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

Age classification has increasingly become very important due to its role in face identification/verification [2]. Up to now, research studies on this topic have been mainly driven towards handling uncontrollable and personalised databases [1]. Despite their success in age classification, the established approaches can be further extended in two ways: firstly, most these approaches are based on single features or a combination of multiple features. 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 attempt to investigate a mechanism that enables us to extract features and adequately select attributes for age classification. 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 facial rotation. 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. Fig. 1 illustrates the proposed feature extraction algorithm step by step. Discrepancies of Radon coefficients appear against different ages (bottom row).

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. DoG is an operator that allows regions of rapid intensity change to be detected, whilst preserving, to some extent, the face texture. To achieve an adequate filtering result, whilst optimally balancing low and high frequency information, we adopt a contrast equalisation scheme similar to that introduced in [3].

Age classification require invariant features against various transformations such as rotation, scale, illumination and deformation. Compared to the state of the art features such as Gabor and SIFT, the Radon transform has unusual advantages in representing lines and curves. The Radon transform of $I(x,y)$ (output of DoG) is defined as

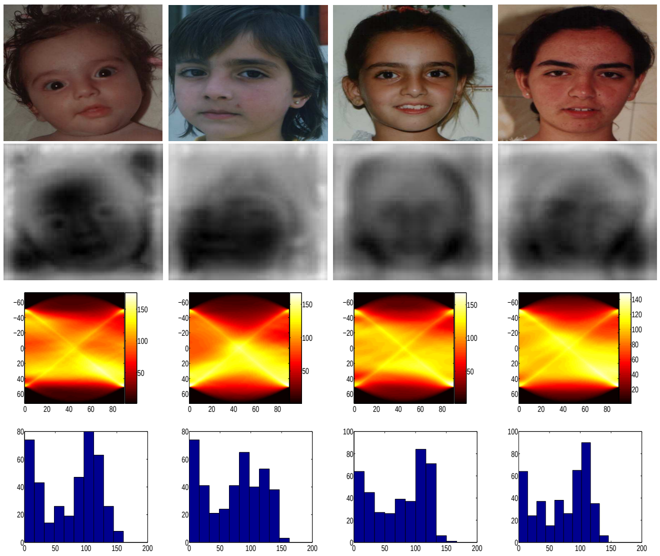
$$R(\mathbf{t},\theta)(I(x,y)) = \int\int I(x,y)\delta(\mathbf{t} - x\cos\theta - y\sin\theta)dxdy, \tag{1}$$

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. $\theta \in [0, \pi/2)$. Let $I_\phi(x,y)$ be the rotated version of $I(x,y)$ with rotation angle ϕ. A parameter namely "Radon projection correlation distance" is defined as $d(I(x,y), I_\phi(x,y))$. It has been proved 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 [4]. For the reason of efficiency, in our system this distance can be simply approximated to be the difference of the numerical sums of the Radon coefficients over the whole image.

In the stage of classification, we search for important and optimal attributes that can improve the performance of SVM classification. Scaling is a preprocessing step that has shown promise in classification applications, e.g. SVM classifiers. 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:

$$f_\sigma(\mathbf{x}) = \mathbf{w}^T\phi_\sigma(\mathbf{x}) + b = \sum_i y_i\alpha_i K_\sigma(\mathbf{x}_i, \mathbf{x}) + b, \tag{2}$$

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.



(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.

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:

$$H_{R2} \approx -\log\Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n K_\sigma(\mathbf{x}_i, \mathbf{x}_j)\Big). \tag{3}$$

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, i.e. (1) PCA, (2) HOG, (3) LBP, (4) DRT, (5) DRTC and (6) HOGSS. For our application, video surveillance, we are interested in two age classes; youths/adolescents and adults. Two publicly accessible datasets, FG-NET Aging Database and MORPH database, are used in the evaluation. The results favor the proposed approach in terms of classification accuracy and convergence speed.

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

[2] 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.

[3] 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.

[4] 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.