Age Estimation using Active Appearance Models and Support Vector Machine Regression

Khoa Luu[1], Karl Ricanek Jr.[2] *Senior Member, IEEE*, Tien D. Bui[1], Ching Y. Suen[1]

***Abstract*—In this paper, we introduce a novel age estimation technique that combines Active Appearance Models (AAMs) and Support Vector Machines (SVMs), to dramatically improve the accuracy of age estimation over the current state-of-the-art techniques. In this method, characteristics of the input images, face image, are interpreted as feature vectors by AAMs, which are used to discriminate between childhood and adulthood, prior to age estimation. Faces classified as adults are passed to the adult age-determination function and the others are passed to the child age-determination function. Compared to published results, this method yields the highest accuracy recognition rates, both in overall mean-absolute error (MAE) and mean-absolute error for the two periods of human development: childhood and adulthood.**

I. INTRODUCTION

The determination of the age of a person from a digital photography is an intriguing problem. It involves an understanding of the human aging process, the biomechanical factors that influence the general patterns of aging that the idiosyncratic nature of aging, which is evident in the facial aging differences of identical twins. This paper will present an overview of the prior works in age estimation (determination) and a novel approach based on a hierarchical model, which infuses a classification system with multiple age estimator functions to create an *industry age-estimation algorithm*.

It is a well-known fact that the biodynamic factors of facial aging are quite different for the two stages of aging: growth and development and adult. During the latter, the major changes in facial complex are due to lengthening and widening factors of cranial complex [2]. The aging factors for adults, ~ 21 years and older, does include some cranial changes, but the primary drivers are the development of wrinkles, lines, creases, and sagging of the skin [1, 3]. As with synthetic face aging techniques, great care should be taken to separate the aging process into growth and development and adults as the factors that contribute to the changes that are being modeled are vastly different [1, 3].

This work presents an approach that, until now, has not been utilized for age estimation, which is to develop a unique aging functions based on the two different aging categories. A system is developed that exploits a single set of feature vectors for age group classification and aging function based on AAM and SVM/SVR.

[1]Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University, Canada ({kh_lu, bui, suen}@cse.concordia.ca).

[2]Department of Computer Science, University of North Carolina Wilmington, North Carolina, USA (ricanekk@uncw.edu).

II. PREVIOUS WORK

One of the first papers related to age estimation from digital images of the face belongs to Kwon & Lobo (1993) [5]. They used 47 high resolution images of a face for classification of the images into one of three age groups: babies, young adults or senior adults. Their approach was based on geometric ratios of key face features and wrinkle analysis. Their experiment resulted in 100% accuracy. However, their database was limited to 47 high resolution photos with that would be very difficult to implement in a practical application. Horng et al. [7] recognized these limitations and used a different approach to solve geometric ratios and wrinkle analysis. Sobel edge operator and region labeling were first applied to find the positions of facial features (eyes, noses, mouths). Then, two geometric features were evaluated as the ratios of the distance between eyes, nose, and mouth, and three different wrinkle features were extracted in five regions: foreheads, two eye corners and two cheeks, by using the Sobel edge magnitudes. Finally, a back-propagation neural network was used to perform classification in four age groups: babies, young adults, middle-adults, and old adults. The accuracy rates achieved were: 90.5% for 116 training images and 81.6% for 114 testing images. Hayashi et al. [8] studied the pattern of face wrinkles for estimating age and gender, from a database of controlled face images taken from 300 of subjects ranging from 15 to 64 years of age. Their approach involved the extraction of skin regions for the purpose of quantifying and identifying the types of wrinkles, short or long. The wrinkles were enhanced by histogram equalization on the raw skin regions and the wrinkle classification was performed by the Digital Template Hough Transform. Finally, a look-up table based on the number and type of wrinkles was used to classify the subject's age. Results showed a 27% accuracy on age estimation and a 83% accuracy on gender classification. However, there is no information on the size and source of the database and the study was performed exclusively on the Japanese people. The work demonstrated that wrinkle distribution is not a good indicator of age.

The research in age estimation (age determination or age classification) has increased significantly since 2002 [9,10, 11,12,13]. The recent methods can be divided into two categories: *local approaches* and *holistic approaches*.

A. *Local Approaches*

Yan et al. [9] introduced a method using coordinate patches and Gaussian Mixture Models (GMMs) to estimate facial ages. According to their analysis, contrary to holistic

978-1-4244-5020-6/09/$25.00 ©2009 IEEE

face interpretation, the local patches have sufficiency to overcome the shortcomings of sensitivity, illumination variations and image occlusions. Moreover, age features are usually encoded by local information, such as: wrinkles on the forehead or at the eye corners. In their method, face image of an individual is encoded as an ensemble of overlapped Spatially Flexible Patches (SFPs), each of which integrates coordinate information together with the local features that are extracted by 2-D Discrete Cosine Transform (DCT). These extracted SFPs are modeled with GMMs to estimate the age of a person in the input facial image, by comparing the sum of likelihoods from total SFPs of the hypothetic age. Their experimental Mean Absolute Errors (MAEs) results reach 8.53 for females and 7.82 for males on the private YAMAHA database, with 8,000 images of 1,600 subjects with ages ranging from 0 to 93. However, an analysis performed in [10] revealed three problems. First, using order-less SFPs will destroy spatial information of features in human faces. Second, the distance between two faces is randomly calculated by discarding the physical positions of parts of human faces, such as: foreheads, mouths, noses, etc. And, finally, they just used a single probability density function to model the distribution of SPFs from an image and supposed these SPFs were i.i.d (independent and identically distributed). To overcome these shortcomings, Zhuang et al. [10] used Hidden Markov Models (HMMs) super vectors to present the face image patches. Each image is modeled by one patch-based HMM supervector and each patch is in a state of HMM. The facial similarity between two images is also defined as the pair-wise Euclidean distance, approximated by the Kullback-Leibler divergence between corresponding patch distributions. The experimental results outperformed the previous approach, with MAEs of 6.333 years for females and 5.397 years for males on the YAMAHA database. Yan et al. [11] also improved their previous approach by using Gaussian Mixture Models to model the distribution of an ensemble consisting of order-less coordinate patches for each image. Then, a Maximum a Posterior (MAP) adaptation is used to further express the specific distribution. The authors also defined a facial similarity measurement, called inter-modality similarity synchronization, learned by a weak learning process. Finally, they employ a kernel regression method for optimizing human age. Their experimental MAE results reached 4.95 on FG-NET database [12], and 4.94 for males and 4.38 for females on the YAMAHA database.

B. Holistic Approaches

Lanitis et al. [13] used the Active Appearance Model (AAM) [14], a statistical face model, to study age estimation problems. In their approach, after AAM parameters were extracted from face images landmarked with 68 points, an "*aging function*" was built using Genetic Algorithms to optimize the aging function. They introduced four different kinds of functions, including: the Global Aging Function, where changes in aging appearance are assumed to be similar for all individuals; the Appearance Specific Aging Function (ASA), where each individual has an aging function in the training set; the Weighted Appearance Aging Function (WAA), where the individual aging functions are combined together with specific weighted parameters; and the Weighted Person Specific Aging Function (WSA), where the life-style profile of each individual is integrated with facial appearance into a weighted aging function. Meanwhile Geng et al. [15] introduced an AGing pattErn Subspace (AGES) to estimate the ages of individuals. Their basic idea is to model the aging pattern, which is defined as a sequence of a particular individual's face images sorted in time order, by constructing a representative subspace. The proper aging pattern for a previously unseen face image is determined by the projection in the subspace that can reconstruct the face image with a minimum reconstruction error, while the position of the face image in that aging pattern will then indicate its age. In their experiments, the MAE of AGES is 6.77 for all ages.

C. Human Face Age-Progression

As suggested by [1, 4, 16, 17], physical age-progression can be divided into the following categories: growth and development (childhood to pre-adulthood), which primarily affects the physical structure of the craniofacial complex in the form of lengthening and widening of bony-tissue; and adult aging that spans from the end of pre-adulthood to senescence, where the primary impacts occur in the soft-tissue, in the formation of wrinkles, lines, ptosis, and soft tissue morphology. Since the two aging periods have fundamentally different aging mechanisms, there are two specific "aging functions" that need to be constructed: the *growth and development function* and the *adult aging function*.

III. Support Vector Machines (SVMs)

We first give a brief overview of the basics of SVMs for binary classification [21]. Then, we explain how this technique can be expanded to deal with the regression problem.

A. Binary Classification

Given N training points (x_1, y_1), (x_2, y_2), ..., (x_N, y_N) with $x_i \in \mathbb{R}^n$ and $y_i \in \{-1,1\}$, $i = 1,...,N$ and suppose these points are linearly separable; we have to find a set of N_s support vectors s_i ($N_s \leq N$), coefficient weights α_i, constant b and the linear decision surface, as in Eq. 1 below, such that the distance to the support vectors is maximized:

$$w.x + b = 0 \qquad (1)$$

where,

$$w = \sum_{i=1}^{N_s} \alpha_i y_i s_i \qquad (2)$$

SVMs can be expanded to become nonlinear decision surfaces by first using a mapping Φ to map these points to some other Euclidian space H, that is linearly separable with a given regularization parameter $C > 0$, $\Phi : \mathbb{R}^n \mapsto H$, and by defining a kernel function K, where $K(x_i,x_j) = \Phi(x_i).\Phi(x_j)$.

Then, the nonlinear decision surface is defined as:

$$\sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b = 0 \quad (3)$$

where α_i and b are the optimal solution of a Quadratic Programming (QP) as follows:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N_s} \xi_i$$
$$y_i(w.x_i + b) \geq 1 - \xi_i \text{ with } \xi_i \geq 0$$

B. Support Vector Regression (SVR)

The goal of the SVR problem is to build a hyperplane "close" to as many of the training points as possible. Given N training points (x_1, y_1), (x_2, y_2), ..., (x_N, y_N) with $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$, $i = 1,...,N$, we have to construct the hyper plane and values of w and b. The hyper plane w is selected with a small norm, while simultaneously minimizing the sum of the distances from these training points to the hyperplane, measured by using Vapnik's ε-insensitive loss function:

$$|y_i - (w.x_i + b)|_\varepsilon = \begin{cases} 0 & \text{if } |y_i - (w.x_i + b)| \leq \varepsilon \\ |y_i - (w.x_i + b)| - \varepsilon & \text{otherwise} \end{cases} \quad (4)$$

The value of ε is selected by the user, and the tradeoff between finding a hyper-plane with a good regression performance is controlled via the given regularization parameter C. The QP problem associated with SVR is described as follows:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N_s} (\xi_i + \xi_i^*)$$
$$y_i - (w.x_i + b) \leq \varepsilon + \xi_i \quad \text{with } \xi_i \geq 0$$
$$-y_i + (w.x_i + b) \leq \varepsilon + \xi_i^* \text{ with } \xi_i^* \geq 0$$

IV. THE NOVEL AGE ESTIMATION METHOD

The feature vectors x are extracted from face images I and work as the inputs of the age classification module. There are two main steps in the classification module. First, a classifier $f(x)$ is built to distinguish between youths (aging from 0 to 20) and adults (aging from 21 to 69). Then, the two aging functions, which include the growth and development function $f_1(x)$ and adult aging function $f_2(x)$, are constructed (Fig. 1).

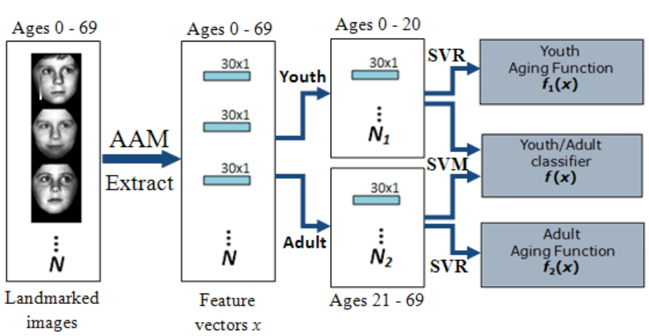

Fig. 1. Training steps

A. Feature Extraction

Each face image is annotated with sixty-eight landmark points as shown in Fig. 2 in a specific anthropometric order as outlined in table 1. Then, AAM is used to extract the feature vector x_i of size of 30×1 for the image I_i. AAMs have been used successfully in several areas of facial interpretation, such as: face recognition [18], face interpretation [19], face aging progression [13], etc. It is statistical models of appearance, generated by combining an Active Shape Model (ASM)[20] which represents the facial structure and a quasi-localized texture model which represents the pattern of pixel intensities (skin texture) across a face image patch. Given a facial image and corresponding landmark points, a statistical model of shape variation can be generated from the points by using Principle Components Analysis (PCA). Given a mean shape of the face, each training example is warped into a 'facial shape-free' patch. A statistical model of the texture variation is then built into this patch. The AAM result will encode correlations between the parameters of the shape model and those of the texture model across the training set. The final training images can be represented by using:

$$X_i = \overline{X} + Px_i \quad (5)$$

where, X_i represents the shape or texture of a training image

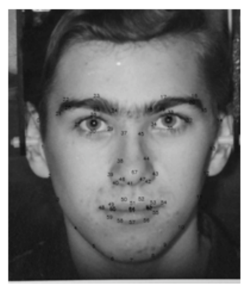

Fig. 2. Landmarked face

Facial parts	Landmark ID	Number of points
Face outside	0 – 14	15
Right eyebrow	15 – 20	6
Left eyebrow	21 – 26	6
Left eye	27 – 30	4
Left iris	31	1
Right eye	32 – 35	4
Right iris	36	1
Nose	37 – 45	9
Nose center	67	1
Nostrils	46, 47	2
Lip outline	48 – 59	12
Top lip	48, 60, 61, 62, 54	5
Bottom lip	54, 63, 64, 65, 48	5

Table 1. The facial landmarks

I_i, $\overline{X}$ is the mean example, P is the eigenvector matrix and x_i is a vector of weights that is called the 30×1 feature vector.

B. Youth/Adult Classification

A binary classifier $f(x)$ (Eq. 6) is first built by SVMs as in section 3 to classify whether an input face image is of a child or an adult. In training steps, the inputs x_i are the 30×1 feature vectors and the corresponding labels $y_i \in \{-1,1\}$ (1 for children, -1 for adults). To configure the SVM parameters, we use Gaussian kernel K (Eq. 7).

$$f(x) = \sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b \quad (6)$$

$$K(x_i, x_j) = e^{-\frac{1}{2\sigma^2}\|x_i - x_j\|^2} \quad (7)$$

Given an output from youth/adult classifier, an appropriate aging function is used to determine the age of the face (as Fig. 3). As discussed above, two aging functions are constructed, i.e. *growth and development function* $f_1(x)$ and *adult aging function* $f_2(x)$.

C. Growth and Development Function

The aging progress can interpreted by an aging function as described in Eq. 8. In [13], the author used a polynomial equation to represent the function $f_1(x)$, and the weight and offset parameters are optimized by a Genetic Algorithm. We propose to use Support Vector Regression (SVR) [22] to construct the function $f_1(x)$ from the training youth images with known ages (range from 0 to 20) and 30×1 feature vectors x_i. SVR model with linear ε-insensitive cost was used to train the aging function $f_1(x)$. To configure the parameters, we also used Gaussian Kernel, λ is 10^{-3} and ε is 0.05.

$$age = f_1(x) \qquad (8)$$

D. *Adult Aging Function*

To construct the adult aging function $f_2(x)$, we follow the steps of building the growth and development function. However, here we select adult images with ages ranging from 21 to 69 to train the function.

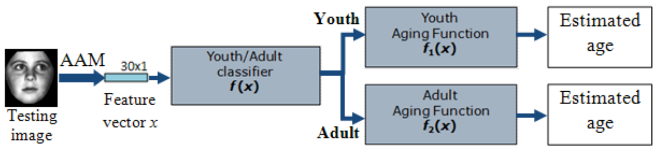

Fig. 3. Age estimation steps

V. EXPERIMENTAL RESULTS

This work used two databases, FG-NET [12], which contains scanned face images from newborns to senescence and the smaller BGC database [16], which is a public database of family members with a strong focus on childhood images. For the performance comparison, we used the Mean Absolute Error (MAE) measurement. MAE is defined as the average of the absolute error between the recognized labels and the ground truth labels:

$$MAE = \frac{\sum_{i=1}^{N_t} \left| \hat{x}_i - x_i \right|}{N_t} \qquad (9)$$

where $\hat{x}_i$ is the recognized age for the i^{th} testing sample, x_i is the corresponding ground truth, and N_t is the total number of the testing samples. Additionally, another popular measure, called Cumulative Score (Eq. 10), was also used to evaluate the estimation results:

$$CumScore(\theta) = \frac{N_{e \le \theta}}{N_t} \times 100\% \qquad (10)$$

where $N_{e \le \theta}$ is the number of recognition results having the absolute errors that are equal to or smaller than θ.

A. *Training Steps*

We used 802 face images from FG-NET with ages ranging from 0 to 69 years for training of the youth/adult classifier. Childhood images were selected to train the growth and development function. The others were used in the training of the adult aging function.

Age Ranges	On training DB (FG-NET)		On testing DB (FG-NET)		On testing DB (BGC DB)	
	# of photos	MAEs	# of photos	MAEs	# of photos	MAEs
0 – 20	588	1.09	140	1.93	193	3.70
21 – 69	214	1.37	60	5.80	0	N/A
0 - 69	802	1.78	200	**4.37**	193	**5.87**(*)

Table 2. MAEs for proposed method

B. *Testing Steps*

The recognition results were evaluated on two different databases. For testing we used a hold out set of 200 face images (not included in the training set) from the FG-NET database. A second test was conducted on 193 faces from the BGC database; however, none of the BGC faces were used for training of the system. The results of age estimation on the FG-NET hold out set and the BGC images are shown in Table 2. The BGC faces were all young faces, i.e. less than 21 years of age; however, the AAM classifier made some erroneously classified some of these images as adults instead of "youths", and hence, used the adult aging function, $f_2(x)$, thereby increasing the MAE for the overall. As shown in the Fig. 4 and Table 3, our proposed method has the lowest MAEs and highest cumulative scores as compared to other published results on the FG-NET data set.

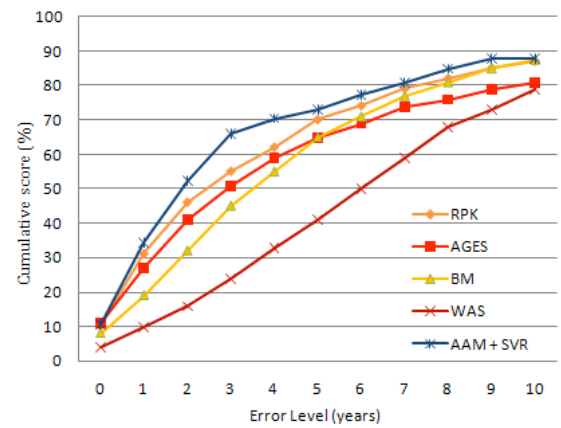

Fig. 4. Comparison of cumulative scores of age estimation methods at error levels from 0 to 10 years in FG-NET database

Methods	MAEs
Our proposed method	4.37
RPK [11]	4.95
AGES [15]	6.77
BM [23]	5.33
WAS [13]	8.06

Table 3. Comparison of estimation results on FG-NET database with ages from 0 to 69 years

VI. CONCLUSIONS

We have presented a novel age estimation method based on characteristics of human craniofacial development. Our method takes advantages of AAMs for extracting facial features based on both shape changes that occur throughout aging as well as the texture changes that are seen only during adult aging. Additionally, SVMs and SVR are used to train the binary classifier and building two regression functions for age determination. The experimental results demonstrate the efficacy of this approach by yielding the best MAE results to date.

However, Table 2 BGC results demonstrate the need to develop a better youth/adult classifier. As it illustrates that the classifier can significantly affect the overall result of the system. Using the correct aging function, $f_1(x)$, for the BGC yielded an overall MAE of 3.70 years.

ACKNOWLEDGMENT

"This study was made possible by use of material from the Burlington Growth Centre, Faculty of Dentistry, University of Toronto, which was supported by funds provided by Grant (1) (No. 605-7-299) National Health Grant (Canada), (data collection); (2) Province of Ontario Grant PR 33 (duplicating) and (3) the Varsity Fund (for housing and collection)".

This work was, in part, made possible by funds provided by the Department of Defense under the ODNI Center for Academic Excellence pilot program CASIS. In addition, Project 322, Ministry of Education and Training, Vietnam Government funded this study.

REFERENCES

[1] M. Albert, K. Ricanek and E. Patterson, A review of the literature on the aging adult skull and face: Implications for forensic science research and applications, Journal of Forensic Science International, April 2007.

[2] L. G. Farkas, Anthropometry of the Head and Face in Medicine, New York, Elsevier North Holland, 1981.

[3] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King, "Aspects of Age Variation in Facial Morphology Affecting Biometrics," Proceedings of the IEEE Conference on Biometrics: Theory, Applications, and Systems, Washington, D.C., September, 2007.

[4] M. G. Rhodes, *Age Estimation of Faces: A Review*, Applied Cognitive Psychology 23, 1-12, 2008.

[5] Y.H. Kwon and N. da Vitoria Lobo, *Locating facial features for age classification*, In proceedings of SPIE – the International Society for Optical Engineering, vol. 2055, pp. 62-72, 1993.

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

[7] W. B. Horng, C. P. Lee and C. W. Chen, *Classification of Age Groups Based on Facial Features*, Tamkang Journal of Science and Engineering, vol. 4, no. 3, pp. 183-192, 2001.

[8] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, *Age and Gender Estimation from Facial Image Processing*, the 41st SICE Annual Conference, vol. 1, pp. 13 -18, Aug. 2002.

[9] S. Yan, M. Liu, T. S. Huang, *Extracting Age Information from Local Spatially Flexible Patches*, ICASSP, 2008.

[10] X. Zhuang, X. Zhou, M. Hasegawa-Johnson, and T. S. Huang, *Face Age Estimation Using Patch-based Hidden Markov Model Supervectors*, ICPR, 2008.

[11] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, T. S. Huang, *Regression from Patch-Kernel*, ICPR 2008.

[12] A. Lanitis, *Comparative Evaluation of Automatic Age-Progression Methodologies*, EURASIP Journal on Advances in Signal Processing, volume 8, issue 2, Jan. 2008.

[13] A. Lanitis, C. J. Taylor, T. F. Cootes, *Modeling the process of ageing in face images*, ICCV, 1999.

[14] T. F. Cootes, G. J. Edwards, C. J. Taylor, *Active Appearance Models*, ECCV, 1998.

[15] X. Geng, Z. Zhou, and K. Smith-Miles, *Automatic age estimation based on facial aging patterns*, PAMI, vol.29, no.12, pp. 2234-2240, 2007.

[16] K. Luu, K. Ricanek, T.D. Bui, and C.Y. Suen, *The Burlington Familial Face Database: A Longitudinal Study of Family-based Growth and Development on Face Recognition*, in the Proceedings of the First IEEE International Workshop on Robust Biometrics: Understanding the Science and Technology, Hawaii, Oct. 2008.

[17] K. Ricanek Jr. and T. Tesafaye, *MORPH: A longitudinal Image Database of Normal Adult Age-Progression*, in Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR '06), pp. 341-345, Southampton, UK, April 2006.

[18] G. J. Edwards, T. F. Cootes, and C. J. Taylor, *Face Recognition Using Active Appearance Models*, ECCV, 1998.

[19] G. J. Edwards, C. J. Taylor and T. F. Cootes, *Interpreting Face Images using Active Appearance Models*, 3*rd* IEEE International Conference on Automatic Face and Gesture Recognition, 1998.

[20] T. F. Cootes, C. J. Taylor, D. H. Copper, and J. Graham, *Active Shape Models – Their training and Application*, Computer Vision Graphics and Image Understanding, vol. 61, no 1, pp 38–59, 1995.

[21] V. Vapnik, *Statistical learning theory*, John Wiley and Sons, New York, 1998.

[22] A. Smola and B. Scholkopf, *A tutorial on support vector regression*, Technical Report TR-1998-030, Neuro Colt Royal Holloway College, 1998.

[23] S. Yan, H. Wang, T. Huang, Q. Yang, and X. T. Xiaoou, *Ranking with Uncertain Labels*, ICME, pp. 96-99, 2007.