How Confident Are You in Your Estimate of a Human Age? Uncertainty-aware Gait-based Age Estimation by Label Distribution Learning

Atsuya Sakata[1], Yasushi Makihara[1], Noriko Takemura[1], Daigo Muramatsu[2,1], Yasushi Yagi[1]
[1]Osaka University [2]Seikei University
{sakata, makihara, takemura, yagi}@am.sanken.osaka-u.ac.jp muramatsu@st.seikei.ac.jp

Abstract

Gait-based age estimation is one of key techniques for many applications (e.g., finding lost children/aged wanders). It is well known that the age estimation uncertainty is highly dependent on ages (i.e., it is generally small for children while is large for adults/the elderly), and it is important to know the uncertainty for the above-mentioned applications. We therefore propose a method of uncertainty-aware gait-based age estimation by introducing a label distribution learning framework. More specifically, we design a network which takes an appearance-based gait feature as an input and outputs discrete label distributions in the integer age domain. Experiments with the world-largest gait database OULP-Age show that the proposed method can successfully represent the uncertainty of age estimation and also outperforms or is comparable to the state-of-the-art methods.

1. Introduction

Gait has been regarded for the last two decades as unique behavioural biometrics which can be used even at a distance from a camera without subjects' cooperation [29]. Due to the above-mentioned advantages, the gait-based person verification/identification has been expected to be applied to many applications such as surveillance, access control, and criminal investigation using CCTV footage. In fact, there have been several cases where the gait recognition has been applied for criminal investigation and forensics [3]. In addition to identities [36, 32, 26, 39], the gait may contain a variety of cues such as gender [40, 15, 19], age [6, 1, 25, 22, 23, 21, 42], ethnicity [41], disease [16], emotion [37], some qualitative gait attribute [33].

Among them, gait-based age estimation has many potential applications. For example, once a visitor's age is estimated in a shopping mall, an advertiser can change a content of a digital signage into more suitable one for the estimated age. An age-based access control is also possible based on the estimated age, e.g., people can prevent people under age from buying alcohol or cigarette. Moreover, in a forensic scenario, a criminal investigator may get a witness regarding a rough age information of a perpetrator/suspect, and thereafter may be able to automatically retrieve candidates from CCTV footage whose estimated age by gait match the witness. It would be also possible to search people like wandering elderly and children got lost in CCTV footage with the help of the gait-based age estimator.

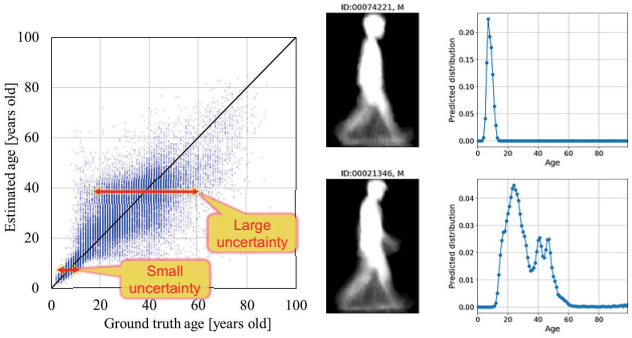


Figure 1: An example of a scatter plot between ground-truth age (horizontal axis) and estimated ages (vertical axis) by [30] (left) and probability distributions of estimated ages by the proposed method (right) given an input of a silhouette-based gait feature (middle). While we obtain a relatively sharp age probability distribution (i.e., one with small uncertainty) for the upper subject (a child), we obtain a relatively diverse probability distribution (i.e., one with large uncertainty) for the lower subject (an adult). The proposed method indicates that more age variations are likely to occur for the lower subject's gait feature than the upper one.

Technically speaking, relevant work on gait-based age analysis mainly falls into two families: age-group classification [6, 1, 27] and age regression [25, 21, 23, 22, 17]. While the above-mentioned studies employ classical machine learning techniques such as support vector machine [2] and support vector regression [35] due to a limited number of training data (e.g., [32]), current studies

978-1-7281-9186-7/20/$31.00

on gait-based age estimation mainly employ deep learning-based approaches [28, 18, 42] thanks to recently released much larger-scale gait databases including over 60,000 subjects [38].

Whatever the approaches to gait-based age estimation are employed, we almost always observe that age estimation accuracies are indeed age-dependent. By taking a closer look at a scatter plot between the ground-truth age and the estimated age with one of the above-mentioned deep learning approaches [30] (see Figure 1 (left)), we can clearly see that the uncertainty (or degree of the diversity of ground truth ages) for each of an estimated age is quite different between a child and an adult. More specifically, while the uncertainty is small for a child (e.g., less than ± 2 years), it is quite large for an adult (e.g., ± 20 years). In fact, it is relatively difficult even for us, humans, to accurately tell how old the subject is from gait.

As a trial, we would like the readers to guess two subjects' ages in Figure 1 (middle) given inputs of silhouette-based gait features called gait energy image (GEI) [10]. As for the upper subject, we expect that the readers can reach relatively fine-grained age estimation range (e.g., from 5 years old to 10 years old). On the other hand, we expect that the readers struggle against the age estimation for the lower subject. A reader would estimate it as 20 years old, another may guess as 30 years old, or some might say 40 years old. The age-dependent uncertainty difference of gait-based age estimation is derived from several points. First, a growth or a decline of gait (and also body shape) is relatively fast for children and the elderly, while it is quite slow for adults, and hence it is relatively difficult to tell the age difference in adulthood due to the slow change of the gait feature. Second, while facial images have texture cues to estimate an age (e.g., wrinkles and dull skin appears more as age progression), a gait image does not have such texture-based cues. Third, although a gait image still has some cues to estimate an age such as stoop and middle age spread, they do not only depend on the age but also highly depend on individuality and individual lifestyle habits. For example, a person in his/her forties, who keeps a slim body shape by good dietary and exercise habits, may look younger than his/her age, and vice versa. In other words, there may be some subjects whose gait features are almost the same but whose ages are different, which results in an uncertain age estimation result from gait.

This kind of uncertainty (or confidence) of the age estimation plays an important role in some applications. Let's consider a person search scenario by age. In case of searching a lost child with five years old, since the system is relatively confident in its estimate of human age for children, it would be enough to show a list of people whose estimated age is within a limited age range (e.g., five ± 2 years). On the other hand, in case of searching a suspect with his/her thirties, it would be necessary to show a large number of people whose estimated age is within a large age range (e.g., from twenties to forties). In short, if the system is aware of the uncertainty (or confidence) of the estimated age, it can appropriately bound a age range of people candidates for a search target.

The answer is 30 years old.

While most of the existing approaches to the gait-based age estimation output a single age as an estimation result, there is an only exception [25] which outputs an estimated age and its uncertainty. The work [25] employs a framework of Gaussian process regression (GPR) for the gait-based age estimation, which outputs a Gaussian distribution for an estimated age (i.e., a mean and a variance) given an input gait feature as well as a training set of gait features and corresponding ages. The uncertainty (i.e., the variance of the estimated age) provided by the GPR framework is mainly derived from the degree of closeness from a test gait feature to the training gait features, i.e., the uncertainty gets smaller as the test gait feature is relatively closer to one of the training gait features, and vice versa. In that sense, the GPR framework cannot handle the age uncertainty derived from similar gait features with different ages as mentioned above. More specifically, the system should ideally return a large uncertainty if a test gait feature is close to a cluster of similar training gait features but with different ages, while the GPR returns small uncertainty even under such a situation.

In order to overcome such a difficulty, we propose a deep neural network-based approach to the gait-based age estimation which outputs not a single estimated age but a probability distribution of the estimated age. The contributions of this work are two-fold.

1. Uncertainty-aware gait-based age estimation using a label distribution Unlike the GPR framework, the proposed method can better cope with the above-mentioned similar gait features but with different ages. More specifically, we introduce a label distribution [7] as an output of an age estimation network given a gait feature as an input. The age estimation network is trained so as to minimize a joint loss function of (1) Kullback–Leibler (KL) divergence [14] for the uncertainty representation and (2) mean absolute error (MAE) for the ordinary preserving property, in the same way that the state-of-the-art facial age estimation [8] does so. As such, we can assign probabilities to the multiple different ages as shown in Figure 1 (right).

2. State-of-the-art accuracy on gait-based age estimation We achieved state-of-the-art accuracy on gait-based age estimation in standard evaluation metrics such as MAE and cumulative score, using OULP-Age [38], the world-largest publicly available gait database containing over 60,000 subjects with wide age diversity ranging from 2 to

90 years old.

2. Related work

2.1. Gait-based age estimation

Early-stage studies on gait-based age analysis mainly handled age group classification problems (e.g., children vs. adults) using body joints-based gait parameters. For example, Davis et al. [6] classified children and adults based on a biological motion cue (i.e., point light sources on the body joints) and Begg et al. [1] classified the younger and the elderly based on minimum foot clearance from the ground [1]. Mannami et al. [27] analyzed the differences of an image-based gait feature such as gait energy image (GEI) [10] (a.k.a. averaged silhouette [20]) and frequency-domain feature [26] from multiple views among four age/gender groups, i.e., children, adult females, adult males, and the elderly.

Thereafter, studies on gait-based age estimation have been conducted since 2010, by using the image-based gait features [10, 26] in conjunction with classical machine learning techniques. For example, Lu et al. proposed ordinary preserving manifold learning in [22], and also proposed its extensions: ordinary preserving linear discriminant analysis (OPLDA) and ordinary preserving margin Fisher analysis (OPMFA) in [23]. Xu et al. [38] constructed the world-largest gait database named OULP-Age and evaluated multiple benchmark methods of gait-based age estimation including [22, 23] and also a method using support vector regression (SVR) [35]. Li et al. [17] proposed a multi-stage framework which combines age group classification and subsequent age regression for each age group by using a couple of machine learning techniques such as support vector machine (SVM), manifold learning, and SVR.

In addition to the above-mentioned classical machine learning-based approaches, researchers have started to employ deep learning-based approaches to the gait-based age estimation. For example, Sakata et al. [30] employed DenseNet [12], one of the start-of-the-art deep neural network architecture and validated its effectiveness in the gait-based age estimation task. In addition to age itself, other attributes such as gender, age group, and/or identity were incorporated in multi-task learning frameworks [28, 42] and a multi-stage learning framework [31]. Moreover, a method of gait-based age estimation robust against carried object was proposed in [18].

While the all the above-mentioned approaches output a single value as an estimated age without taking the uncertainty into consideration, work [25] employed a GPR framework [4] for gait-based age estimation, which outputs a Gaussian distribution of the estimated age, more specifically, its mean and variance (i.e., a sort of uncertainty). As we have already addressed in the introduction section, the GPR framework can, however, not better handle the uncertainty induced by similar gait features but with different ages. We will later describe more details of this issue with a preliminary simulation experiment in section 3. Unlike the GRP framework, our method can better handle the uncertainty by employing a label distribution framework, where the system outputs a discrete probability distribution over integer age labels as an output given a gait feature as an input.

The proposed method may have a similarity in terms of representation of integer age labels to a previous study [21]. More specifically, Lu et al. [21] cast an age estimation problem into a classification problem of integer-age classes (labels) using multi-label guided (MLG) subspace learning. Although they employ the integer age labels similarly to the proposed method, they still output a single integer age label as a result of a classification process, unlike the proposed method output a probability distribution of age.

2.2. Face-based age estimation using a label distribution

Since a face is the most frequently used biometric modality for age estimation, the face-based age estimation enjoys a rich body of literature. Among them, we briefly introduce the existing work of face-based age estimation using a label distribution as more closely relevant work with us.

Geng et al. [9] has introduced a concept of label distribution learning into a face-based age estimation for the first time and proposed two algorithms of learning, named improved iterative scaling-learning from label distribution (IIS-LLD) and conditional probability neural network (CPNN). Zhao and Wang [43] proposed a strategic decision-making learning from label distribution, which copes with different types of age label distributions such as Gaussian-type, triangle-type, and box-type distributions. He et al. [11] proposed a data-dependent label distribution learning, where neighboring training samples to a test sample are first selected based on the face affinity graph and the label distribution is then constructed based on the cross-age correlations among neighboring face samples.

Gao et al. [7] has introduced a deep learning-based approaches to the facial age estimation with a label distribution, named deep label distribution learning (DLDL) which employs KL-divergence to measure distribution similarity. They also extended DLDL in [8] so as to cope with a regression problem as well as the label distribution estimation at the same time by introducing the joint loss function of an MAE and the KL divergence.

As mentioned above, label distributions have been already employed in facial age estimation community. The gait analysis community has never employed the useful label distribution for age estimation and hence this work introduces it for the purpose of gait-based age estimation the

first time, to the best of our knowledge.

3. Observation on GPR-based approach [25]

Before moving on the proposed method, we would like to briefly mention to the most closely relevant work, i.e., the GPR-based approach to uncertainty-aware gait-based age estimation which outputs Gaussian distribution of an estimated age, i.e., an expectation and a variance, in order to clarify its drawback. We may refer the readers to [25] for more details.

In the GPR-based approach, we assume that a training set $D = [X, \boldsymbol{y}]$ is given, where $X = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N]$ is a set of N samples of gait features and $\boldsymbol{y} = [y_1, \cdots, y_N]$ is a set of the corresponding ground truth ages. In addition, an affinity/similarity function, i.e., an inner product between two feature vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined, often by a non-linear kernel function such as a radial basis function (RBF)

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j; r) = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2r^2}\right), \quad (1)$$

where $\|\cdot\|$ is the L_2 norm and r is a hyper-parameter for the RBF kernel. Note that the function $k(\cdot, \cdot)$ measures the closeness between two input arguments and that it approaches 1 if the two inputs are more similar, while it approaches 0 if they are more different.

Thereafter, given an input gait feature $\boldsymbol{x}_*$, the GPR estimates the posterior probability distribution of age y_* corresponding to the gait feature $\boldsymbol{x}_*$ based on the training set D. According to the theory of Gaussian processes [4], once we assume each age y_i in the training set D follows a Gaussian distribution $\mathcal{N}(y; y_i, \Delta^2)$, where Δ^2 is a variance of age observation noise, the posterior probability distribution $P(y_*|\boldsymbol{x}_*, D)$ also follows a Gaussian distribution $\mathcal{N}(y_*; \mu_y, \sigma_y^2)$, where mean μ_y and variance σ_y^2 are defined as

$$\mu_y = \boldsymbol{k}_*^T (K + S)^{-1} \boldsymbol{y} \quad (2)$$

$$\sigma_y^2 = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^T (K + S)^{-1} \boldsymbol{k}_* + \Delta^2, \quad (3)$$

where K is an $N \times N$ square matrix whose (i, j)-th component is $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{k}_*$ is an N-dimensional vector whose i-th row is $k(\boldsymbol{x}_i, \boldsymbol{x}_*)$, S is an $N \times N$ diagonal matrix whose (i, i)-th component is the age observation noise Δ^2.

According to Eq. (3), we notice that the variance (i.e., the uncertainty) is dependent not on ground-truth ages $\boldsymbol{y}$ in the training set but on relations, more specifically, affinity/similarity/closeness between the input gait feature $\boldsymbol{x}_*$ and those in the training data X, which appear in K and $\boldsymbol{k}_*$ in the above-defined equations.

In order to better observe and understand this, let us show a preliminary simulation experiment with 10 training data of one-dimensional feature vectors (i.e., a scalar value) and an estimation target (e.g., an age) in Fig. 2. We can see

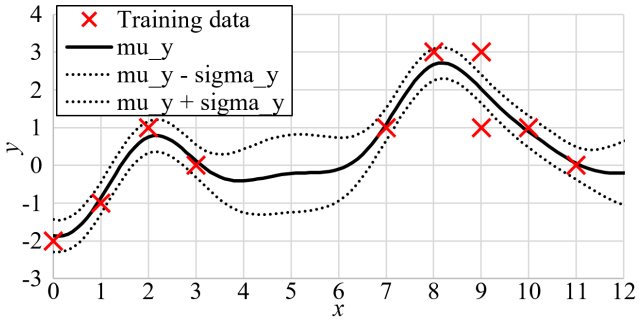


Figure 2: A preliminary simulation experiment for GPR with 10 training data and hyper-parameters $r = 1.0$ and $\Delta = 0.1$. The uncertainty range $\mu_y \pm \sigma_y$ gets larger if an input is more distant from any of the training data (e.g., $x = 5$). On the other hand, the uncertainty does not get large even if an input is closer to the training data with multiple outputs (e.g., $x = 9$ with $y = 1$ and $y = 3$).

that the uncertainty gets larger as an input get more distant from any of the training data (e.g., $x = 5$). Moreover, in this simulation experiment, a training sample $x = 9$ has multiple outputs $y = 1$ and $y = 3$, which can be a sort of analogy to subjects who have similar gait features but have different ages. Despite it is preferable to well represent the uncertainty due to the different ages for the sample, the uncertainty derived from the GPR does not get large. This is the very drawback of the GPR framework.

On the other hand, the label distribution learning framework can successfully cope with this kind of uncertainty because it can be trained to output multiple ages. For example, given the sample $x = 9$ in the above-mentioned simulation experiment, we intuitively expect that probability 0.5 is assigned to each of $y = 1$ and $y = 3$.

4. Proposed method

4.1. Overview

The proposed method tries estimating an age of a target person from gait. We introduce an overview of the proposed method at first as shown in Fig. 3. Given a gait feature as an input, a backbone network (i.e., an age estimation model) output a label distribution of an estimated age. Once the label distribution is obtained, we can also compute an expectation of the label distribution as a single estimated age. We will explain details of individual components in the following subsections.

4.2. Input data

The first step to realize the gait-based age estimation, is to prepare input data to the age estimation model. For this purpose, we need to firstly capture gait data, and then extract an efficient gait feature from the gait data. We can

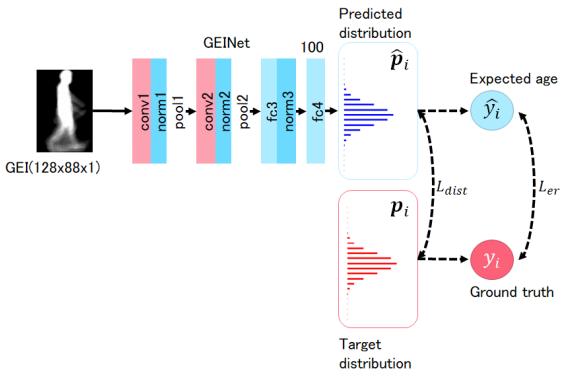


Figure 3: Schematic of the proposed age estimation model with label distribution learning

Table 1: Layer configurations

Layer	#Kernels	Kernel size	Stride
conv1	81	5×5	1
pool1		3×3	2
conv2	45	7×7	1
pool2		2×2	2
fc3	1024		
fc4	100		

consider multiple sensors such as image sensors (cameras), depth-sensor, and inertial-sensor. Among them, cameras are the most popular sensors and already installed in many places (e.g., CCTVs in public spaces), we therefore focus on gait image sequence captured using cameras as gait data. We then extract GEI [10] from the gait image sequence, which is the most widely used image-based gait feature.

4.3. Network structure

As a backbone network, we employ a convolutional neural network designed for gait recognition named GEINet [34] and modify it for age estimation. The modified GEINet is composed two sequential triplets, which include convolution, pooling, and normalization layers, fully connected layers with normalization, and another fully connected layer with activation function of Softmax. Layer configurations of the modified network are summarized in Table 1. The number of units for the fully connected layer fc4 is set so to be the same as the number of bins for discrete label distribution explained in subsection 4.4.

4.4. Output representation

There are two major output/ground-truth representations. One is a scalar value for regression-based methods [22, 23, 17, 30, 31], and the other is a one-hot vector for classification-based methods [21]. Different from these

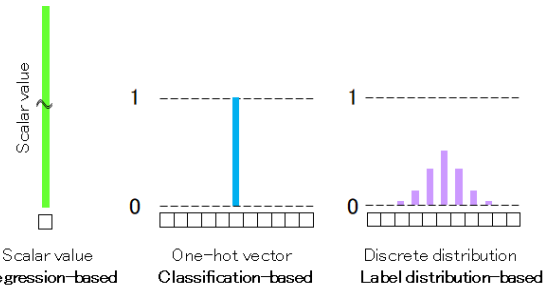


Figure 4: Conceptual difference of output representations

ways, in this paper, we incorporate the idea of label distribution [9, 8]. A label distribution-based method assumes that the ground-truth is neither a scalar value nor a one-hot vector, but a discrete distribution of age. Figure 4 shows a conceptual difference of these tree representations.

Let K be the number of bins for a discrete probability distribution of integer ages; we set the minimum and maximum ages to 0 and $K-1$, respectively. Let $y_i \in \mathbb{R}$ be a ground-truth age of the i-th training data. Besides, let $\boldsymbol{p}_i$ and $p_{i,k}$ be assigned discrete probability distribution and probability for age k associated with i-th data, respectively. In the proposed method, we set the $p_{i,k}$ so as to follow a Gaussian distribution whose mean and standard deviation is the ground-truth age y_i and σ as

$$p_{i,k} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(k-y_i)^2}{2\sigma^2}\right\}. \tag{4}$$

Here, σ is a hyper-parameter to control the uncertainty of the ground-truth age.

4.5. Loss function

Because a ground-truth label is described by a discrete probability distribution, we can consider two different criteria for evaluating the goodness of the trained parameters. The first criterion measures a similarity between an assigned target probability distribution and an estimated probability distribution. The other criterion is the dissimilarity between the ground-truth age and an expected age calculated from the estimated probability distribution.

Let $\hat{\boldsymbol{p}}_i = [\hat{p}_{i,0}, \ldots, \hat{p}_{i,K-1}]^T \in \mathbb{R}^K$ be an estimated discrete probability distribution for the i-th training data ($i = 1, \ldots, N$), where N is the number of training data. Note that the integer age for the k-th age label is k and its probability is $\hat{p}_{i,k}$.

For the first criterion, we consider KL divergence [14] between two distributions and set the loss function L_{KL} as follows:

$$L_{\mathrm{KL}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{K-1} p_{i,k} \ln \frac{p_{i,k}}{\hat{p}_{i,k}}. \tag{5}$$

For the other criterion, we calculated the expected age from the estimated distribution, and measure an MAE between the ground-truth age and the expected age as below:

$$L_{\mathrm{MAE}} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \tag{6}$$

$$\text{where } \hat{y}_i = \sum_{k=0}^{K-1} k\hat{p}_{i,k}. \tag{7}$$

We finally define a joint loss function L using the two loss functions as

$$L = \lambda_{\mathrm{KL}}L_{\mathrm{KL}} + \lambda_{\mathrm{MAE}}L_{\mathrm{MAE}}, \tag{8}$$

where λ_{KL} and λ_{MAE} are hyper-parameters for balancing the two loss functions.

5. Experiments

5.1. Data set

We used the OU-ISIR Gait Database, Large Population Dataset with Age (*OULP-Age*) [38] to evaluate the performance of the proposed method. *OULP-Age* is the world's largest gait database that includes gait images as well as the ground-truth of age and gender. It consists of 63,846 gait images (31,093 males and 32,753 females) with age ranging from 2 to 90 years. GEIs of 88×128 pixels extracted for a side-view gait are provided for each subject. Based on the predefined protocol, the database was divided into a training set composed of 31,923 subjects (15,596 males and 16,327 females) and a test set composed of 31,923 subjects (15,407 males and 16,426 females).

5.2. Training

The network was trained so as to minimize the joint loss function (Eq. (8)) using adaptive moment estimation (Adam) [13] with the batch size of 128 and 100 epochs. An initial learning rate was set to 0.001. The number of age labels is set to $K = 101$, i.e., the minimum and maximum ages are 0 and 100 years old. The hyper-parameter σ for label distribution was set to 1.0 and the both balancing parameters λ_{KL} and λ_{MAE} of the joint loss function were set to the same value, 1.0.

5.3. Evaluation measure

We evaluated the accuracy of gait-based age estimation using two criteria. One is MAE between the estimated age (i.e., an expectation of the age label distribution) and a ground-truth age, which is similarly computed as the MAE-based loss function (Eq. (6)).

The other one is a cumulative score (CS) which shows an error tolerance ratio. Specifically, we define the number of test samples whose absolute error between an estimated age and the ground-truth age is less than or equal to y as $n(y)$, and then CS of the y-year absolute error as

$$\mathrm{CS}(y) = \frac{n(y)}{N}. \tag{9}$$

5.4. Qualitative evaluation on the label distribution

First of all, as the most important analysis, we show and compare the age distributions by the GPR-based method [25] and the proposed method in Fig. 5 in order to see how the age estimation uncertainty is well represented.

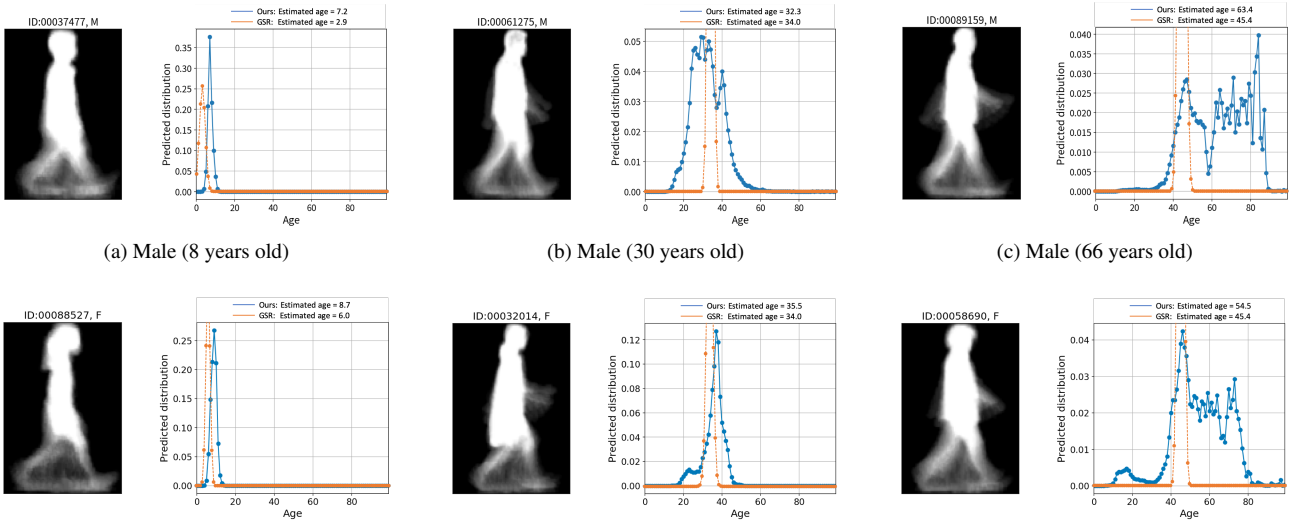
As for the GPR-based method, we can clearly see that the uncertainty (i.e., the variance of the Gaussian distribution, depicted by orange curves in Fig. 5) does not change much among children, adults, and the elderly, despite they should change so as to reflect the difference of ages in the uncertainty as described and shown in Fig. 1 (i.e., the small uncertainty for children and the large uncertainty for adults and the elderly). This is because the uncertainty by the GPR-based approach is mainly dependent just on the closeness of an input gait feature to any of the training gait feature, and hence cannot appropriately handle similar gait features but with different ages as we have already discussed in section 3.

On the other hand, the proposed label distribution learning-based approach can successfully cope with the age-dependent uncertainty. More specifically, the proposed method returns sharp distribution (i.e., small uncertainty) for children (Fig. 5 (left)), it returns diverse distributions (i.e., large uncertainty) for adults (Fig. 5 (middle)) and the elderly (Fig. 5 (bottom)), which is consistent with a common insight on gait-based age estimation as well as the scatter plots between the ground-truth and estimated ages (see Fig. 6). In addition, we can see that the ground-truth age for each subject is covered by relatively high probabilities in the estimated age distribution. Consequently, we confirmed that the proposed method is suitable for the purpose of the uncertainty-aware gait-based age estimation.

Moreover, we notice that the subject in Fig. 5 (e) has smaller uncertainty than the subject in Fig. 5 (b), who are both 30 years old. This may be because the outfit of the subject (e), more generally speaking, a sort of age generation-specific fashion style, could be a cue to narrow the possible age range. As such, we can see that the proposed method is potentially capable of handling not only age-dependent uncertainty but also sample-dependent uncertainty.

5.5. Comparison with state-of-the-arts

We compared the proposed method with benchmarks of the existing gait-based age estimation. As a baseline algorithm, a method using GPR with RBF kernel [25] and an active set method in the same way as in [24], where k nearest neighbors for each test sample are used for GPR



(a) Male (8 years old) (b) Male (30 years old) (c) Male (66 years old)

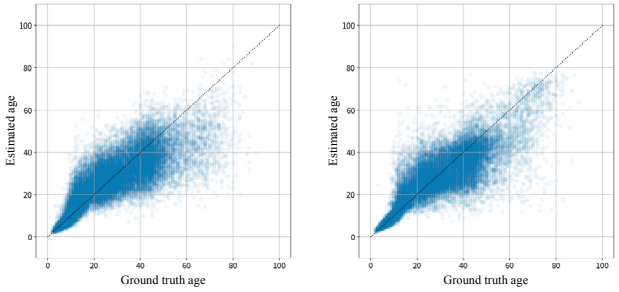
(d) Female (8 years old) (e) Female (30 years old) (f) Female (60 years old)

Figure 5: Pairs of GEIs and estimated age distributions by the GPR-based method [25] (depicted by orange) and the proposed method (depicted by cyan) for males/females and children/adults/the elderly. In each caption, ground-truth gender and age (in bracket) are provided. The GPR-based method returns similar ranges of the uncertainty (variance) regardless of the age. On the other hand, the proposed method successfully returns small and large uncertainties for children and adults/the elderly, which coincides with the intuition of the gait-based age estimation accuracy addressed in the introduction section and also as with the scatter plots in Fig. 6.

Table 2: MAEs [years] and CSs [%] at 1, 5, and 10 years absolute errors. Bold and Italic bold indicate the best and the second best performances, respectively. "-" indicates not provided in the original papers. The last five methods are deep learning-based approaches.

Method	MAE	CS(1)	CS(5)	CS(10)
MLG [21]	10.98	16.7	43.4	60.8
GPR ($k = 10$) [25]	8.83	9.1	38.5	64.7
GPR ($k = 100$) [25]	7.94	10.5	43.3	70.2
GPR ($k = 1000$) [25]	7.30	10.7	46.3	74.2
SVR (Liner)	8.73	7.9	38.2	67.6
SVR (Gaussian)	7.66	9.4	44.2	73.4
OPLDA [23]	8.45	7.7	37.9	67.6
OPMFA [23]	9.08	7.0	34.9	64.1
ADGMLR [17]	6.78	18.4	54.0	76.2
DenseNet [30]	5.79	22.5	55.9	80.4
Multi-task [42]	***5.47***	-	-	-
Multi-stage [31]	5.48	**25.3**	**62.6**	***82.0***
GEINet [34]	6.22	17.2	55.9	79.2
Ours	**5.43**	***24.8***	***62.3***	**82.4**

and $k = 10, 100, 1000$ were evaluated. We also evaluated existing methods of gait-based age estimation using conventional machine learning techniques such as SVR with linear and Gaussian kernels, denoted as SVR (linear) and SVR (Gaussian), MLG [21], OPLDA [23], OPMFA [23], and AGDMLR [17]. In addition, we employed a slightly modified version of GEINet [34] as a deep learning-based approach to age estimation. More specifically, the original GEINet outputs class (i.e., subject) likelihoods and hence the number of nodes at the last layer is equal to the number of subjects. On the other hand, the modified version of GEINet outputs an age and hence the last layer has just a single node. Moreover, we also evaluated the other state-of-the-art deep learning approaches [30, 31, 42] to gait-based age estimation.



(a) GEINet [34] (b) Proposed method

Figure 6: Scatter plot between ground truth age and estimated age. Diagonal lines show equal lines of ground truth age and estimated age.

MAEs and CSs for 1, 5, and 10 years tolerance are summarized in Table 2. As a result, deep learning-based

methods (i.e., GEINet [34], DenseNet [30], multi-task [42], multi-stage [31], and the proposed method) significantly outperform the other conventional machine learning-based methods. In addition, the proposed method outperforms or at least comparable to the state-of-the-art deep learning-based approaches [34, 30, 42, 31] despite the proposed method employs a relatively simple backbone network, which is the same as GEINet [34]. In fact, if we compare the proposed method with a regression-based method under the same backbone network, i.e., GEINet [34], we can see the proposed method clearly outperforms it by a large margin (e.g., the proposed method improves MAE and CS(1) by 0.79 years and 7.6%, respectively). This is because the proposed method with the age label distribution has a more powerful and flexible expression capability than the regression-based method which outputs a single age value. More specifically, while the regression-based model (e.g., GEINet [34]) is highly affected by outlier subjects who look much younger/older for his/her ages, the proposed method can mitigate the effect of the outliers by assigning probabilities to multiple age labels.

For further analysis, scatter plots between the ground-truth age and the estimated ages are shown for deep learning-based methods with the same backbone network, i.e., GEINet [34] as a regression-based approach and the proposed method as a label distribution-based approach in Fig. 6. In both cases, we can observe a common property in the gait-based age estimation that the age estimation uncertainty is small for children, while it is large for adults and the elderly. We then take a closer look at differences between them. While the estimated ages for GEINet (Fig. 6 (a)) is more deviated than the proposed method (Fig. 6 (b)) for all the age range, particularly, for the children under 15 years old. In addition, the proposed method has significantly improved the estimation accuracy for over 60 years old compared with GEINet. More specifically, while most of subjects over 60 years old are under-estimated (i.e., biased to younger direction) for GEINet, the proposed method successfully mitigates the under-estimate. This leads to the performance improvement of quantitative criteria such as MAEs and CSs.

5.6. Sensitivity analysis

Since the standard deviation σ of label distribution is one of key hyper-parameters for the proposed method, we analyze its sensitivity on gait-based age estimation accuracy.

Table 3 (left) shows MAEs when changing the standard deviation σ. According to Table 3 (left), the larger the σ is, the worse the MAE is. This would be partly because the ground-truth age in *OULP-Age* was given in the integer domain. Actually, the age annotation of *OULP-Age* was provided by each subject's self-declaration who participated in a long-run exhibition of video-based gait analysis [24], and hence the upper-bound of age annotation error is ideally less than one year unless he/she provided a wrong information. The smallest standard deviation $\sigma = 1$ yielded the best accuracy.

Table 3: MAEs [years] for sensitivity analysis of the standard deviation σ of the ground-truth label distribution (left) and those for ablation studies of the two loss functions L_{KL} and L_{MAE} (right).

σ	MAE
1.0	**5.43**
2.0	5.47
3.0	5.51
5.0	5.58
10.0	5.64

L_{KL}	L_{MAE}	MAE
✓		5.79
	✓	5.80
✓	✓	**5.43**

5.7. Ablation studies of loss functions

Moreover, we evaluated the effects of the two loss functions, i.e., the KL divergence-based loss function L_{KL} and the MAE-based loss function L_{MAE}.

Table 3 (right) shows MAEs when turning on and off the two loss functions (i.e., set the balancing parameter 1 and 0, respectively). The result shows that the MAE gets worse when we make either of the two loss functions invalid, in other words, both the KL divergence-based and the MAE-based loss play important roles in a complementary way each other. As a result, the joint loss function composed of the two loss functions yielded the best accuracy.

6. Conclusions

This paper described a described a method of uncertainty-aware gait-based age estimation. Unlike the existing uncertainty-aware approach such as the GPR-based method [25], the proposed method can successfully and naturally handle similar gait features but with different ages by introducing a label distribution learning framework. Experiments with the world-largest gait database *OULP-Age* showed that the proposed method can well represent the uncertainty of the age estimation and outperformed or was comparable to the state-of-the-art methods.

Since the proposed method employs a relatively simple backbone network [34], we will further improve the accuracy by introducing more state-of-the-art backbone network (e.g., [5]). Moreover, in the proposed method, the standard deviation σ which is a hyper-parameter to control the uncertainty of the target distribution, is the same for all the training data. We will therefore try realizing a better label distribution learning by employing an age-/sample-dependent σ to improve accuracy as another future work.

Acknowledgment This work was supported by JSPS KAKENHI Grant No. JP18H04115, JP19H05692, and JP20H00607.

References

[1] R. Begg. Support vector machines for automated gait classification. *IEEE Trans. on Biomedical Engineering*, 52(5):828–838, May 2005.

[2] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A discriminant analysis for undersampled data. In *Proc. of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152, 1992.

[3] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, 56(4):882–889, 2011.

[4] C. K. I. W. Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[5] H. Chao, Y. He, J. Zhang, and J. Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. In *Proc. of the 33rd AAAI Conf. on Artificial Intelligence (AAAI 2019)*, 2019.

[6] J. Davis. Visual categorization of children and adult walking styles. In *Proc. of the Int. Conf. on Audio- and Video-based Biometric Person Authentication*, pages 295–300, Jun. 2001.

[7] B. Gao, C. Xing, C. Xie, J. Wu, and X. Geng. Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing*, 26(6):2825–2838, 2017.

[8] B.-B. Gao, H.-Y. Zhou, J. Wu, and X. Geng. Age estimation using expectation of label distribution learning. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 712–718. International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[9] X. Geng, C. Yin, and Z. H. Zhou. Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2401–2412, 2013.

[10] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316– 322, 2006.

[11] Z. He, X. Li, Z. Zhang, F. Wu, X. Geng, Y. Zhang, M. Yang, and Y. Zhuang. Data-dependent label distribution learning for age estimation. *IEEE Transactions on Image Processing*, 26(8):3846–3858, 2017.

[12] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, July 2017.

[13] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. pages 1–15, 2014.

[14] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

[15] L. Lee and W. Grimson. Gait analysis for recognition and classification. In *Proc. of the 5th IEEE Conf. on Face and Gesture Recognition*, volume 1, pages 155–161, 2002.

[16] M. Lemke, T. Wendorff, B. Mieth, K. Buhl, and M. Linnemann. Spatiotemporal gait patterns during over ground locomotion in major depression compared with healthy controls. *Journal of Psychiatric Research*, 34:277–283, 2000.

[17] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren. Gait-based human age estimation using age group-dependent manifold learning and regression. *Multimedia Tools and Applications*, 77(21):28333–28354, Nov 2018.

[18] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren. Make the bag disappear: Carrying status-invariant gait-based human age estimation using parallel generative adversarial networks. In *Proc. of the IEEE 10th Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS 2019)*, pages 1–9, Sep. 2019.

[19] X. Li, S. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. *Trans. on Systems, Man, and Cybernetics, Part C*, 38(2):145–155, Mar. 2008.

[20] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. of the 17th International Conference on Pattern Recognition*, volume 1, pages 211–214, Aug. 2004.

[21] J. Lu and Y.-P. Tan. Gait-based human age estimation. *IEEE Trans. on Information Forensics and Security*, 5(4):761–770, Dec. 2010.

[22] J. Lu and Y.-P. Tan. Ordinary preserving manifold analysis for human age estimation. In *IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics 2010*, pages 1–6, San Francisco, CA, USA, Jun. 2010.

[23] J. Lu and Y.-P. Tan. Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Transactions o Human-Machine Systems*, 43(2):249–258, March 2013.

[24] Y. Makihara, T. Kimura, F. Okura, I. Mitsugami, M. Niwa, C. Aoki, A. Suzuki, D. Muramatsu, and Y. Yagi. Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, June 2016.

[25] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi. Gait-based age estimation using a whole-generation gait database. In *Proc. of the Int. Joint Conf. on Biometrics (IJCB2011)*, pages 1–6, Washington D.C., USA, Oct. 2011.

[26] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of the 9th European Conference on Computer Vision*, pages 151–163, Graz, Austria, May 2006.

[27] H. Mannami, Y. Makihara, and Y. Yagi. Gait analysis of gender and age using a large-scale multi-view gait database. In *Proc. of the 10th Asian Conf. on Computer Vision*, pages 975–986, Queenstown, New Zealand, Nov. 2010.

[28] M. J. Marín-Jiménez, F. M. Castro, N. Guil, F. de la Torre, and R. Medina-Carnicer. Deep multi-task learning for gait-based biometrics. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 106–110, Sep. 2017.

[29] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Int. Series on Biometrics. Springer-Verlag, Dec. 2005.

[30] A. Sakata, Y. Makihara, N. Takemura, D. Muramatsu, and Y. Yagi. Gait-based age estimation using a densenet. In *Proc. of the Int. Workshop on Attention/Intention Understanding (AIU 2018)*, pages 55–63, Dec. 2018.

[31] A. Sakata, N. Takemura, and Y. Yagi. Gait-based age estimation using multi-stage convolutional neural network. *IPSJ*

Transactions on Computer Vision and Applications, 11(1):4, 2019.

[32] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. G. ther, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[33] A. Shehata, Y. Hayashi, Y. Makihara, D. Muramatsu, and Y. Yagi. Does my gait look nice? human perception-based gait relative attributes estimation by dense trajectory analysis. In *Proc. of the 5th Asian Conf. on Pattern Recognition (ACPR 2019)*, pages 1–14, Nov. 2019.

[34] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016)*, number O19, pages 1–8, Halmstad, Sweden, Jun. 2016.

[35] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, Aug. 2004.

[36] S. Stevenage, M. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, 13(6):513–526, Dec. 1999.

[37] N. F. Troje. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2:371–387, 2002.

[38] C. Xu, Y. Makihara, G. Ogi, X. Li, Y. Yagi, and J. Lu. 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, Dec 2017.

[39] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proc. of the 18th Int. Conf. on Pattern Recognition*, volume 4, pages 441–444, Hong Kong, China, Aug. 2006.

[40] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE Trans. on Image Processing*, 18(8):1905–1910, Aug. 2009.

[41] D. Zhang, Y. Wang, and B. Bhanu. Ethnicity classification based on gait using multi-view fusion. In *IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics 2010*, pages 1–6, San Francisco, CA, USA, Jun. 2010.

[42] S. Zhang, Y. Wang, and A. Li. Gait-based age estimation with deep convolutional neural network. In *2019 International Conference on Biometrics (ICB)*, pages 1–8, 2019.

[43] W. Zhao and H. Wang. Strategic decision-making learning from label distributions: An approach for facial age estimation. *Sensors*, 16(7):994, Jun 2016.