Age classification using Radon transform and entropy based scaling SVM

Huiyu Zhou[1]
h.zhou@ecit.qub.ac.uk

Paul Miller[1]
p.miller@ecit.qub.ac.uk

Jianguo Zhang[2]
jgzhang@computing.dundee.ac.uk

[1] The Institute of Electronics, Communications and Information Technology Queen's University Belfast

[2] School of Computing University of Dundee United Kingdom

Abstract

This paper mainly addresses the problem of age classification. Image features can be extracted using a difference of Gaussian filter followed by Radon transform. The relevance and importance of these features are determined in a scaling support vector machine classifier, where zero weights are assigned to irrelevant variables. To enhance the quality of feature selection, we introduce entropy estimation to the scaling classifier. Experimental results demonstrate that the proposed algorithm leads to better recognition accuracy than the state of the art.

1 Introduction

Aging has a significant impact on the appearance of a face [8]. Consequently, face identification or verification can be significantly affected by variations in appearance due to age [6]. To maintain identification/verification performance in the presence of age variation, some researchers have attempted to address this issue classifying the subjects age [10][31]. Unfortunately, age classification itself is very challenging due to the anatomical changes in the cranio-facial region, the bony portion of the head and the overlying soft-tissue caused by the aging progress [33]. For example, a more angular face shape emerges with the growth of the nose and nasal bridge; or, the eyes of a young child appear larger against the face [28]. In general, anatomical changes may occur in the regions called "internal features" such as eyes, mouth, nose and nasal bridge; and "external features" such as the chin, face outline and hair [2].

Age classification has increasingly become more important due to its role in face identification/verification [30]. Up to now, research studies on this topic have been mainly driven towards handling uncontrollable and personalised databases [9]. For example, Kwon and Vitoria Lobo [19] utilised cranio-facial development theory and facial skin wrinkle analysis to categorise a face into several age groups. Aging pattern subspace analysis was proposed in [11],[32] to deal with highly incomplete data due to the problem of data acquisition. On the other hand, facial features were extracted from an appearance based shape-texture model

[20]. These features were then represented by a set of fitted model parameters, which were produced using a regression approach.

Despite their success in age classification, these established approaches can be further extended in two ways: firstly, most these approaches are based on single features or a combination of multiple features, e.g. Active Appearance Model (AAM) and wrinkles. Less effort has been made in order to retain important cues from individual descriptors. Secondly, fundamental studies on feature selection have not been substantially achieved, resulting in a lack of knowledge about the importance of individual features in the age classification. In this paper, we mainly address the latter issue. In other words, we attempt to investigate a mechanism that enables us to extract features and adequately select attributes for age classification rather than undertaking a comprehensive evaluation on the facial features reported in the literature.

In our system, we propose to extract perceptual features by applying difference of Gaussians (DoG) filtering to a face image. These features are then processed using a Radon transform in order to diminish the effects of in-plane facial rotation, which often occurs in realistic face images. Afterwards, to achieve correct age classification using appropriate attributes, we propose an improved adaptive scaling approach for feature selection in a support vector machine (SVM) classifier. The proposed approach is efficient, unsupervised, and does not require face segmentation. To our knowledge, this is the first attempt to adaptively allocate the attributes of the selected features for the purpose of age classification. Fig. 1 illustrates the proposed feature extraction algorithm step by step. Discrepancies of Radon coefficients appear against different ages (bottom row). Note that the proposed feature selection stage has not been shown in this graph but will be introduced in Section 3.

This paper is structured as follows. Section 2 introduces related work on the topic of feature extraction and selection. In Section 3 we propose a novel framework for age classification. Experiments related to the implementations are presented in Section 4. Section 5 contains the conclusions and points out possible areas of future work.

2   Related work

DoG has been frequently used as a filtering technique to enhance image edges by convolving the original gray-scale image with Gaussian kernels of different standard deviations. Applications of DoG have been commonly found in object tracking and pattern recognition. For example, to localise keypoints for feature tracking, Lowe [21] suggested smoothing the input image with Gaussian filters at various scales and then calculating the DoG that presents a fast approximation of the Laplacian operator. To generate a local feature based representation, DoG has been used with PCA-SIFT [18] to handle the classification problem [24]. It has also been pointed out in [34] that DoG is vital for the analysis of bandpass behaviours.

The Radon transform, which is conceptually similar to the Hough transform, can be used to generate linear features using a low pass filter. It is based on the parameterization of straight lines of the image, and evaluates the integrals of the image along these [22]. Evidence shows that the lower frequency components contribute to the global description, while the higher frequency components contribute to the finer details required in the identification process [41]. In practice, the Radon transform is normally used to derive directional facial features in a manner similar to Gabor filters. This capability complements the performance of DoG in the presence of in-plane facial rotation. There exists a remarkable number of applications of the Radon transform in classification and object recognition. For example,

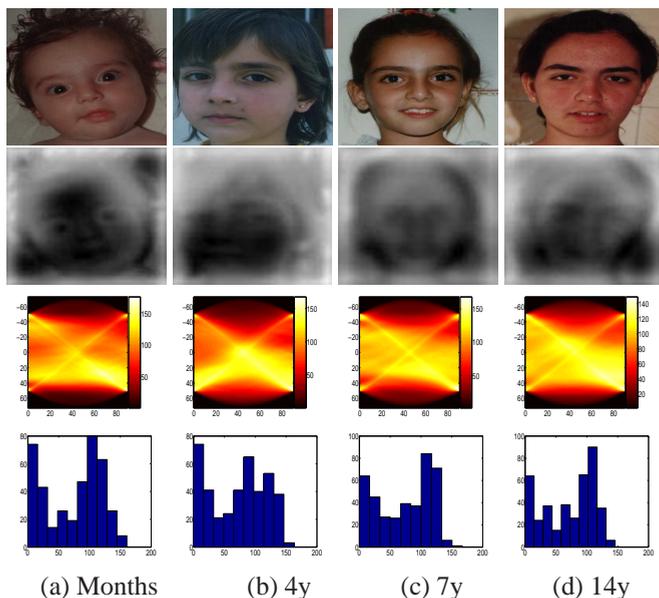|			
(a) Months	(b) 4y	(c) 7y	(d) 14y

Figure 1: Illustration of feature extraction against different ages of the same person: row 1 - original, row 2 - DoG, row 3 - Radon maps, and row 4 - histograms of Randon maps, where x-axis is intensity and y-axis stands for bins and hereafter. Better view in colour.

Jafari-Khouzani and Soltanian-Zadeh [16] presented a new approach for texture classification, utilising the Radon transform to detect the principal texture direction, followed by a wavelet transform to extract texture features. Wu and Ye [42] combined the Radon transform with neural networks for classification of infra-red finger images. Rahtu and Heikkila [26] reported affine invariant transforms using the Radon transform for object recognition. Boulgouris and Chi [3] used the Radon transform of binary silhouettes for gait recognition.

DoG filtering and the Radon transform used in our study only provide primitive descriptions (e.g. wrinkles, shapes and textures) of the image that lack semantic meaning. Without proper training and/or correspondence, it is difficult to associate these features with age determination in different genders etc. Therefore, our aim is to discover what features proactively contribute to age classification. Classical feature selection approaches can be categorised into supervised and unsupervised. For example, Jain and Zongker [17] compared several established feature selection algorithms, e.g. sequential floating, genetic and branch-and-bound algorithms. Viola and Jones [37] described a real-time face detection system designed using the AdaBoost learning algorithm. Zhao and Liu [44] proposed a spectral graph theory based framework to determine the relevance of a feature by its consistency with the structure of the graph. Peng *et al*. [25] presented a new algorithm, namely mRMR, to select good features according to the maximal statistical dependency criterion based on mutual information. Grandvalet and Canu [12] introduced an effective algorithm for automatic relevance determination of input variables in kernelised SVMs.

3   Proposed algorithm

As mentioned in the first section, the proposed age classification algorithm begins by DoG filtering. This is followed by applying the Radon transform to the DOG filtered images, leading to so-called "Radonmaps" containing Radon coefficients of the face image. The Radon coefficients are then dynamically and adaptively optimised for age classification within the framework of an automatic scaling SVM. Detailed descriptions are as follows.

3.1   Adaptive DoG filtering

A face image usually contains background clutter, noise and illumination variation. We wish to limit the influence of these factors in an efficient and effective manner. However, existing edge detectors may not be capable of fully meeting this requirement because of loss of texture information. DoG is an operator that allows regions of rapid intensity change to be detected, whilst preserving, to some extent, the face texture. DoG has been used in face classification as a preprocessing tool to overcome the problems of lighting variations and multiple scales [15]. To achieve an adequate filtering result, whilst optimally balancing low and high frequency information, we define the DoG standard deviations to be: $\sigma_1 = \frac{\sigma_0}{8}$, and $\sigma_2 = \frac{\sigma_0}{16}$, where σ_0 is the SD of the entire face image. This determination is different from that reported in [35] in which a fixed SD suits only specific datasets. It has been reported that DoG filtering may cause some image blur in shadow regions. To compensate for this loss, Tan and Triggs [35] suggested Gamma correction before DoG filtering. We have observed that in the presence of bright light and strong reflections, Gamma correction is necessary. In our application, however, the image contrast is normal so that Gamma correction does not significantly boost system performance. Therefore, we omit the Gamma correction stage.

To further enhance the performance of DoG filtering under differing illumination conditions, we adopt a contrast equalisation scheme similar to that introduced in [35]. Usually, intensity artefacts include specular reflections and small dark regions, e.g., nostrils. Our solution is as follows:

$$\begin{cases} I_1(x,y) \leftarrow \frac{I_0(x,y)}{(median(|I_0(x',y')|^\alpha))^{\frac{1}{\alpha}}}, \\ I(x,y) \leftarrow \frac{I_1(x,y)}{(median(min(\beta,|I_1(x',y')|^\alpha))^{\frac{1}{\alpha}}}. \end{cases} \tag{1}$$

where α is a scaling factor (e.g. 0.1) that limits the effect of relatively large intensity values, β is a threshold (e.g. 10) used to truncate the first phase of the normalisation, and $I_0(x,y)$ is the DoG output. In our approach, we take a median value instead of the mean because the former has better performance in reducing the impact of outliers in the distribution's tails. Finally, a hyperbolic tangent function $I(x,y) \leftarrow \beta \tanh\left(\frac{I(x,y)}{\beta}\right)$ is applied, which limits I to the range of $(-\beta, \beta)$. Fig. 2 shows some examples of contrast equalisation where, regardless of reflections and angular variations, the contrast equalisation scheme results in a successful intensity balance.

3.2   Rotation-invariant features using Radon transform

Object recognition and classification require invariant features against various transformations such as rotation, scale, illumination and deformation. Many successful studies can be found in the research literature. For example, Mel [23] proposed a perceptually driven object
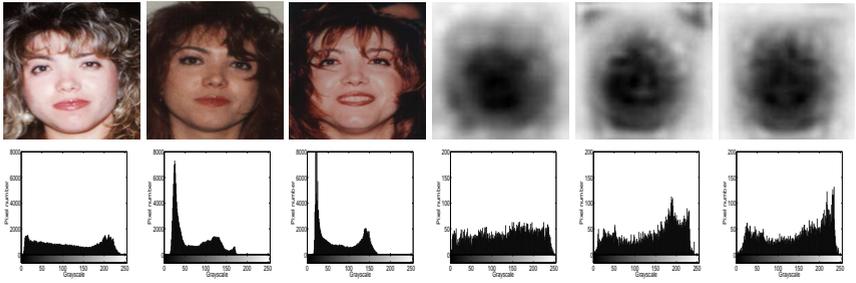
Figure 2: Illustration of contrast equalisation. Row 1: images 1-3 - original, and images 4-6 - outcomes of contrast equalisation over images 1-3, respectively. Row 2: corresponding histograms of row 1. Better view in colour.

recognition system using multiple low-level attributes. Lowe [21] introduced the SIFT features that appear to be to some extent rotational and scale invariant. A local descriptor based on sets of oriented Gaussian derivatives was used in [43] due to their selectivity to specific orientations and frequencies. A rotation- and scale-invariant Gabor representation was proposed in [13], where each representation refers to the summations of the conventional Gabor filter impulse responses. Compared to these established approaches, the Radon transform has unusual advantages in representing lines and curves. This advantage is extremely valuable as facial curves such as face outline and wrinkles help improve the performance of age classification [14],[7]. In the meantime, facial sketch curves have been successfully applied to face recognition [38]. The Radon transform of $I(x,y)$ is defined as [39]:

$$R(\mathbf{t}, \theta)(I(x,y)) = \int \int I(x,y)\delta(\mathbf{t} - x\cos\theta - y\sin\theta)dxdy, \qquad (2)$$

where $\delta(\mathbf{t})$ is the Dirac function, $\mathbf{t}$ is the perpendicular distance of a straight line from the origin, and θ is the angle between the distance vector and the x-axis, e.g. $\theta \in [0, \pi/2)$. Now, we take a look at the properties of an image translation or rotation in the Radon transform domain. Let $I_\phi(x,y)$ be the rotated version of $I(x,y)$ with rotation angle ϕ. Thus, we have $R_\phi(\mathbf{t}, \theta) = R(\mathbf{t}, \theta + \phi)$, where $R_\phi(\mathbf{t}, \theta)$ is the Radon transform of $I_\phi(x,y)$. The correlation function of $R(\mathbf{t}, \theta)$ and $R_\phi(\mathbf{t}, \theta)$ is given by

$$\mathcal{C}(\tau, \mathbf{t}) = \int_o^{2\pi} R_\phi(\mathbf{t}, \theta)R(\mathbf{t}, \theta + \tau)d\theta. \qquad (3)$$

Let $\gamma = \theta + \phi$, then $d\theta = d\gamma$. Eq. 3 can be re-written as:

$$\mathcal{C}(\tau, \mathbf{t}) = \int_o^{2\pi} R(\mathbf{t}, \gamma)R(\mathbf{t}, \gamma + \tau - \phi)d\gamma. \qquad (4)$$

A parameter namely "Radon projection correlation distance" is defined as $d(I(x,y), I_\phi(x,y))$. For each value of $\mathbf{t}$, the function $\mathcal{C}(\tau, \mathbf{t})$ has the maximum value when $\tau = \phi$. We also have $d(I(x,y), I_\phi(x,y)) = 0$, and the rotation angle ϕ is equivalent to μ (μ indicates the mean value of the overall maximum values of $\mathcal{C}(\tau_i, \mathbf{t})$). This indicates that the Radon projection correlation distance $d(I(x,y), I_\phi(x,y))$ can be used as a criterion for rotation invariant age classification [40]. For the reason of efficiency, in our system this distance can be simply
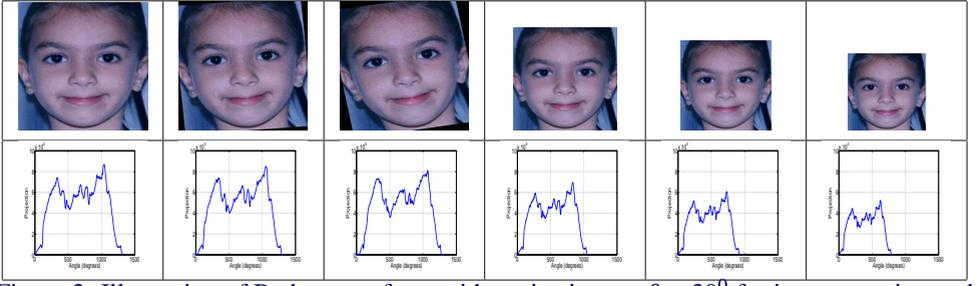
Figure 3: Illustration of Radon transform with projections at $\theta = 30^0$ for image rotation and scaling: images 1-3 of row 1 - original and the rotated images of 5^0 and 10^0, images 4-6 of row 1 - 0.8, 0.7 and 0.6 scaling of the original image, and row 2 - corresponding projections of row 1.

approximated to be the difference of the numerical sums of the Radon coefficients over the whole image. Fig. 3 demonstrates the rotation (not scaling) invariance of the Radon transform over different angular variations in this application.

3.3   Entropy based scaling SVM classification

SVM classifiers are frequently used in pattern recognition and classification. In contrast to logistic regression, which relies on a deterministic model to predict the occurrence of a binary event by fitting data to a logistic curve, SVM separates two classes by generating a hyperplane that optimally separates classes in a high-dimensional space. Current research in the field of feature selection primarily addresses the search for important and optimal attributes that can improve the performance of SVM classification. Successful examples can be found in, e.g. [27]. These methods mainly address the feature ranking problem.

Scaling is a preprocessing step that has shown promise in classification applications, including SVM classifiers [12]. A common approach is to compute the generalisation error of regularly spaced hyper-parameters, and then select the best solution across the overall trials. For an input $\mathbf{x}$, the SVM decision function $sign(f_\sigma(\mathbf{x}))$, where f_σ is defined as:

$$f_\sigma(\mathbf{x}) = \mathbf{w}^T \phi_\sigma(\mathbf{x}) + b = \sum_i \mathbf{y}_i \alpha_i K_\sigma(\mathbf{x}_i, \mathbf{x}) + b, \qquad (5)$$

where α_i are Lagrange multipliers, $\mathbf{y}_i$ are labels, K_σ is a kernel, and the weight and bias $(\mathbf{w}, b)$ can be derived if this optimisation problem can be solved: $\min_{\mathbf{w}, b, \eta} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \eta_i$, subject to $\mathbf{y}_i(\mathbf{w}^T \phi_\sigma(\mathbf{x}_i) + b) \geq 1 - \eta_i$ and $\eta_i \geq 0$ $(i = 1,...,n)$, where n is the number of inputs, $\phi_\sigma(\mathbf{x})$ is defined as $\phi(\sum x)$, η_i are slack variables, C and the kernel bandwidth σ are tunable hyper-parameters which need to be defined by the user.

Let the weight $\mathbf{w} = \sum_{j=1}^n \alpha_j \mathbf{y}_j \phi_\sigma(\mathbf{x}_j)$, assuming the Lagrange multipliers are less affected by the updates of σ. We then have a different representation of the optimisation problem in the scaling SVM:

$$\min_{\mathbf{w}, b, \eta} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j K_\sigma(\mathbf{x}_i, \mathbf{x}_j) + C \sum_{i=1}^n \eta_i, \qquad (6)$$

subject to these constraints: $\mathbf{y}_i \left(\sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j K_\sigma(\mathbf{x}_i, \mathbf{x}_j) + b \right) \geq 1 - \eta_i$, $\eta_i \geq 0$, $\sigma_0^q = \frac{\sum_{k=1}^d \sigma_k^q}{d}$ (d and q are two numbers) and $\sigma_k \geq 0$ $(i = 1,...,n$ and $k = 1,...,d)$ [12]. Given σ and α, η

can be obtained solving the equation: $\min_{b,\eta} C \sum_{i=1}^n \eta_i$ with the same constraint. Finally, σ can be updated using a conjugated reduced gradient gradient technique, provided that μ, $C \sum_{i=1}^n \eta_i$ and its derivative with respect to σ have been computed. How Lagrange multipliers α_i are updated is omitted here due to the space limit. The reader can refer to [12] for further details.

The kernel K_σ has a significant impact on the determination of the weight $\mathbf{w}$. In our approach, to seek an adequate weight, we propose to use entropy to measure the average information of the overall inputs in the SVM classification domain. In particular, we use *Rényi* entropy due to the diversity and randomness of the measures, which has been defined as the rank of probability density function [4]:

$$H_{R\varphi} = \frac{1}{1-\varphi} \log \int p^\varphi(x) dx. \tag{7}$$

Using a quadratic *Rényi* entropy we have: $H_{R2} = -\log \int p^2(x) dx$. Due to $\int p^2(x) dx \approx \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_\sigma(\mathbf{x}_i, \mathbf{x}_j)$, we then obtain the quadratic *Rényi* entropy as follows:

$$H_{R2} \approx -\log \left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_\sigma(\mathbf{x}_i, \mathbf{x}_j) \right). \tag{8}$$

Our approach leads to a significant change in the weight that positively affects the final outcome of the classification due to the augmented attributes. The proposed *Rényi* entropy based SVM classification algorithm is shown in Algorithm 1.

Algorithm 1 The proposed entropy based scaling SVM classifier.

repeat
    1. Obtain the weight $\mathbf{w}$ and kernel K_σ using the classical adaptive scaling SVM [12].
    2. Calculate *Rényi* entropy $H_{R2}(\mathbf{v})$ by Eq. 8, where $\mathbf{v}$ indicates the overall inputs.
    3. Select a portion $\tilde{X}_i$ of $\mathbf{v}$, whose *Rényi* entropy is larger than (mean+2*standard deviation) of $H_{R2}(\mathbf{v})$.
    4. $\mathbf{w} = \mathbf{w} \cup \{\tilde{X}_i\}$.
    5. Calculate b, α and errors.
until
    6. Stop if criteria are met.

---

4 Experimental work

We evaluate the proposed age classification algorithm, with the entropy based scaling SVM classifier, by comparing its performance against other state of the art techniques. For our application, video surveillance, we are interested in two age classes; youths/adolescents and adults. This is because most anti-social behaviour is due to the former group. Hence, we categorise the face images into two groups; less than twenty years old, and older. Other categorisations will follow the same strategy as described in this paper. The overall approach involved in this comparison consists of 5-fold cross validation SVM classifiers based on (1) principal component analysis (PCA) [36], (2) Local Binary Patterns (LBP) [31], (3) histogram of oriented gradient (HOG) [5], (4) DoG and Radon transform without feature
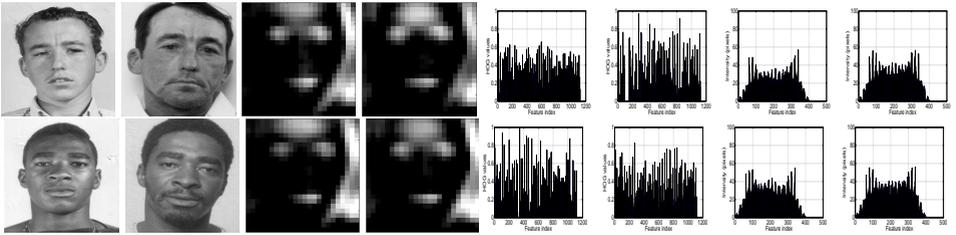
Figure 4: Illustration of feature extraction using different algorithms in the MORPH database. Cols 1-2: original (different ages for two persons), cols 3-4: PCA (reconstructions using top 50 eigenvectors), cols 5-6: HOG, and cols 7-8: proposed DoG and Radom transform.

selection, (5) DoG, Radon transform with the classical feature selection [12], (6) HOG features with the proposed feature selection algorithm and (7) DoG, Radon transform with the proposed feature selection algorithm. For convenience, these algorithms can be abbreviated as (1) PCA, (2) HOG, (3) LBP, (4) DRT, (5) DRTC, (6) HOGSS and (7) DRTP.

Two publicly accessible datasets, FG-NET Aging Database [1] and MORPH database [29], are used in the evaluation. The FG-NET Aging Database contains 1002 face images from 82 subjects, where the image number of youths is 706. In the MORPH database, there are 1724 face images from 515 subjects, where the image number of youths is 342. Each subject has around three images taken at different ages. The ages in both databases are distributed highly unevenly over wide age ranges: 0-69 for FG-NET and 15-68 for MORPH. The entire evaluation includes two stages: firstly, the recognition rates of the seven algorithms on age classification will be compared. Secondly, error rates of classification against the training set sizes are compared using the DRTC and DRTP algorithms. In this study, we repeat the same experiments fifty times and the results are then averaged.

4.1   MORPH database

Fig. 4 shows examples of the feature extraction by PCA, HOG, and the proposed DoG with Radon transform. Observation reveals that the forehead, mouth, and eyes have high contrast in the PCA reconstructions. The HOG features of various faces form significantly different histograms. The feature values from the proposed DoG+Radon transform are less than sixty pixels (mean values have been deducted from the overall estimates and the images have been resized to 20-by-20). A close look at the histograms in the last two columns shows that the shape boundaries are distinctive from one another.

Using these extracted features, we now evaluate the performance of various algorithms against the MORPH database. We here show the average recognition rates and test error rates against the number of training sets by the proposed DRTP algorithm. Fig. 5 demonstrates that (1) the eyes, mouth and forehead (outlined by ellipses) seem to contain important features determined in the feature selection stage. (2) The proposed DRTP scheme has the highest recognition rate (app. 86%) while PCA only has app. 77%. (2) Using the proposed DRTP algorithm, we obtain less errors as the training sizes increase. Compared with the classical DRTC approach, the proposed DRTP algorithm has faster convergence speed. We also observe that around 15% of the features do not contribute to the age classification in the
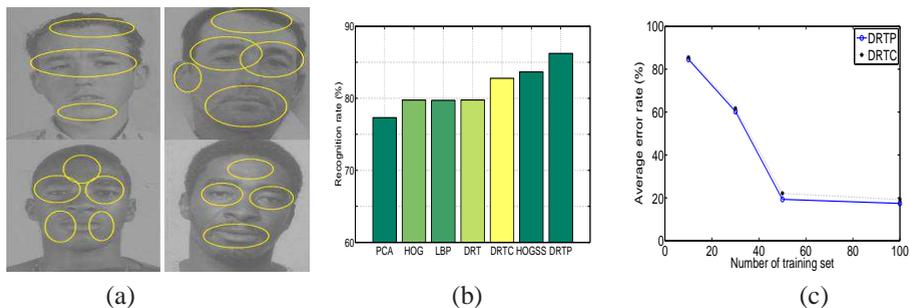
Figure 5: Performance of various algorithms in age classification (MORPH): (a) Examples of important features, outlined by the ellipses for the images shown in Fig. 4, (b) recognition rates, (c) DRTP vs. DRTC: average error rates against the number of training sets.

last iteration round.

4.2 FG-NET database

Example images and associated feature extraction for FG-NET database have been shown in Fig. 1. We focus here on the performance comparison of age classification. Fig. 6 illustrates that (1) as before, the eyes, mouth and forehead, tend to play a more active role than other features. (2) The proposed DRTP approach has the best recognition rate due to the feature selection. Without this feature selection (i.e. the DRT algorithm), the outcome is similar to LBP and HOG. (3) The DRTP scheme can reach an error rate of around 25%. Once again, DRTP arguably converges faster than DRTC. In the final iteration round, about 30% of the features are considered to be irrelevant.

5 Conclusions and future work

In this paper we have introduced a method to perform automatic feature extraction and selection in nonlinear SVMs. Our approach uses DoG filtering followed by the Radon transform for feature extraction. To properly determine the importance of the extracted features for age classification, we use a *Rényi* entropy based SVM classification algorithm that adaptively changes the weights in the SVM decision function. Experimental results show that this new approach has better performance than other state of the art techniques in terms of recognition accuracy and convergence speed. In the future, we intend to combine a psychophysical model with our approach. The motivation for this is driven by the need that the relevance between features must be properly used in order to achieve better classification accuracy. A thorough study on the psychophysics will allow us to capture the most important features for the purpose of age classification. This research work is ongoing.

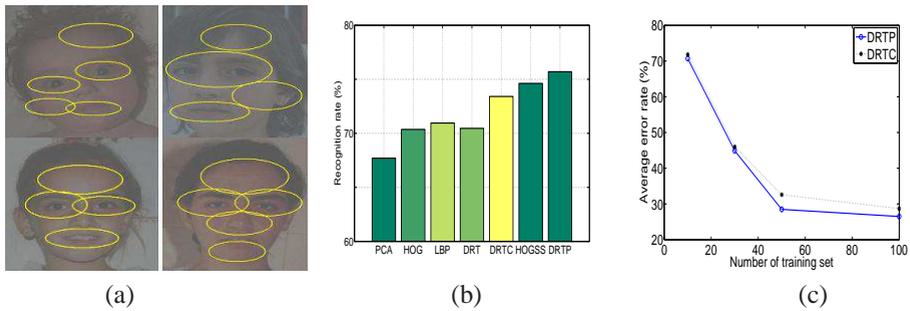(a)   (b)   (c)

Figure 6: Performance of various algorithms in age classification (FG-NET): (a) Examples of important features from the images shown in Fig. 1, (b) recognition rates, (c) DRTP vs. DRTC: average error rates against the number of training sets.

References

[1] The FG-NET aging database. http://sting.cycollege.ac.cy/~alanitis/, 2002.

[2] T.J. Andrews, J. Davies-Thompson, A. Kingstone, and A.W. Young. Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *The J. of Neurosci.*, 30(9):3544–3552, 2010.

[3] N.V. Boulgouris and Z.X. Chi. Gait recognition using radon transform and linear discriminant analysis. *IEEE Trans. on Image Proc.*, 16(3):731–740, 2007.

[4] G.C. Cawley and N.L.C. Talbot. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Netw.*, 17(10):1467–1475, 2004.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of CVPR*, pages 886–893, 2005.

[6] H. Dehon and S. Bredart. An 'other-race' effect in age estimation from faces. *Perception*, 30(9):1107–1113, 2001.

[7] M.M. Dehshibi and A. Bastanfard. A new algorithm for age recognition from facial images. *Signal Process.*, 90:2431–2444, August 2010.

[8] D. Enlow. *The handbook of facial growth (2nd ed.)*. London: W.B. Saunders, 1982.

[9] Y. Fu and T.S. Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimedia*, 10(4):578–584, 2008.

[10] Y. Fu, G. Guo, and T.S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:1955–1976, November 2010.

[11] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Trans. PAMI*, 29(12):2234–2240, 2007.

[12] Y. Grandvalet and S. Canu. Adaptive scaling for feature selection in svms. In *NIPS*, pages 553–560, 2002.

[13] J. Han and K.-K. Ma. Rotation-invariant and scale-invariant gabor features for texture image retrieval. *Image Vision Comput.*, 25(9):1474–1481, 2007.

[14] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu. Age and gender estimation from facial image processing. In *Proc. of he 41st SICE Annual Conference*, pages 13–18, 2002.

[15] A. J. Howell and H. Buxton. Towards unconstrained face recognition from image sequences. In *Proc. of International Conference on Automatic Face and Gesture Recognition*, pages 224–229, 1996.

[16] K. Jafari-Khouzani and H. Soltanian-Zadeh. Radon transform orientation estimation for rotation invariant texture analysis. *IEEE Trans. PAMI*, 27(6):1004–1008, 2005.

[17] A. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. PAMI*, 19(2):153–158, 1997.

[18] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proc. CVPR*, 2004.

[19] Y.H. Kwon and Niels da Vitoria Lobo. Age classification from facial images. *Comput. Vis. Image Underst.*, 74(1):1–21, 1999.

[20] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern. Part B*, 34(1):621–628, 2004.

[21] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int'l J. of Comput. Vis.*, 60(2):91–110, 2004.

[22] F. Matus and J. Flusser. Image representation via a finite radon transform. *IEEE Trans. PAMI*, 15:996–1006, 1993.

[23] B.W. Mel. Seemore: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *NeurComp*, 9(4):777–804, May 1997.

[24] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *Proc. ICCV*, 2005.

[25] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. PAMI*, 27: 1226–1238, 2005.

[26] E. Rahtu and J. heikkila. A new affine invariant image transform based on ridgelets. In *Proc. British Machine Vision Conference*, 2006.

[27] A. Rakotomamonjy. Variable selection using svm based criteria. *J. Mach. Learn. Res.*, 3:1357–1370, 2003.

[28] M.G. Rhodes. Age estimation of faces: a review. *Applied Cognit. Psychol.*, 27(2): 1–12, 2009.

[29] K. Ricanek Jr. and T. Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. In *Proc. of FGR*, pages 341–345, 2006.

[30] J.J. Ruiz-Hernandez, J.L. Crowley, and A. Lux. "how old are you?" : Age estimation with tensors of binary gaussian receptive maps. In *British Machine Vision Conference*, pages 1–11, 2010.

[31] C. Shan. Learning local features for age estimation on real-life faces. In *Proc. of ACM International Workshop on Multimodal Pervasive Video Analysis*, pages 23–28, 2010.

[32] J. Suo, X. Chen, S. Shan, and W. Gao. Learning long term face aging patterns from partially dense aging databases. In *Proc. ICCV*, pages 622–629, 2009.

[33] M.A. Taister, S.D. Holliday, and H.I.M. Borrman. Comments on facial aging in law enforcement investigation. *Foren. Sci. Comm.*, 2(2):11, 2000.

[34] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Int'l Workshop on Analy. and Modeling of Faces and Gestures*, pages 168–182, 2007.

[35] X. Tan and B. Triggs. Fusing Gabor and LBP feature sets for kernel-based face recognition. In *Int'l Workshop on Analysis and Modeling of Faces and Gestures*, pages 235–249, 2007.

[36] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cog. Neurosci.*, 3(1):71–86, 1991.

[37] P. Viola and M.J. Jones. Robust real-time face detection. *Int'l J. Comput. Vision*, 57 (2):137–154, 2004.

[38] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Trans. PAMI*, 31(11):1955–1967, 2009.

[39] X. Wang, B. Xiao, J.-F. Ma, and X.-L. Bi. Scaling and rotation invariant analysis approach to object recognition based on radon and fourier-mellin transforms. *Pattern Recogn.*, 40(12):3503–3508, 2007.

[40] X. Wang, F.-X. Guo, B. Xiao, and J.-F. Ma. Rotation invariant analysis and orientation estimation method for texture classification based on radon transform and correlation analysis. *J. Vis. Comun. Image Represent.*, 21(1):29–32, 2010.

[41] G.I. Webb and X. Yu, editors. *Advances in Artificial Intelligence, 17th Australian Joint Conference on Artificial Intelligence*, volume 3339 of *Lecture Notes in Computer Science*. Springer, 2004.

[42] J.-D. Wu and S.-H. Ye. Driver identification using finger-vein patterns with radon transform and neural network. *Exp. Syst. with Applic.: An Intern. J.*, 36(3):5793–5799, 2009.

[43] J.J. Yokono and T. Poggio. Oriented filters for object recognition: an empirical study. In *FGR*, page 755, 2004.

[44] Z. Zhao and H. Liu. Spectral feature selection for supervised and unsupervised learning. In *Proc. of ICML*, pages 1151–1157, 2007.