CrossMark

Gait-based human age estimation using age group-dependent manifold learning and regression

**Xiang Li[1,2] · Yasushi Makihara[2] · Chi Xu[1,2] ·
Yasushi Yagi[2] · Mingwu Ren[1]**

Abstract Human age estimation from gait is expected to be an important technology for a variety of applications such as automatic customer counting for marketing research or automatic age-based access control restriction for a specific area because the gait can be observable at a distance from a camera (e.g., CCTV). Although the aging process of gait significantly differs among age groups (e.g., children, adults, and the elderly), previous studies on gait-based human age estimation employ a single age group-independent estimation model that suffers from large estimation errors when the age variation increases. We therefore propose an age group-dependent gait-based human age estimation method for better accuracy. Specifically, in the training phase, we first compose age groups that are well-separated from each other by clustering gait features along with their age labels. We then learn a classifier that classifies the gait features for multiple age groups using a directed acyclic graph support vector machine. Next, we learn an age regression model for each age group using support vector regression with a Gaussian kernel in conjunction with a manifold learning technique, i.e., orthogonal locality preserving projection, to better

✉ Yasushi Makihara
makihara@am.sanken.osaka-u.ac.jp

Xiang Li
lixiangmzlx@gmail.com

Chi Xu
xuchisherry@gmail.com

Yasushi Yagi
yagi@am.sanken.osaka-u.ac.jp

Mingwu Ren
renmingwu@mail.njust.edu.cn

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

[2] The Institute of Scientific and Industrial Research, Osaka University, Osaka 567-0046, Japan

⚛ Springer

characterize the gait feature. In the test phase, given a gait feature, it is first classified into an age group and then its age is estimated with the age regression model of the classified age group. Experimental results on a gait database that has the world's largest population of participants ranging from 2 to 90 years old demonstrate the state-of-the-art performance of the proposed method.

1 Introduction

Gait biometrics have recently received more attention because of increasing demands for visual surveillance. Compared with other biometrics (e.g., irises, faces, and finger veins), gait has many advantages. For instance, it can be perceivable even at a long distance from a camera (i.e., from a low-resolution image). Moreover, gait is an unconscious behavior (i.e., people generally do not conceal their gait intentionally) and it does not require the subject's cooperation. Most gait-based studies focus on gait-based authentication or identification [17, 30, 34, 36, 38], which can be directly used for many applications such as surveillance, forensics, and criminal investigations [2, 16, 22].

Moreover, studies on recognizing other human attributes (e.g., age, gender, and ethnicity), also play an important role. Among them, human age estimation is an interesting and active research area. Existing methods of human age estimation mainly rely on facial images [4, 9–12, 40]. Face-based human age estimation may, however, not work in surveillance scenarios because captured facial images may be of low-resolution, have limited texture information, or even consist of faces covered by a mask. In contrast, gait-based human age estimation has its own unique advantages, particularly in surveillance scenarios, because gait can still be well perceived under such conditions. Gait-based human age estimation therefore provides many potential applications, such as an automatic customer counting scenario in which age group is of interest for product marketing research or automatic age-based access control to a specific area.

Besides, gait, or, more strictly speaking, a composite of gait and shape, is feasible for human age estimation because it contains age discriminative clues. For example, Davis [7] showed the gait difference between an adult and a child in terms of leg length, stride width, and stride frequencies. Ince et al. [14] showed that the head-to-body ratio of a child is different from an adult. Similar findings were also reported in [25] by analyzing the most widely used appearance-based gait representation (known as gait energy images (GEIs) [13] or averaged silhouettes [18]), which contains both gait and shape information. To provide more evidence for this, GEIs averaged over individuals for each age range are shown in Fig. 1. As we can see, there are obvious changes in head-to-body ratio as children grow. In addition, as people get older, a middle-aged spread and stoop appear. Hence, such changes in appearance-based gait representation provide the possibility of gait-based human age estimation.

Existing gait-based human age estimation methods typically consist of two modules: feature representation and human age estimation algorithms. As for feature representation, GEI is the most commonly used [19, 25] because of its simple yet effective properties. Some other approaches [20, 21] apply age manifold learning techniques on GEI to help find a low-dimensional representation that captures the intrinsic data distribution and geometric structure of GEI. As for human age estimation algorithms, there are mainly two
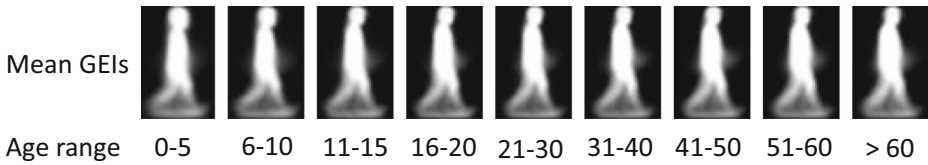
Fig. 1 Mean GEIs of subjects with different age ranges

categories: classification-based approaches [19] and regression-based approaches [21, 25]. The classification-based approaches usually regard each age label as an individual class, then use multi-class classification algorithms to solve the human age estimation problem, while the regression-based approaches directly solve a regression problem from a gait feature to human age, which is more natural because age is essentially represented by a continuous value.

However, these studies on gait-based human age estimation all employed a single age group-independent estimation model, regardless of the fact that the aging process of gait significantly differs among age groups (e.g., children, adults, and the elderly), which can be easily seen in Fig. 1 (e.g., the growth speed of children is much faster than adults). Thus, a single age group-independent estimation model may not handle well the differences among age groups, which suffers from large estimation errors when the age variation becomes large.

Therefore, we propose an age group-dependent framework for gait-based human age estimation to handle this problem, which can be regarded as a fusion of classification and regression-based approaches. The major contributions of this paper are summarized as follows:

1) An age group-dependent framework for gait-based human age estimation.

The proposed age group-dependent framework is a fusion of classification-based and regression-based approaches based on a coarse-to-fine principle. More specifically, we employ a directed acyclic graph support vector machine (DAGSVM) for age group classification and support vector regression (SVR) in conjunction with orthogonal locality preserving projection (OLPP) for age group-dependent age regressions, respectively. We can better handle the differences in age progression among age groups thanks to the proposed age group-dependent framework, which leads to higher age estimation accuracy.

2) State-of-the-art accuracy on the world's largest gait database.

The proposed age group-dependent framework is shown to achieve the best accuracy in terms of mean absolute error (MAE) between an estimated age and a ground truth age, compared with other state-of-the-art approaches, through experiments on the OU-ISIR Gait Database, Large Population Dataset with Age (OULP-Age) [39], which has a larger population (more than 60,000 subjects) than any other dataset and ages that range from 2 to 90 years old.

3) Low computational cost.

Due to its coarse-to-fine framework, the proposed age group-dependent framework has low computational cost, which is more easily applicable in real-world applications.

The rest of this paper is organized as follows. In Section 2, we briefly review related work on human age group classification and human age estimation using gait features. In

Section 3, we introduce the age group-dependent framework for gait-based human age estimation. In Section 4, we show the experimental results of the proposed methods compared with other benchmarks. Finally, Section 5 concludes this paper and suggests future research directions.

2 Related work

2.1 Gait-based human age group classification

Methods of gait-based human age group classification usually employ static and kinematic features. For example, Davis [7] utilized the properties of leg length, stride width, and stride frequency to classify two age groups: children (3–5 years old) and adults (30–52 years old). Begg et al. [1] classified younger people (28.4 years mean age and 6.4 years standard deviation) and the elderly (69.2 years mean age and 5.1 years standard deviation) using minimum foot clearance data. In [5], features such as head-to body ratio, leg length, and stature were used to classified children and adults based on manually labeled datasets. In [29], spatiotemporal longitudinal and transverse projections of the silhouette during a gait cycle were used to represent the arms' swing, the head's pitch, the hunched posture and the stride's length, which showed a considerable discrimination between young and elderly people. Unlike these studies, some methods used appearance-based gait features for human age group classification. For example, Mannami et al. [27] used frequency-domain features [24] to classify three age groups: children (under 15 years old), adults (between 15 and 65 years old), and the elderly (over 65 years old).

However, there are some limitations of the previous human age group classification methods, e.g., insufficient experimental validation due to limited age range and a small number of very coarse age groups.

2.2 Gait-based human age estimation

Gait-based human age estimation is a relatively new research area, and hence a limited number of studies have been done so far. Lu and Tan [19] published the earliest study on this topic. They first converted each age value into a binary sequence through an effective label encoding scheme. Then, multilabel-guided (MLG) subspace learning was applied on the GEIs and their Gabor representations to better characterize and correlate the age information of a person for estimating human age. Last, they performed multilabel k-nearest neighbors classification instead of traditional classification methods that regard each age label as a class and decoded the age information of the label vector.

Subsequently, Makihara et al. [25] proposed a baseline algorithm for gait-based human age estimation using Gaussian process regression (GPR) [33], which has showed the great successes in the face-based human age estimation field [40], in conjunction with a silhouette-based gait feature, i.e., GEI. The experimental results using a whole-generation gait database including 1728 subjects with a wide age range (from 2 to 94 years old) indicated the potential possibility for gait-based human age estimation in real-world applications.

More recently, another work presented by Lu and Tan [21] employed ordinary preserving manifold analysis methods (i.e., ordinary preserving linear discriminant analysis (OPLDA) and ordinary preserving margin Fisher analysis (OPMFA)) to find a low-dimensional discriminative subspace for human age estimation. Specifically, samples with similar age

values were projected to be as close as possible and those with dissimilar age values were projected as far as possible, simultaneously. Subsequently, a quadratic regression model was applied to uncover the relationships among these low-dimensional features and the ground-truth age values.

There is one more work using the fusion of gait and face features to estimate human age [32]. They first individually fused gait features from several gait periods and face features from several angels using averaging function, then concatenated these individually fused features into a single-feature vector. Finally, a similar age estimation method as [19] was utilized to get the estimated age.

However, these studies all employed a single age group-independent estimation model, regardless of the fact that aging process of gait significantly differs among age groups (e.g., children, adults, and the elderly), and hence suffer from large estimation errors when the age variation increases.

3 Proposed method

The flowchart of our proposed method is shown in Fig. 2. It contains three modules: multi-class classification, age group-dependent manifold learning, and age group-dependent regression. More specifically, in the training stage, we first train a set of support vector machines (SVMs) as a multi-class classifier for the age groups. Second, for each classified age group, we carry out age group-dependent manifold learning to map the original GEI into a low-dimensional subspace for better regression at the next stage. Third, we train a non-linear SVR model for each age group using all the samples that have been classified into the specific age group to mitigate the effect of mis-classified test samples. In the testing stage, for each test sample, we successively conduct multi-class classification, age group-dependent manifold mapping, and prediction to estimate age. More details of these three modules are given in the rest of this section.

3.1 Gait feature representation

Among the large number of gait feature representations, GEI [13] is the most popular gait feature representation and has been frequently used in many gait recognition algorithms and gait-based human age estimation algorithms due to its simple yet effective properties. We therefore also choose GEI as our gait representation. A GEI is a gait template obtained by
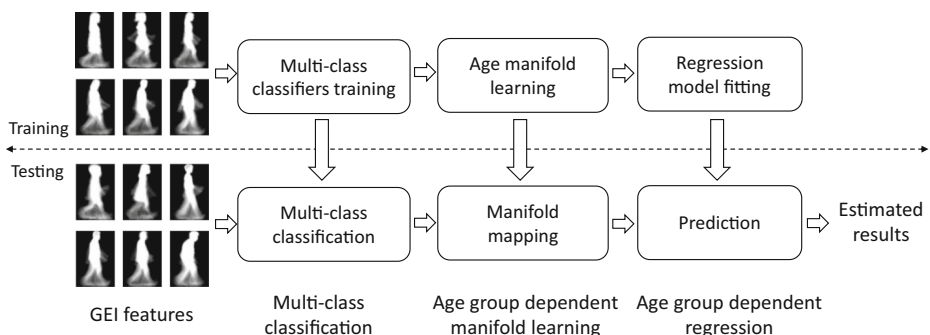
Fig. 2 Flowchart of our proposed method

averaging size-normalized and registered silhouettes over a complete gait period (cycle) T as

$$I(x, y) = \frac{1}{T} \sum_{t=1}^{T} B(x, y, t), \tag{1}$$

where $B(x, y, t)$ is a size-normalized and registered binary silhouette at the position (x, y) at the t-th frame, and $I(x, y)$ is a gait energy at the position (x, y). Clearly, GEI effectively represents dynamic gait features, which are the movement of arms and legs, using pixels with grayscale intensities as well as static gait features using pixels with intensities of 0 or 255, which represent background only or static body parts, respectively.

3.2 Age group definition

In this subsection, we introduce how to define appropriate age groups to obtain a good trade-off between age group classification and age regression in the following stage. In this trade-off, if we prepare higher numbers of age groups, we can train a more precise age group-dependent age regressor while age group classification gets more difficult, and vice versa. By analyzing the human growth process, we initially divide age, in 5-year intervals, up to 20 years old because people tend to grow quickly in their youth. We use 10-year intervals after 20 years old because, as they age, their growth progresses more slowly; for people over 60, we simply group them together due to the relative lack of training data for older people. As a result, we have nine age groups (i.e., 0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, and over 60 years). It is, however, difficult to classify these nine age groups correctly because some age groups may have very similar gait features due to slow growth. Thus, we intend to further merge some age groups to obtain a moderate number of age groups to balance the trade-off between age group classification and age regression.

Specifically, we first calculate the L2 distances of the mean GEIs between every two adjacent age groups because smaller L2 distances indicate more similar age groups. We then choose a threshold and combine those age groups whose L2 distance is under a threshold. Finally, we obtain merged age groups that are more easily classified. The results of age group definition are presented in Section 4.4.

3.3 Age group classification

We choose DAGSVM [31] to solve our multiple age group classification problem. Although the basic SVM [6] was originally designed for binary classification, DAGSVM can integrate multiple binary SVM classifiers to solve the problem of multi-class classification. Additionally, DAGSVM is more suitable for groups with ordered information by using a rooted binary directed acyclic graph. The graph is constructed from root node to leaves by considering the difficulty of classification for each two groups (from easy to difficult).

Suppose there are K age groups, where the 1st age group is children and the Kth age group is the elderly. In the training stage, DAGSVM constructs $K(K-1)/2$ binary classifiers using a linear kernel and every classifier is trained by the samples from two age groups. In the testing stage, the method constructs a rooted binary directed acyclic graph that has $K(K-1)/2$ internal nodes and K leaves, as shown in Fig. 3. In the graph, each node is a binary SVM classifier for the ith and jth age groups; each leaf indicates an age group decision. The root node is the easiest classifier between children and the elderly, while the node in the four layer is the most difficult classifier between two adjacent age groups. Given a test sample x, we first perform classification at the root node classifier. After evaluation,
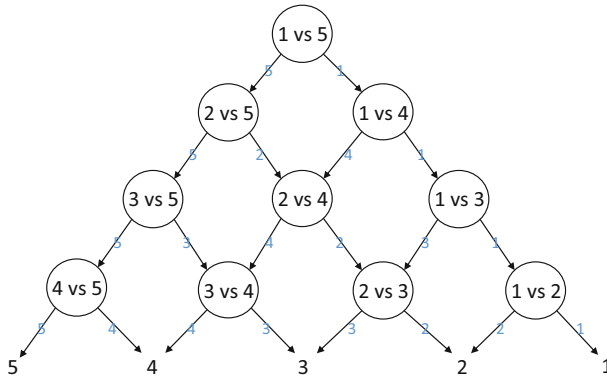
Fig. 3 Rooted binary directed acyclic graph for DAGSVM (using $K = 5$ as an example, the 1st age group is children and the Kth age group is the elderly). Each node is a binary SVM classifier for the ith and jth age groups; each leaf indicates an age group decision

the algorithm moves either left or right depending on the output value. We then traverse the graph and repeat the evaluation. Finally, we end at the leaf node that indicates the predicted age group.

Because the numbers of subjects in different age groups tend to be unbalanced, we adjust the penalty parameters for different binary classifiers. For this purpose, we set a larger penalty parameter for an age group with smaller number of subjects, and vice versa. More specifically, we set the penalty parameters to be inversely proportional to the number of subjects. Let n_k and c_k be the number of subjects and a penalty parameter of the kth age group G_k ($k = 1, 2, \ldots, K$). We then set the penalty parameters as

$$c_k = \frac{\alpha}{n_k}, \tag{2}$$

where α is a penalty coefficient that is common for all the age groups.

3.4 Age manifold learning

The purpose of manifold learning is to find a low dimensional subspace that maintains the intrinsic data distribution and geometric structure with respect to different ages for GEIs. In this paper, we adopt a typical manifold leaning method named OLPP [3] as our age manifold learning method. In this subsection, we briefly describe the procedure to obtain the manifold and we refer the reader to [3] for more details.

Suppose $\{z_1^k, z_2^k, \ldots, z_n^k\} \in \mathbb{R}^D$ and $\{l_1^k, l_2^k, \ldots, l_n^k\} \in \mathbb{R}$ are respectively a set of GEIs and their ground truth age labels in classified age group G_k. Here, each GEI is represented as a D-dimensional column vector. Similar to [3], we first project the GEIs z_i^k onto their principal component analysis (PCA) [37] subspace x_i^k by keeping a pre-defined cumulative variance contribution rate. We denote the transformation matrix of PCA by W_{PCA}. Then, the projection is represented as

$$x_i^k = W_{\text{PCA}}^T (z_i^k - \bar{z}^k), \tag{3}$$

where $\bar{z}^k$ is the mean of all the GEIs in G_k.

Next, we define a similarity matrix S to model the local structure of the gait manifold, where its component at position (i, j) is represented as

$$S_{ij} = \begin{cases} e^{-(\|x_i^k - x_j^k\|^2/t)} & (l_i^k = l_j^k) \\ 0 & (l_i^k \neq l_j^k). \end{cases} \qquad (4)$$

An optimal projection a^* for OLPP is obtained as

$$a^* = \arg\min_a \sum_{i=1}^{n} \sum_{j=1}^{n} (a^T x_i^k - a^T x_j^k)^2 S_{ij}$$

$$= \arg\min_a a^T X^k L (X^k)^T a,$$

$$\text{s.t.} \quad a^T X^k D (X^k)^T a = 1, \qquad (5)$$

where $L = D - S$ is a Laplacian matrix, D is a diagonal matrix defined as $D_{ii} = \sum_j S_{ij}$, and X^k is a matrix of projected vectors in the PCA space defined as $X^k = [x_1^k, x_2^k, \ldots, x_n^k]$.

Let $\{a_1, a_2, \ldots, a_p\}$ be the orthogonal basis vectors. The first projection vector a_1 is obtained by solving eigenvalue problem of (5); then, the remaining projection vectors are obtained in an asymptotic way. The whole procedure is as follows. (1) Compute a_1 as the eigenvector of the following Q associated with the smallest eigenvalue:

$$Q = (X^k D (X^k)^T)^{-1} X^k L (X^k)^T. \qquad (6)$$

(2) Compute a_p as the eigenvector of the following $M^{(p)}$ associated with the smallest eigenvalue:

$$M^{(p)} = Q - (X^k D (X^k)^T)^{-1} A^{(p-1)} [B^{(p-1)}]^{-1} [A^{(p-1)}]^T Q, \qquad (7)$$

where

$$A^{(p-1)} = [a_1, a_2, \ldots, a_{p-1}],$$
$$B^{(p-1)} = [A^{(p-1)}]^T (X^k D (X^k)^T)^{-1} A^{(p-1)}. \qquad (8)$$

Once we have obtained the projection matrix of OLPP $W_{\text{OLPP}} = [a_1, a_2, \ldots, a_p]$, we project the original GEI z_i^k into dimension reduced feature y_i^k as

$$y_i^k = W_{\text{OLPP}}^T W_{\text{PCA}}^T (z_i^k - \bar{z}^k). \qquad (9)$$

3.5 SVR

Given the low dimensional feature y_i^k in the kth classified age group, an SVR [35] function is used to characterize the relationship between the feature and corresponding age label l_i^k. The goal of SVR is to find a function $f(y^k)$ that has a deviation of at most ϵ from the actually obtained targets l^k for all the training data, and, at the same time, is as smooth as possible. In other words, we only care about errors that are larger than ϵ.

Suppose $\mathcal{D} = \{(y_1^k, l_1^k), (y_2^k, l_2^k), \ldots, (y_n^k, l_n^k)\}$ are the data in the kth classified age group. We first describe function $f(y^k)$ in a linear form as

$$f(y^k) = \langle w^k, y^k \rangle + b, \qquad (10)$$

where $\langle ., . \rangle$ denotes the dot product.

Introducing slack variables ξ_i, ξ_i^* to cope with otherwise infeasible constraints, the optimization is formulated as

$$\min \quad \frac{1}{2}\|\boldsymbol{w}^k\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*).$$

$$\text{s.t.} \quad \begin{cases} l_i - \langle \boldsymbol{w}^k, \boldsymbol{y}^k \rangle - b \leq \epsilon + \xi_i \\ \langle \boldsymbol{w}^k, \boldsymbol{y}^k \rangle + b - l_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{11}$$

The constant $C > 0$ determines the trade-off between the flatness of f and the amount to which deviations larger than ϵ are tolerated. The formulation above corresponds to dealing with a so-called ϵ-insensitive loss function $|\xi|_\epsilon$, described by

$$|\xi|_\epsilon = \begin{cases} 0, & \text{if } |\xi| < \epsilon \\ |\xi| - \epsilon, & \text{otherwise.} \end{cases} \tag{12}$$

It turns out that the optimization problem (11) can be solved very easily in its dual formulation. The final solution is given by

$$\boldsymbol{w}^k = \sum_{i=1}^{n}(\alpha_{\text{SV}(i)} - \alpha_{\text{SV}(i)}^*)\boldsymbol{y}_{\text{SV}(i)}^k, \tag{13}$$

and therefore

$$f(\boldsymbol{y}^k) = \sum_{i=1}^{n}(\alpha_{\text{SV}(i)} - \alpha_{\text{SV}(i)}^*)\langle \boldsymbol{y}_{\text{SV}(i)}^k, \boldsymbol{y}^k \rangle + b, \tag{14}$$

where α_i and α_i^* are Lagrange multipliers and $\text{SV}(i)$ denotes the sample index for the ith support vector in $\mathcal{D}$. The value of b is computed by exploiting Karush–Kuhn–Tucker conditions. More details can be found in [35].

In this paper, we use a nonlinear SVR function with a Gaussian kernel (i.e., the radial basis function) because its nonlinear property models the complex aging process well. The radial basis function is defined as

$$k(\boldsymbol{y}, \boldsymbol{y}') = e^{-\frac{\|\boldsymbol{y}-\boldsymbol{y}'\|^2}{2\sigma^2}}, \tag{15}$$

where kernel scale σ is a constant to adjust the width of the Gaussian function. Replacing $\langle \boldsymbol{y}_{\text{SV}(i)}^k, \boldsymbol{y}^k \rangle$ with the kernel function, the solution in (14) becomes

$$f(\boldsymbol{y}^k) = \sum_{i=1}^{n}(\alpha_{\text{SV}(i)} - \alpha_{\text{SV}(i)}^*)k(\boldsymbol{y}_{\text{SV}(i)}^k, \boldsymbol{y}^k) + b. \tag{16}$$

3.6 Training sample selection for regression

When preparing training samples for the age group-dependent regression, we need to consider the effect of mis-classification at the preceding age group classification. Suppose that we train a regression model for a specific age group (e.g., with an interval from 5 to 9 years old; called age group A). If we use training samples whose ground-truth age belongs to the age group, the ages estimated by the regression model are usually bounded by the age group (e.g., between 5 and 9 years old). If a sample from another age group (e.g., with an interval from 10 to 14 years old; called age group B) is mis-classified to age group A, which is

inevitable because perfect age group classification is naturally impossible, we have almost no chance to estimate age from the correct age group B.

To mitigate the effect of the mis-classification, when we train a regression model for a specific age group, we use all the samples that are classified into the age group regardless of their ground truth age groups. In other words, we include a certain proportion of samples from other age groups when training an age group-dependent SVR model. Using this strategy, we have a chance to estimate the correct age of these mis-classified samples, which reduces the effect of mis-classification.

In the best practice of this training sample selection approach, we also consider generalization errors of the age group classification. If we use the same training samples for the regression model as those used for the age group classifier, the number of mis-classified samples are often underestimated, i.e., more samples may be mis-classified at the test phase due to the generalization errors. To avoid this underestimation, we use a sort of validation set as another training set for the regression model that is disjoint from the training set for age group classification.

4 Experiments

4.1 Dataset

We conducted the experiments on the OULP-Age dataset [39]. This dataset was collected by a gait measurement system [26] in an experience-based long-run exhibition at a science museum. Each participant was required to walk using their preferred speed without carrying any items. In addition, they were also asked to declare their informed consent to the use of the collected data for research purposes and provide age and gender information as the ground truth for performance evaluation. After collecting the gait video sequences, we extracted the GEI feature of each participant in three steps: (1) extraction of gait silhouettes using a background subtraction-based graph-cut segmentation [23]; (2) use of the region center to obtain size-normalized (128×88 resolution) and registered silhouettes [24]; and (3) detection of gait period [24] and averaging of the silhouettes within one gait period.

There are a total of 63,846 subjects in the dataset, including 31,093 males and 32,753 females with ages ranging from 2 to 90 years old. The statistics of the subjects' age and gender in 5-year intervals are shown in Fig. 4. The statistics show that the dataset has an extremely large population (approximately 16 times larger than the publicly available large-scale gait database [15]), covers more than four generations (from 2 to 90 years old), and has a good gender balance (the ratio of males to females is close to one). With these advantages, the dataset is an ideal choice for evaluating the performance of the proposed age group-dependent human age estimation method. We then randomly divided the entire dataset into two disjoint subsets: a training set and a testing set. As mentioned in Section 3.5, the training set is further divided into a training set for age group classification and a training set for the regression model (the same as the validation set for the age group classification). The training set for classification includes 15,961 subjects (7829 males and 8132 females) and the training set for regression includes 15,962 subjects (7767 males and 8195 females), whereas the testing set contains 31,923 subjects (15,497 males and 16,426 females).
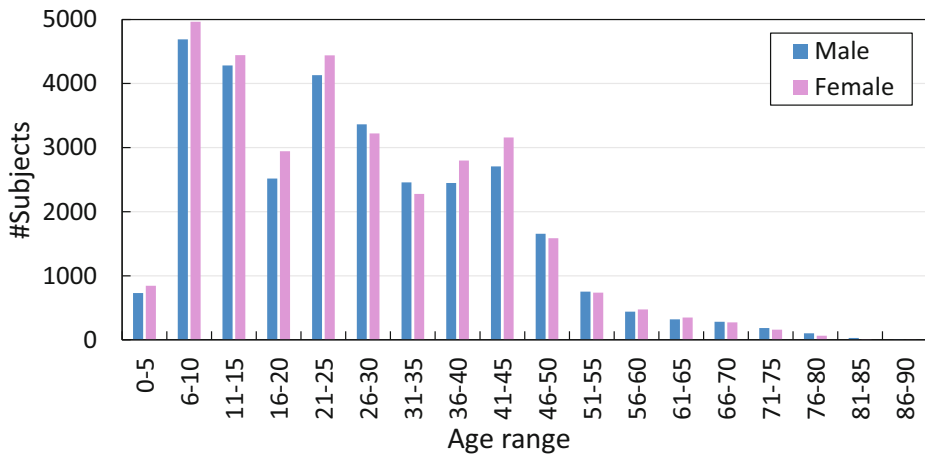
Fig. 4 Statistics of subjects' age and gender in OULP-Age dataset

4.2 Parameter settings

There are several hyper-parameters in our proposed method: (1) penalty coefficient α for DAGSVM in the age group classification module; (2) the cumulative variance contribution rate of PCA and dimension p of the OLPP projection matrix in the age group-dependent manifold learning module; and (3) penalty parameter C, tolerance ϵ, and kernel scale σ for SVR in the age group-dependent regression module.

For DAGSVM, we determined the penalty coefficient α by grid search using the validation set for age group classification. The search range of $c_{k*} = \alpha/n_{k*}$, where k^* is an index for the age group with the largest subjects, was set to $[10^{-7}, 10^{-6}, \ldots, 10^0]$. To evaluate the accuracy of multiple age groups classification, we used macro-average F_1 measure [8], which is the traditional arithmetic mean of the F_1 measure computed for each age group. It gives equal weight to each age group because we expect each age group to have a high correct classification rate regardless of the sample size. Finally, the optimal c_{k*} was set to 10^{-6}, and the penalty parameters c_k of the other age groups were then set using (2).

For manifold learning, we experimentally set parameter t in (4) to 1, as in [3], and the dimension p of OLPP for every age group was set equal to the dimension of GEI after PCA projection that maintains a 99% cumulative variance contribution rate.

For SVR, we experimentally set a small $\epsilon = 0.1$, which means we tolerated an estimated error within 0.1 years of age, and then determined C and σ by a grid search through 4-fold cross-validation on the training set for regression. The search range for both parameters was set to $[2^1, 2^2, \ldots, 2^{10}]$, and the optimal C and σ were both set to 2^5 as a result.

4.3 Evaluation metrics

We measured the performance of human age estimation by two widely used measures: the MAE and cumulative score (CS) [11, 19, 21, 26]. The MAE is defined as the average of the absolute errors between the estimated ages and ground truth ages using the formulation

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\hat{l}_i - l_i|, \tag{17}$$

where l_i is the ground truth age for test sample i, $\hat{l}_i$ is the estimated age, and N is the total number of test subjects. The cumulative score for j-years absolute error tolerance CS(j) is defined as

$$\text{CS}(j) = \frac{N_{e<j}}{N}, \tag{18}$$

where $N_{e<j}$ is the number of test samples whose absolute errors are less than j years.

4.4 Evaluation on age group classification

As mentioned in Section 3.2, we first defined nine age groups (i.e., 0–5, 6–10, 11–15, 16–20, 21–30, 31–40, 41–50, 51–60, and over 60 years) on the training set as a result of an analysis of the human growth process. We then merged adjacent age groups that have similar gait features. The L2 distances between every pair of adjacent age groups are shown in Fig. 5. Subsequently, we experimentally chose a threshold of 8.0×10^4, represented by the red dotted line, and we combined those age groups whose differences are smaller than the threshold. Finally, we obtained five age groups (0–5, 6–10, 11–15, 16–60, and over 60 years) which should realize a good tradeoff between the age group classification and age group-dependent regression compared with the original nine age groups. The numbers of samples for these five age groups are 830, 4,740, 4,373, 21,143, and 837, respectively.

Next, we applied DAGSVM for age group classification using these five age groups. Table 1 shows the confusion matrix of the classification result on the testing set. We can see from this table that the highest correct classification rate (CCR) is 79.84% for age group 0–5 years and the lowest CCR is 63.48% for age group 11–15 years. In general, the average CCR for all age groups is over 70.00%. Moreover, the mis-classified samples are mostly predicted as belonging to the neighboring age groups. For example, in age group 0–5 years, 20.03% samples are predicted as belonging to the neighboring age group 6–10 years. This is
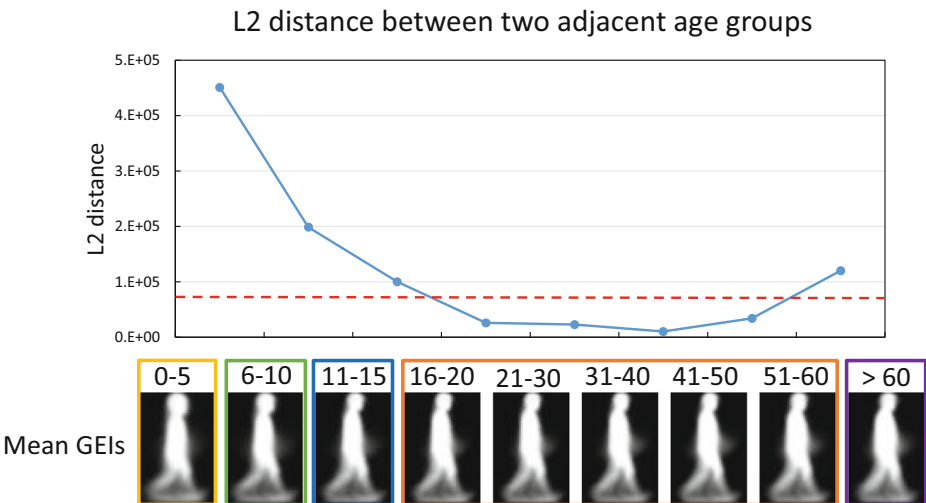
Fig. 5 L2 distance between every pair of adjacent age groups. The red dotted line is the threshold (8.0×10^4) that determines which age groups need to be combined. The age groups marked by rectangles of different colors indicate the age groups combined at that threshold, that is, five age groups (0–5, 6–10, 11–15, 16–60, and over 60 years) after combination

Table 1 Confusion matrix of classification rate [%] on the testing set

Actual \ Predict	0–5	6–10	11–15	16–60	Over 60
0–5	**79.84**	20.03	0.00	0.13	0.00
6–10	8.14	**76.88**	13.06	1.12	0.79
11–15	0.11	14.82	**63.48**	17.15	4.44
16–60	0.03	0.71	13.51	**75.16**	10.60
Over 60	0.00	2.01	6.86	25.34	**65.79**

Bold fonts indicate the correct classification rate (CCR) of each age group

because there is no clear boundary between these age groups. Thus, there is some possibility that some samples will be classified into neighboring age groups.

4.5 Evaluation of human age estimation

Individual component analysis We first conducted experiments to evaluate the individual components of the proposed method. The proposed method is treated as a basic age group-independent regression-based method (regarded as the baseline) plus three individual components: the age group-dependent framework, training sample selection for regression (called SS), and manifold learning method (i.e., OLPP). We then represent the baseline and baseline plus various individual components (1, 2, or 3 components) as "Age group-independent", "Age group-dependent", "Age group-dependent + SS", "Age group-independent + OLPP", "Age group-dependent + OLPP", and "Age group-dependent + SS + OLPP", respectively. In addition, we kept the parameters of every individual component unchanged throughout the experiments. The MAEs of all these methods are shown in Table 2. As a result, the age group-dependent framework, in conjunction with SS, can outperform the baseline, which shows the importance of SS in the age group-dependent framework. Moreover, when all the individual components are included, the proposed method achieves the best performance.

In addition, we compared the proposed method and the baseline, i.e., age group-independent method using two other means of evaluation, in further detail. First, scatter plots for the ground truth ages versus the corresponding estimated ages are shown in Fig. 6. We can clearly see from the figure that the proposed method is more closely distributed around the ground truth, particularly for small ground truth ages from 0 to 10 years. Second, MAE with respect to the ground truth age and mean signed error (MSE) with respect to the estimated age at each interval of 5 years are shown in Fig. 7. As a result, the MAEs of the proposed method in small ground truth age ranges (0–5 and 6–10 years) are much lower than the baseline, which is consistent with the result revealed by Fig. 6. Moreover, the

Table 2 MAE [years old] of individual components

Method	MAE [years old]
Age group-independent	7.66
Age group-dependent	8.52
Age group-dependent + SS	7.13
Age group-independent + OLPP	***6.96***
Age group-dependent + OLPP	8.34
Age group-dependent + SS + OLPP (proposed)	**6.78**

Bold and italic bold fonts indicate the best and the second-best, respectively. This convention is consistent throughout this paper
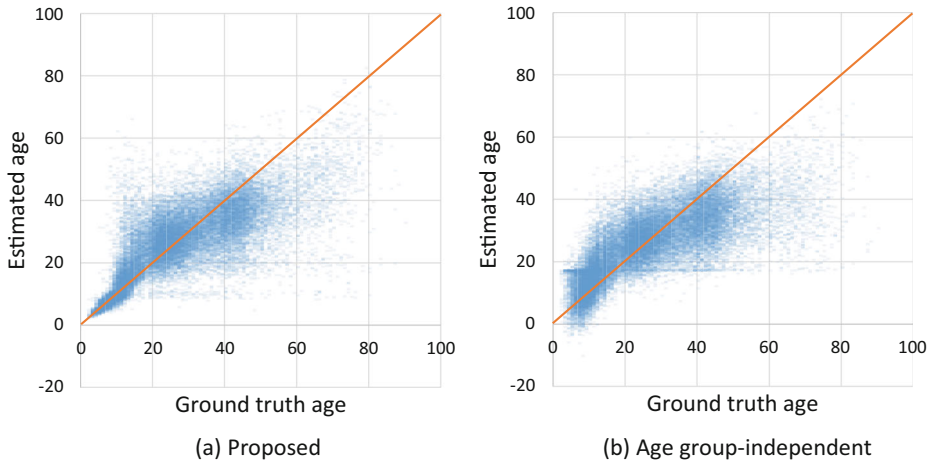
Fig. 6 Scatter plots for ground truth age versus the corresponding estimated ages of the proposed method and the basic age group-independent method. The orange diagonal line indicates where the estimated age is equal to the ground truth age

proposed method also yields lower MAEs in larger ground truth age ranges (over 60 years). As for the MSEs, the estimated age range of the baseline is from -10 to 75 years, which is inconsistent with our knowledge that human ages are non-negative. However, the proposed method can handle the problem well and yields a more reasonable estimated age range from 0 to 85 years.

Comparison with the state-of-the-art methods We compared the proposed method with other state-of-the-art methods: the classification-based method (MLG [19]), regression-based methods (GPR [25] and SVR [35]), and age manifold learning-based methods (OPLDA and OPMFA [21]). Specifically, we implemented three different $K \in \{10, 100, 1,000\}$ for GPR, which is a parameter that determines the number of neighboring training samples to the test sample. We also implemented SVR with both linear and Gaussian kernels.

The MAEs of the proposed method and other comparison methods are shown in Table 3. We can clearly see that the proposed method achieves the state-of-the-art performance, with an MAE that is 0.52 lower than that of the second-best method (GPR with K = 1,000) and with an MAE that is 4.20 lower than the worst method (MLG). Furthermore, we show the cumulative scores of all the methods with absolute error tolerance values from 1 to 15 years in Fig. 8. The proposed method outperforms other state-of-the-art methods across almost the entire range of absolute errors (from 1 to 14 years). More specifically, the proposed method can reach the highest accuracy of 18.41% for an absolute error tolerance of 1 year, which is approximately twice that of other methods excluding MLG. Although MLG achieves the second-best accuracy of 16.71% for an absolute error tolerance of 1 year, the gap between MLG and the proposed method becomes more obvious as the absolute error tolerance increases.

4.6 Sensitivity analysis of hyper-parameters

In this subsection, we analyze the sensitivity of cumulative variance contribution rate r for PCA in the age group manifold learning module and tolerance ϵ for SVR in the
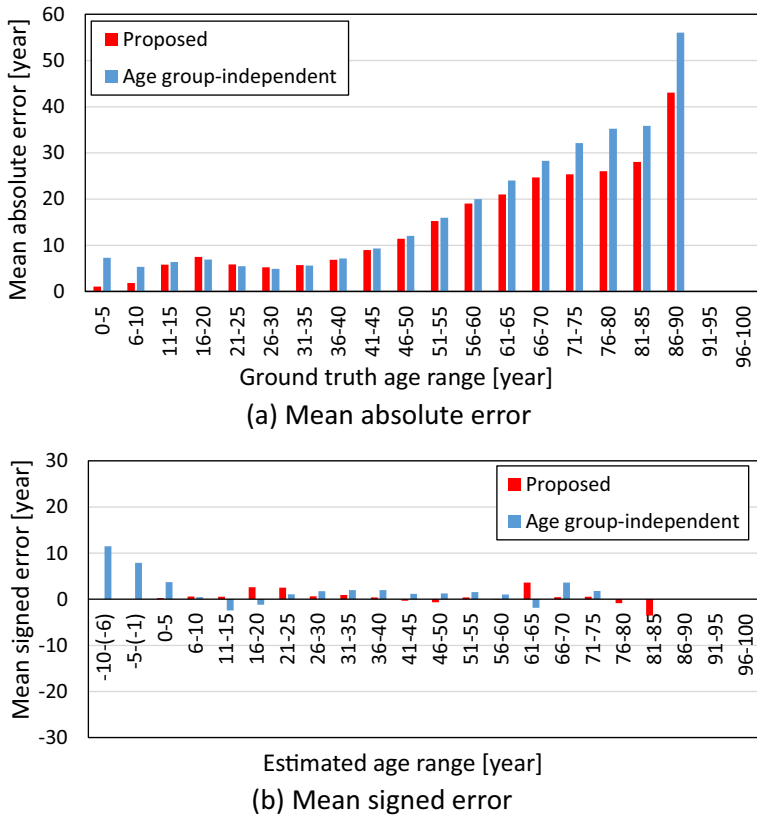
Fig. 7 MAE and MSE of the proposed method and the basic age group-independent method in intervals of 5 years. The upper figure shows the MAEs with respect to the ground truth age, while the lower one shows the MSEs with respect to the estimated age

age group-dependent regression module on the human age estimation performance in OULP-Age. More specifically, we set the default values for all the hyper-parameters based on the criteria mentioned in Section 4.2, then analyzed the sensitivity by changing either r or ϵ.

	Method	MAE [years old]
Table 3 Comparison of MAEs [years old] of the proposed method and state-of-the-art methods	MLG	10.98
	GPR (K = 10)	8.83
	GPR (K = 100)	7.94
	GPR (K = 1,000)	*7.30*
	SVR (linear)	8.73
	SVR (Gaussian)	7.66
	OPLDA	8.45
Bold and bold-italic fonts indicate the best and second-best results	OPMFA	9.08
	Proposed	**6.78**
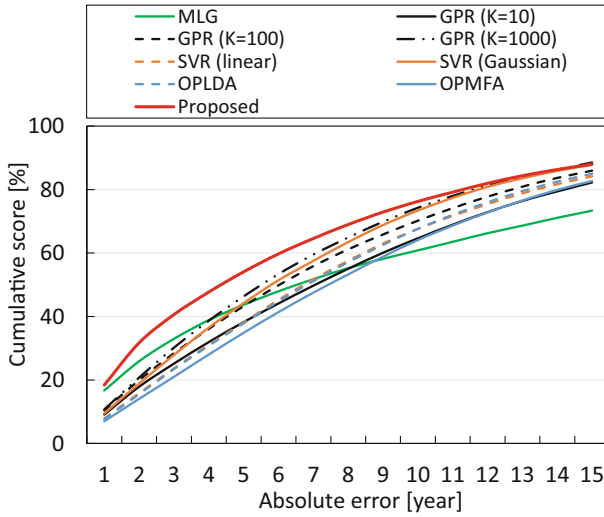
Fig. 8 Cumulative scores for the comparison with state-of-the-art methods with an absolute error tolerance from 1 to 15 years.
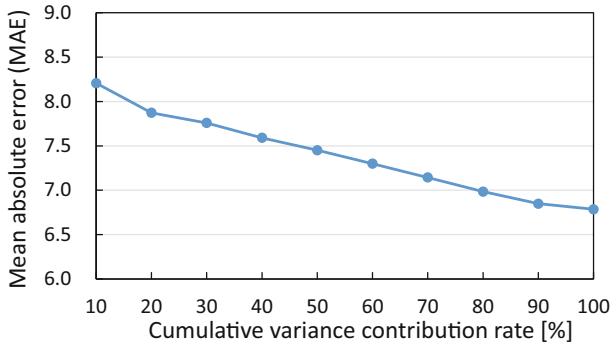
(a) Sensitivity analysis of r
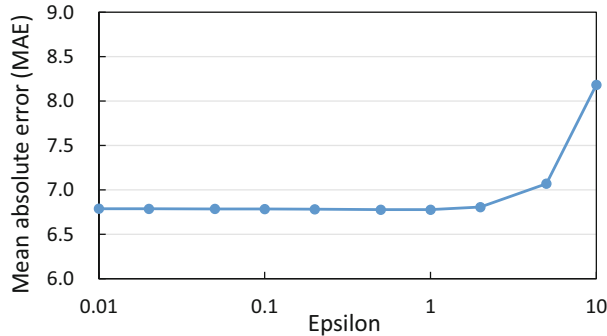
(b) Sensitivity analysis of ϵ

Fig. 9 Sensitivity analysis of cumulative variance contribution rate $r\%$ for PCA and ϵ for SVR

Table 4 Computational time of the proposed method, age group-independent method, and GPR ($K = 1,000$)

Method	Averaged computational time per test sample [ms]
GPR (K = 1,000)	4749.21
Age group-independent	***25.38***
Proposed	**2.69**

Bold and bold-italic fonts indicate the best and second-best results

We varied r within the range of $[10, 20, \ldots, 100]$ and ϵ within the range of $[0.01, 0.02, 0.05, 0.1, \ldots, 10]$. The MAEs for the sensitivity analysis are shown in Fig. 9. Figure 9a shows that the proposed method achieves lower MAE as the contribution rate $r\%$ increases because the gait feature includes more information with more cumulative variance. When the variance changes from 90 to 100%, the MAE changes very slightly from 6.85 to 6.78. Figure 9b shows that the proposed method is insensitive to ϵ as long as it is smaller than a certain value (e.g., 1).

4.7 Evaluation on computational time

To evaluate the computational time, we ran the MATLAB code of the proposed method on a PC with an Intel Core i7 4.00 GHz processor and 32 G RAM. We also computed two other methods for comparison. One is GPR ($K = 1,000$) [25], which achieves the second-best MAE. The other one is the age group-independent regression-based method. Averaged computational times per test sample for each method are shown in Table 4. It is evident that the proposed method has a much lower computational time than the second-best method (GPR ($K = 1,000$)). Moreover, because of the relatively small number of training samples for regression in the age group-dependent framework, the proposed method is approximately 10 times faster than the age group-independent method. Thus, the proposed method is more suitable for real applications.

5 Conclusion

In this paper, we described an age group-dependent manifold learning and regression method for gait-based human age estimation. Specifically, we first defined five optimal age groups to balance the tradeoff between age group classification and age group-dependent age regression. We then learned a classifier for multiple age groups using DAGSVM. We finally trained the age group-dependent SVR with a Gaussian kernel for human age estimation on classified test samples, in conjunction with age group-dependent OLPP, to better characterize the gait feature. Experimental results on OULP-Age show the state-of-the-art performance of the proposed method.

For future work, we plan to define more appropriate age groups using a learning-based method, and design more efficient age group classifiers to enhance the performance of the proposed method for human age estimation. Moreover, while we demonstrated the effectiveness of age group-dependent framework for gait-based human age estimation under the basic regressor (i.e., SVR) in this paper, the age group-dependent framework, however, could also be easily incorporated with deep learning-based approaches (i.e., replace the DAGSVM for age group classification and SVR for age estimation by deep learning

models). We therefore plan to leave combination of the deep learning-based approaches and the proposed age group-dependent framework for our future studies.

Additionally, because the proposed method can be regarded as a fusion of classification-based and regression-based method, its application is not limited to human age estimation. It could also be of great importance for the medical domain, particularly, on predicting Parkinson's disease, where typical symptoms (e.g., tremors) appear in gait patterns [28].

References

1. Begg R (2005) Support vector machines for automated gait classification. IEEE Trans Biomed Eng 52(5):828–838
2. Bouchrika I, Goffredo M, Carter J, Nixon M (2011) On using gait in forensic biometrics. J Forensic Sci 56(4):882–889
3. Cai D, He X, Han J, Zhang HJ (2006) Orthogonal laplacianfaces for face recognition. IEEE Trans Image Process 15(11):3608–3614
4. Chang KY, Chen CS (2015) A learning framework for age rank estimation based on face images with scattering transform. IEEE Trans Image Process 24(3):785–798
5. Chuen BKY, Connie T, Song OT, Goh M (2015) A preliminary study of gait-based age estimation techniques. In: 2015 Asia-pacific signal and information processing association annual summit and conference (APSIPA), pp 800–806
6. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
7. Davis JW (2001) Visual categorization of children and adult walking styles. Springer, Berlin, pp 295–300
8. Forman G (2003) An extensive empirical study of feature selection metrics for text classification. J Mach Learn Res 3:1289–1305
9. Fu Y, Huang TS (2008) Human age estimation with regression on discriminative aging manifold. IEEE Trans Multimedia 10(4):578–584
10. Geng X, Zhou ZH, Smith-Miles K (2007) Automatic age estimation based on facial aging patterns. IEEE Trans on Pattern Recognition and Machine Intelligence 29(12):2234–2240
11. Guo G, Fu Y, Dyer CR, Huang TS (2008) Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Trans Image Process 17(7):1178–1188
12. Guo G, Mu G, Fu Y, Huang TS (2009) Human age estimation using bio-inspired features. In: Proceedings of the 22nd IEEE conference on computer vision and pattern recognition, Miami, FL, USA, pp 112–119
13. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322
14. Ince OF, Park J, Song J, Yoon B (2014) Child and adult classification using ratio of head and body heights in images. International Journal of Computer and Communication Engineering 3(2):120
15. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
16. Iwama H, Muramatsu D, Makihara Y, Yagi Y (2013) Gait verification system for criminal investigation. IPSJ Transactions on Computer Vision and Applications 5:163–175

17. Li X, Makihara Y, Xu C, Muramatsu D, Yagi Y, Ren M (2016) Gait energy response function for clothing-invariant gait recognition. In: Proceedings of the 13th Asian conference on computer vision (ACCV 2016), Taipei, Taiwan, pp 257–272

18. Liu Z, Sarkar S (2004) Simplest representation yet for gait recognition: averaged silhouette. In: Proceedings of the 17th international conference on pattern recognition, vol 1, pp 211–214

19. Lu J, Tan YP (2010) Gait-based human age estimation. IEEE Trans Inf Forensics Secur 5(4):761–770

20. Lu J, Tan YP (2010) Ordinary preserving manifold analysis for human age estimation. In: IEEE computer society and IEEE biometrics council workshop on biometrics 2010, San Francisco, CA, USA, pp 1–6

21. Lu J, Tan YP (2013) Ordinary preserving manifold analysis for human age and head pose estimation. IEEE Transactions on Human-Machine Systems 43(2):249–258

22. Lynnerup N, Larsen P (2014) Gait as evidence. IET Biom 3:47–54

23. Makihara Y, Yagi Y (2008) Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In: Proceedings of the 19th international conference on pattern recognition. Tampa, Florida, USA

24. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: Proceedings of the 9th european conference on computer vision, Graz, Austria, pp 151–163

25. Makihara Y, Okumura M, Iwama H, Yagi Y (2011) Gait-based age estimation using a whole-generation gait database. In: Proceedings of the international joint conference on biometrics (IJCB2011), Washington D.C., USA, pp 1–6

26. Makihara Y, Kimura T, Okura F, Mitsugami I, Niwa M, Aoki C, Suzuki A, Muramatsu D, Yagi Y (2016) Gait collector: an automatic gait data collection system in conjunction with an experience-based long-run exhibition. In: Proceedings of the 8th IAPR international conference on biometrics (ICB 2016), Halmstad, Sweden, O17, pp 1–8

27. Mannami H, Makihara Y, Yagi Y (2010) Gait analysis of gender and age using a large-scale multi-view gait database. In: Proceedings of the 10th Asian conference on computer vision, Queenstown, New Zealand, pp 975–986

28. Miljkovic D, Aleksovski D, Podpečan V, Lavrač N, Malle B, Holzinger A (2016) Machine Learning and Data Mining Methods for Managing Parkinson's Disease. Springer International Publishing, Cham, pp 209–220

29. Nabila M, Mohammed AI, Yousra BJ (2018) Gait-based human age classification using a silhouette model. IET Biometrics 7(2):116–124

30. Nixon MS, Tan TN, Chellappa R (2005) Human identification based on gait. Int. Series on Biometrics. Springer, Berlin

31. Platt JC, Cristianini N, Shawe-Taylor J (2000) Large margin dags for multiclass classification. In: Solla SA, Leen TK, Müller K (eds) Advances in neural information processing systems 12. MIT Press, pp 547–553

32. Punyani P, Gupta R, Kumar A (2018) Human age-estimation system based on double-level feature fusion of face and gait images. International Journal of Image and Data Fusion 0(0):1–15

33. Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning, The MIT Press, Cambridge

34. Sarkar S, Phillips J, Liu Z, Vega I, ther PG, Bowyer K (2005) The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27(2):162–177

35. Vapnik VN (1998) Statistical learning theory. Wiley-Interscience

36. Wang L, Tan T, Hu W, Ning H (2003) Automatic gait recognition based on statistical shape analysis. IEEE Trans Image Process 12(9):1120–1131

37. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemometr Intell Lab Syst 2(1):37–52

38. Xu C, Makihara Y, Li X, Yagi Y, Lu J (2016) Speed invariance vs. stability: cross-speed gait recognition using single-support gait energy image. In: Proceedings of the 13th Asian conference on computer vision (ACCV 2016), Taipei, Taiwan, pp 52–67

39. Xu C, Makihara Y, Ogi G, Li X, Yagi Y, Lu J (2017) The ou-isir gait database comprising the large population dataset with age and performance evaluation of age estimation. IPSJ Transactions on Computer Vision and Applications 9(1):24

40. Zhang Y, Yeung DY (2010) Multi-task warped gaussian process for personalized age estimation. In: Proceeding of IEEE computer society conferene on computer vision and pattern recognition 2010, San Francisco, CA, USA, pp 1–8
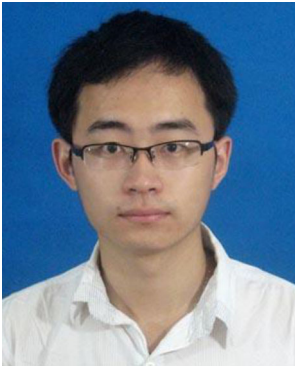
Xiang Li received the B.S. degree in computer science and technology in Nanjing University of Science and Technology (NUST) in China in 2012. He is currently a Ph.D. candidate at NUST. Since Jan. 2016, he has been with the Institute of Scientific and Industrial Research, Osaka University as a visiting researcher. His research interests are gait recognition, image processing and machine learning.

Yasushi Makihara received the B.S., M.S., and Ph.D. degrees in Engineering from Osaka University in 2001, 2002, and 2005, respectively. He is currently an Assistant Professor of the Institute of Scientific and Industrial Research, Osaka University. His research interests are gait recognition, morphing, and temporal super resolution. He is a member of IPSJ, RJS, and JSME.

Chi Xu received the B.S. degree in computer science and technology from Nanjing University of Science and Technology (NUST), China, in 2012. She is currently a Ph.D. candidate in pattern recognition and intelligent system at NUST. Since Jan. 2016, she has been with the Institute of Scientific and Industrial Research, Osaka University, Japan, as a visiting researcher. Her research interests are gait recognition, machine learning and image processing.

Yasushi Yagi is the Executive Vice President of Osaka university. He received his Ph.D. degrees from Osaka University in 1991. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a Research Associate in 1990, a Lecturer in 1993, an Associate Professor in 1996, and a Professor in 2003 at Osaka University. He was also Director of the Institute of Scientific and Industrial Research, Osaka university from 2012 to 2015. International conferences for which he has served as Chair include: FG1998 (Financial Chair), OMINVIS2003 (Organizing chair), ROBIO2006 (Program co-chair), ACCV2007 (Program chair), PSVIT2009 (Financial chair), ICRA2009 (Technical Visit Chair), ACCV2009 (General chair), ACPR2011 (Program co-chair) and ACPR2013 (General chair). He has also served as the Editor of IEEE ICRA Conference Editorial Board (2007–2011). He is the Editorial member of IJCV and the Editor-in-Chief of IPSJ Transactions on Computer Vision and Applications. He was awarded ACM VRST2003 Honorable Mention Award, IEEE ROBIO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 Finalist for Best Vision Paper, MIRU2008 Nagao Award, and PSIVT2010 Best Paper Award. His research interests are computer vision, medical engineering and robotics. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.
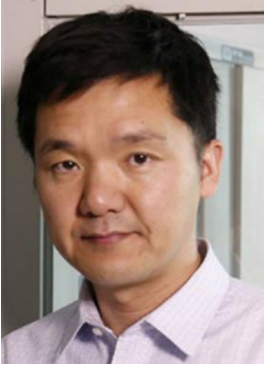
Mingwu Ren received the Ph.D. degree in pattern recognition and intelligent system from Nanjing University of Science and Technology (NUST), Nanjing, P.R. China in 2001. He is currently a Professor with the School of Computer Science and Engineering, NUST. His current research interests include computer vision, image processing, and pattern recognition.