, a BIM (bio-inspired model) is employed to extract facial aging features because it can not only capture discriminative local and global features, but also overcome interferences of some 2D deformations to some extent. Then, LDA (linear discriminant analysis) is used for reducing the BIF (bio-inspired features) to lower dimensions and extracting more discriminative information at the same time. Finally, CS-OHRank (cost-sensitive ordinal hyperplane rank), which tackles with sparse data well and reflects the cumulative attributes of aging, is applied as the estimation method. Experimental results on benchmark dataset FG-NET show that our framework combining BIF, LDA and CS-OHRank is competitive among the state of the art, with MAE (mean absolute error) = 4.72 years. .

Keywords: Age estimation, Bio-inspired features, Linear discriminant analysis, Cost-sensitive ordinal hyperplane rank

1 Introduction

Age is one of human's most important attributes and its estimation is particularly useful in many scenarios. For age invariant face recognition [1,2], which can be used to identify missing children after many years later, age should be estimated accurately before synthesizing the missing children's image at the estimated age for recognition. Searching a criminal suspect in a large amount of surveillance monitoring videos manually is a tedious and time-consuming work. However, if videos can be filtered according to the age of the criminal suspect in advance, workloads will decrease sharply [11]. Other important potential applications of age estimation include: 1)Age-based human-computer interaction, e.g., Internet pages or material access control for juveniles, user-friendly interfaces for people at different age groups; 2)Image and video retrieval or saving according to age, which can apply not only to electronic albums, but also to search engines; 3)Demographic profiling or customer management systems, which can calculate numbers of people at different age groups in a mall, thus help managers to make right sale programs; 4)Video content comprehension because age is an important semantic information[6,7,11].

Recent years, facial age estimation has gained a lot of attention while many related literatures are published. Many models are proposed to extract facial aging features. Y. Kwon et al. [3] used a geometric model in his pioneering work, which calculates the ratios of the distances between some key facial landmarks to extract age information. The geometric model [3] works for young for craniofacial growth in human's early stage but not older for not considering texture information. And obviously, the model is sensitive to head rotation and location accuracy of face landmarks. AAM (active appearance model) [4] is a traditional facial aging features extraction model, which analyzes both shape and texture and has been used in many age estimation literatures[7,8,10,18,26]. A disadvantage of AAM is that it encodes the intensity of the whole image using PCA (principle component analysis), which cannot describe subtle changes of local areas in facial images. Recently, G. Guo et al. [5] proposed a bio-inspired model by modifying HMAX model [6] to extract facial aging features.

Many existing literatures view age estimation as regression or multi-classification problem or combination of these two [10,11]. G. Guo et al. [7] proposed LARR (locally adjusted robust regression) by combining SVR and SVM. The combination improves the age estimation accuracy, however, the LARR method cannot determine the local search range for the classifier SVM. G. Guo et al. [8] proposed a PFA (probabilistic fusion approach) that fuses the regression and classification probabilistic outputs, which can determine the search range for the classifier automatically. Mohamed Y. EI Dib [9] et al. used a much more complex combination of SVR and SVM to estimate ages

978-1-4799-4719-5 /14/$31.00 ©2014 IEEE

of facial images. It gets a very low MAE, however, the method has high complexity in computation, and it is almost impossible to find the optimal values through training samples for so many parameters in many SVMs and SVRs.

K. Chang et al. [10,11] viewed ages as relative orders and proposed using ranking approaches to estimate human age. Manifold learning methods, i.e., OLPP (orthogonal locality preserving projection) in [12], were also introduced to find low dimensions aging manifold from facial images. Although it performs well in the private large balanced Asian people dataset UIUC-IFP-Y [12], it's hard for OLPP to find an aging manifold when using benchmark dataset FG-NET [13] because contrary to the fundamental assumption of manifold learning, samples in FG-NET is small, sparse and imbalanced [14]. N. Fan [14] proposed an iterative algorithm to learn a distance metric function implemented by a compositional pattern producing network [15] for reconciling the discrepancy between the feature space and semantic space, i.e., the AAM feature space and age label space in [14], for age estimation. Chen et al. [16] introduced CA (cumulative attributes), an intermediate-level representation for age estimation. Firstly, low-level facial aging features is mapped onto the CA space using a multi-output regression function, and then ages are predicted based on cumulative attributes using common scalar-output regression [16]. Considering that aging is a personalized process, Geng et al. [17] introduced the AGES (aging pattern subspace), which is images of one person in chronological order. Zhang et al. [18] proposed the MTWGP (multi-task wrapped gaussian process) regression to learn common changes shared between persons and specific characteristics of each person for personalized age estimation.

2 Motivation

Facial aging feature extraction, dimension reduction (or feature selection) and estimation method are three cascade modules for facial age estimation. Many existing literatures [9,10,11,12,13,26] focus on the first or last module while for an age estimation system, it's also important to construct a reasonable framework. Our motivation derives from two seminal works [5,11]. G. Guo et al. [5] proposed BIM to extract facial aging features and used PCA to reduce the feature dimensions. LARR was used as the estimation method. Because the dimension of BIF is very high, about 8000 in [5], feature selection method is of great importance. In their work, performance deteriorates a lot when number of selected features, i.e., features after dimension reduction using PCA, is smaller than 300. The target of PCA is find a direction on which the projections of all data points have biggest variance, i.e., a direction that is effective for representation while the target of LDA is agree with fisher discrimination principle so LDA is more efficient for discrimination, e.g., Fisherface outperforms Eigenface in face recognition [19]. So LDA rather than PCA is used in our framework to get more discriminative information and lower dimensions. Another shortcoming in [5] is that LARR cannot determine the search range of the classifier automatically based on training sets but tries different search ranges manually, which is a crucial disadvantage when constructing an auto-system. Recently, K. Chang et al. [11] proposed a ranking approach called CS-OHRank for age estimation, which not only can determine optimal parameters automatically, but also has at least three advantages as follows: 1) Transforming age estimation problem to a series easy-processed binary problems; 2) Reflecting that human aging is a cumulative process; 3) Using full training samples in every binary problem, which mitigates the sparse problem of the FG-NET dataset. Thus, CS-OHRank is chosen as the method of the last estimation module. From above analyses, we propose a new age estimation framework, which as shown in Figure 1.

The rest of the paper is organized as follows. Section 3 elaborately describes each part of our age estimation framework. Section 4 gives experimental results and analyses. Finally, Section 5 concludes this paper.

3 Our age estimation framework

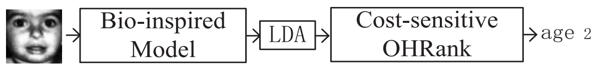

Figure 1 Our age estimation framework

In this section, each part of the proposed framework will be described elaborately except the commonly used LDA. Details of BIM and CS-OHRank are as follows.

3.1 Bio-inspired model

Bio-inspired model [5] derives from HMAX model [6], which has achieved excellent performance in objects detection [6]. HMAX is with four layers (S1, C1, S2 and C2) and mimics the responses of visual cortex for object detection, where S and C layers intimate the simple and complex cells in visual cortex respectively [6]. BIM uses S1 and C1 layer of HMAX and two modifications of original HMAX for age estimation are as follows: 1) Gabor filters with smaller sizes; 2) Nonlinear operation "STD" rather than "MAX" in C1 layer [5]. Feature extraction procedures of BIM are shown in Figure 2. Complete process rather than C1 layer only in [5] are shown for intuitively analysis. For clarity, only 5 orientations of the 3th scale bands are shown.

In S1 layer of BIM, multi-orientation and multi-scale filters can capture changes globally (e.g. craniofacial shape and facial features ratios) and locally (e.g., wrinkles on forehead and facial muscles' sagging and droop). As shown in Figure 2.a, filters with 36 degree orientation (in the middle) response strongly on eye bags' sloop and nasolabial grooves.

In C1 layer of BIM, through the scale and space neighborhood operation, BIM can overcome scale and shift deformation to some extent respectively [6,20]. Compared with the local MAX operation in the original

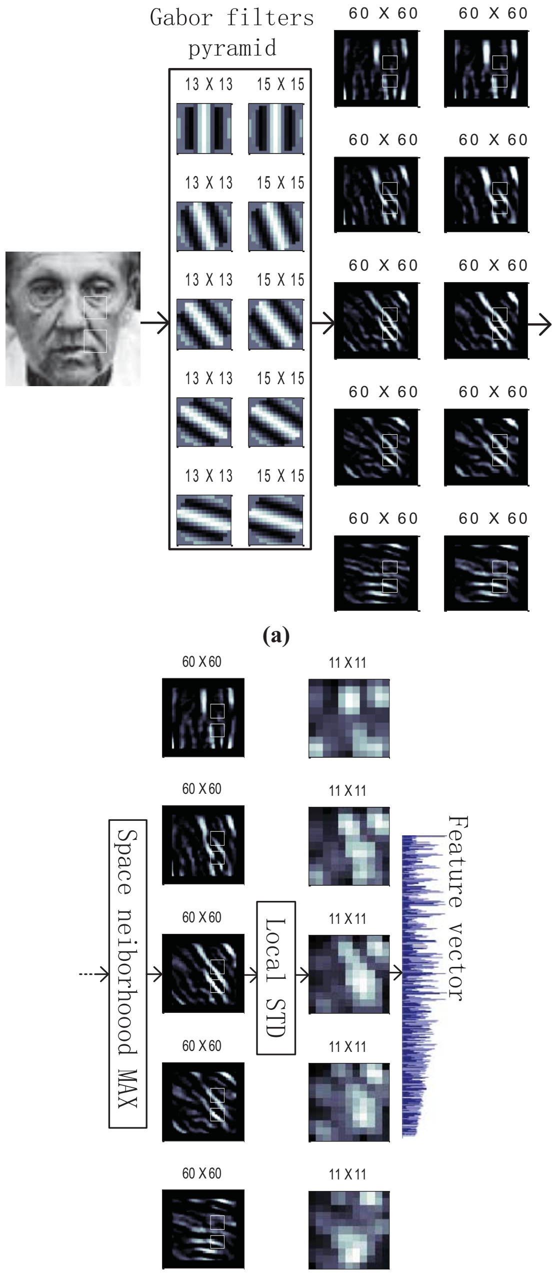

Figure 2 BIF's extracting process **(a)** C1 layer **(b)** S1 layer.

model HMAX [6, 20], local STD describes subtle differences better, which can be seen clearly in Figure 2.b.

3.2 Cost-sensitive ordinal hyperplane rank

Age is not a metric value, e.g., we cannot say that the ages of two 15 years old people are equal to the age of a 30 years old people, which is opposite to the prerequisite for regression. So regression methods, e.g., SVR cannot perform well for age estimation [11].

It is also difficult for multi-classification methods using the OVO (one versus one) strategy [25] to tackle with FG-NET dataset, which is heavily imbalanced in age distribution and has small and sparse samples, i.e., 1002 images of people with different races and genders, because of not enough samples for OVO training of multi-classification method. Additionally, multi-classification methods wrongly assume that age labels are independent from each other.

Ranking approaches treats ages as ordinal information [10,11]. OHRank [11] transforms the age estimation problem to a series of binary classification sub problems, i.e., older or younger than a certain age. For every binary classifier, all training samples are used for training it, which mitigates the sparse problem of FG-NET. Finally, all binary classification results are integrated using sum function.

Formulations are as follows. Firstly, age labels were transformed to corresponding ranks $\{1,2,\ldots,Q\}$. Denote $S=\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ as the training set of N samples with selected feature $\mathbf{x}_i$ and their associated ordinal rank y_i, $y_i \in \{1,\ 2,\ldots,\ Q\}$. Q-1 binary classifiers should be trained and we denote q^{th} binary classifier as $f_q(.)$.

For cost sensitive ordinal hyperplane rank, samples with different ages have different misclassification costs in every binary classification. Cost of misclassifying sample i in q^{th} binary classification are defined as

$$\mathrm{cost}_q(y_i) = |y_i - (q + 0.5)| \tag{1}$$

After introduced cost function (1), target function and constraint conditions of binary classification method, i.e., SVM in the experiments, are as follows [11].

$$\min_{\mathbf{w}_q, b_q, \varepsilon_q} \ \frac{1}{2}\mathbf{w}_q{}^T\mathbf{w}_q + \sum_{i=1}^{N} C_q \mathrm{cost}_q(y_i)\,\varepsilon_i$$

$$\text{s.t.} \begin{cases} z_q(y_i)\left(\mathbf{w}_q{}^T\emptyset(\mathbf{x}_i) + b_q\right) \geq 1 - \varepsilon_i \\ \varepsilon_i \geq 0 \end{cases} \tag{2}$$

In formula (2), C_q is a rescaling parameter and C_q should not be too large to prevent overwhelming other parameters in the target function. $Z_q(y_i) \in \{+1,\ -1\}$ denotes label of y_i in the q^{th} binary classification, i.e., $z_q(y_i) = +1$ if $y_i > q$ and -1 otherwise.

For a test sample $\mathbf{x}$, the q^{th} binary classification result $f_q(\mathbf{x})$=+1 if the q^{th} binary classifier predicts that the test sample's rank > q and 0 otherwise. Then rank of sample $\mathbf{x}$ is achieved by integrating all binary classification results using formula as follows,

$$r(\mathbf{x}) = 1+\sum_{q=1}^{Q} f_q(\mathbf{x}) \tag{3}$$

which reflects the cumulative attributes of aging. Finally, convert the rank to its corresponding age.

4 Experiment results and analyses

4.1 Dataset and Settings

Benchmark dataset FG-NET [13] and LOPO (Leave One Person Out) [21] test strategy are used in the experiments. Evaluation metric is the widely used criterion MAE [21], which is the average of absolute errors between estimated ages and ground truth ages. Before feature extraction, images are first normalized to 60×60 gray-level images.

AAM-API [22] is used to implement AAM. Variations saving ratios of PCA are all 0.95 for shape, intensity and combination of them. At last, the feature dimension of AAM is 43. As a benchmark face representation method, AAM is for comparison with BIF. BIM is with 8 scale bands and 10 orientations filters in the experiments. The original BIF dimension is 8940. After feature reduction by LDA, the dimension is 150.

OHRank is implemented using LIBSVM [23] and RBF kernel SVM is used as the binary classifier in every comparison. 5-fold cross validation is used to find the best penalty cost C in the target function and RBF kernel parameter γ. Search range of C and γ is $2^0, 2^{0.5}, 2^1, \dots, 2^{10}$ and $2^{-10}, 2^{-9.5}, 2^{-9}, \dots, 2^0$ respectively. For multiclass RBF kernel SVM and SVR, which are used for comparison, same optimal parameters search method and ranges are used.

Cost-sensitive OHRank is implemented using svm_learn.exe and svm_classify.exe of SVMlight [24]. We do not try 5-fold cross validation parameter optimization for every binary classifier, but only use SVMlight's default γ, and $\{C_q = 0.3\}_{q=1}^{Q-1}$. Actually, cost transformation technology can be used to make common binary classification algorithms, e.g., SVM, be cost-sensitive [25], and then common tool, e.g., LIBSVM [23] can be used to solve cost-sensitive binary classification problems.

4.2 Experimental results

Table I Comparison of different estimation methods and features

Age group	SVM		SVR		OHRank	
	AAM	BIF	AAM	BIF	AAM	BIF
0-9	2.95	2.18	3.33	4.13	2.18	2.48
10-19	6.51	3.27	2.97	4.71	3.27	3.29
20-29	10.65	7.73	5.50	5.70	5.72	5.38
30-39	17.44	15.85	14.03	8.88	11.11	10.13
40-49	27.22	23.89	22.09	13.32	19.04	15.65
50-59	37.07	35.60	32.00	19.69	26.27	21.80
60-69	43.38	43.38	41.76	31.93	38.25	33.38
MAE	8.35	6.25	5.96	5.80	**5.18**	**4.91**

Estimation results when features are AAM or BIF + LDA using different estimation methods are shown in Table I. Note that BIF denotes BIF+LDA in Table I. For AAM+SVR, linear kernel rather than RBF kernel was used in the experiment because we encountered over-fitting problem when using RBF kernel SVR.

As can be seen, OHRank outperforms SVM and SVR for either AAM or BIF+LDA features. Additionally, experimental results when using either SVR, SVM or OHRank all demonstrate that BIM's ability to extract facial aging feature is stronger than AAM.

In our experiments, LDA is used to select discriminative information from bio-inspired features and reduce the feature dimensions. The performance is good when selected features are 150D in our experiment while in the seminal work [5] using PCA, performance deteriorate obviously when dimensions are lower than 300.

Table II Comparison OHRank with CS-OHRank

Age group	OHRank	CS-OHRank
0-9	2.48	2.51
10-19	3.29	3.4
20-29	5.38	5.21
30-39	10.13	9.96
40-49	15.65	13.41
50-59	21.8	17.47
60-69	33.38	29
total	4.91	**4.72**

According to results in Table I, BIF+LDA+OHRank is a better framework compared with others in the table. After cost function (1) is introduced, the accuracy increases a lot as shown in Table II. Different from seminal work [11] using AAM, the feature in Table II is BIF+LDA, which has better aging feature extraction ability.

Table III MAE comparison of different methods

Method	MAE
AGES[17]	6.77
AAM + Rank[10]	5.79
AAM+RUN[26]	5.78
AAM+LARR[7,12]	5.07
AAM+PFA[8]	4.97
AAM+MTWGP[18]	4.83
BIF+PCA+LARR[5]	4.77
Our method	**4.72**

4.72 years in Table II is the minimum MAE we have got in experiments so far using the proposed framework BIF + LDA +CS-OHRank. Because of not doing complex parameters optimization, obviously 4.72 is not the actually minimum MAE. However, compared with most methods proposed so far as shown in Table III, our framework is competitive. Especially, our framework outperforms BIF+PCA+LARR in [5], which demonstrates that CS-OHRank is a more suitable estimation method for bio-inspired features.

Part of the estimation results are shown in Figure 3. The absolute errors are 0-6 in the figure. Our method performances well although encountering race or gender differences, expression changes, head rotation or bad quality of some images.

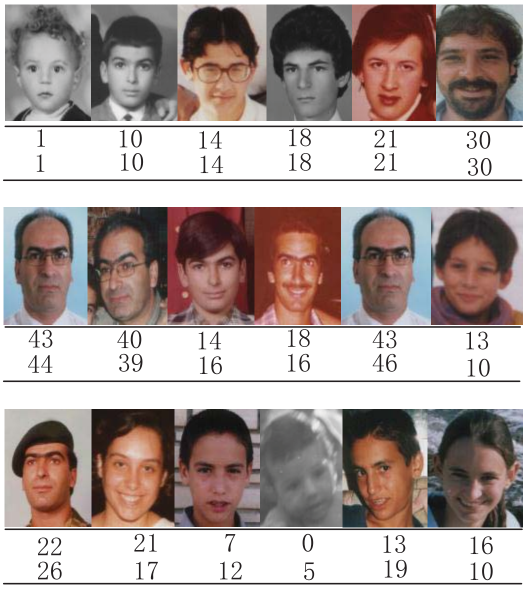

Figure 3 Part of estimation results (number upper and lower are ground true ages and estimated ages respectively)

5 Conclusions

In this paper, an age estimation framework with high accuracy has been proposed by choosing method for every module according to discrimination abilities, the robustness to inferences, e.g., 2D transformation, substance of variables, i.e., the attribute of age, et al., in which BIM is chosen to extract facial aging features rather than the geometric model and AAM, LDA is used as the feature selection method for BIF rather than PCA, ages are treated as relative-order information rather than metric values or classifications and cost functions are used in every sub problem of rank. Our analyses are based on gray images and further researches on color images can be done because skin color or hair color or age pigment also includes aging information.

References

[1] Gong D, Li Z, Lin D, Tang X. Hidden Factor Analysis for Age Invariant Face Recognition. In *ICCV*, pp. 2872- 2879, 2013.

[2] Park U, Tong Y, Jain A K. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32(5): 947-954, 2010.

[3] Kwon Y H, da Vitoria Lobo N. Age classification from facial images. *Computer Vision and Image Understanding*, vol. 74(1), pp.1-21, 1999.

[4] Cootes T F, Edwards G J, Taylor C J. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23(6), pp. 681-685, 2001.

[5] Mu G, Guo G, Fu Y, Huang, T S. Human age estimation using bio-inspired features. In *CVPR*, pp. 112-119, 2009.

[6] Serre T, Wolf L, Bileschi S, Poggio T. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29(3), pp. 411-426, 2007.

[7] Guo G, Fu Y, Huang T S, Dyer C. Locally adjusted robust regression for human age estimation. In *WACV*, pp.1-6, 2008.

[8] Guo G, Fu Y, Dyer C R, Huang T S. A probabilistic fusion approach to human age prediction. In *CVPRW*, pp.1-6, 2008.

[9] El Dib M Y, Onsi H M. Human age estimation framework using different facial parts. *Egyptian Informatics Journal*, vol. 12(1), pp. 53-59, 2011.

[10] Chang K Y, Chen C S, Hung Y P. A ranking approach for human ages estimation based on face images. In *ICPR*, pp. 3396-3399, 2010.

[11] Chang K Y, Chen C S, Hung Y P. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *CVPR*, pp. 585-592, 2011.

[12] Guo G, Fu Y, Dyer C R, Huang T S. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, vol. 17(7), pp. 1178-1188, 2008.

[13] The fg-net aging database, available at http://www-prima. imag.fr /FGnet/.

[14] Fan N. Learning nonlinear distance functions using neural network for regression with application to robust human age estimation. In *ICCV*, pp. 249-254, 2011.

[15] Stanley K O. Compositional pattern producing networks: A novel abstraction of development. *Genetic programming and evolvable machines*, vol. 8(2), pp. 131-162, 2007.

[16] Chen K, Gong S, Xiang T, Loy C C. Cumulative attribute space for age and crowd density estimation. In *CVPR*, pp.2467-2474, 2013.

[17] Geng X, Zhou Z H, Smith-Miles K. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29(12), pp.2234-2240, 2007.

[18] Zhang Y, Yeung D Y. Multi-task warped gaussian process for personalized age estimation. In *CVPR*, pp. 2622-2629, 2010.

[19] Cai D, He X, Han J, Zhang H J. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, vol. 15(11), pp. 3608-3614, 2006.

[20] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nature neuroscience*, vol. 2(11), pp.1019-1025,1999.

[21] Facial Image Process and Analysis (FIPA), http://face. cs. kit.edu/433.php

[22] Stegmann M B, Ersbll B K, Larsen R. FAME-a flexible appearance modeling environment. *IEEE Transactions on Medical Imaging,* vol. 22(10),pp. 1319-1331, 2003. Soft ware available at http://www2.imm.dtu.dk/~aam/aamapi/.

[23] Chang C C, Lin C J. LIBSVM: a library for support vector machinesl. *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol.2(3), pp. 27, 2011.

[24] Schölkopf B, Burges C J, and Smola A J (Eds.). Advances in kernel methods: support vector learning. *MIT press*,1999.

[25] Lin H T. From ordinal ranking to binary classification . Doctoral dissertation, California Institute of Technology, 2008.

[26] Yan S, Wang H, Tang X, Huang T S. Learning auto-structured regressor from uncertain nonnegative labels. In *ICCV,* pp.1-8, 2007.