Research Article
A Hierarchical Framework for Facial Age Estimation

Yuyu Liang, Xianmei Wang, Li Zhang, and Zhiliang Wang

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Xianmei Wang; plum-wang@tom.com

Age estimation is a complex issue of multiclassification or regression. To address the problems of uneven distribution of age database and ignorance of ordinal information, this paper shows a hierarchic age estimation system, comprising age group and specific age estimation. In our system, two novel classifiers, sequence k-nearest neighbor (SKNN) and ranking-KNN, are introduced to predict age group and value, respectively. Notably, ranking-KNN utilizes the ordinal information between samples in estimation process rather than regards samples as separate individuals. Tested on FG-NET database, our system achieves 4.97 evaluated by MAE (mean absolute error) for age estimation.

1. Introduction

Age is a significant attribute of humans and plays an important role in interpersonal communication. It reveals people's personal conditions and social background, as well as conducting human behaviors. Therefore, age estimation has become a hot research area in the field of computer vision. It has many potential applications in real applications, such as electronic customer relationship management, security control and surveillance monitoring, human-computer interaction, biometrics, and criminal investigation [1, 2].

Over the past decade, many approaches have been proposed for age estimation. Geng et al. [3] defined aging pattern as the sequence of a particular individual's face images sorted in time order, by constructing a representative subspace. Given an unseen face image, they projected it into a subspace that could best reconstruct it with minimum reconstruction error, and the position of the face image in that aging pattern was the age value. Guo et al. [4] proposed a manifold learning method for age estimation. The original face image space was mapped into a low-dimensional subspace by using special subspace learning method. Then, they designed a locally adjusted robust regression algorithm to learn and predict human age. Afterwards, Guo and Wang [5] found that there existed relations between age prediction and expression changes and also built a robust system to conduct cross-expression age estimation. Wu et al. [6] made emphasis

on facial shapes, which were modeled as landmarks on Grassmann manifold. Then, these points were projected onto tangent space, and age estimation was fulfilled by Tangent-Space Regression. Guo et al. [7] established hybrid features, combining features, such as shape feature, texture feature, and frequency feature. They utilized SVM and SVR to predict human age.

However, there are two problems confronted by researchers on age estimation at present. (1) Uneven distribution is common for facial aging databases. (2) Intrinsic correlation of ages is often overlooked, since samples are regarded as independent individuals and ordinal information is lost. The issues both have a great influence on the accuracy of age prediction, whereas they are often neglected. Most researches, including some of the papers listed above, enhance prediction performance mainly by extracting features containing more affluent age information and utilizing more robust and complex classifiers.

Till now, some works have considered the ordinal information of samples. Chao et al. [8] proposed "label-sensitive" concept. Instead of treating each class independently, they considered the samples with similar class labels. The weights of similar samples were also assigned based on the label similarity. In addition, they presented locality preserving projection algorithm to avoid overfitting and explore the connections between facial features and aging labels. Lu and Tan [9] aimed to use ordinary characteristics of age information

to learn discriminative features in low dimensions. They proposed an ordinary preserving manifold analysis method to build low-dimensional subspace, on which samples were projected. Finally, to gain the relation between low-dimensional features and sample values for age estimation, a multiple linear regression model was formed through learning. Li et al. [10] tried to preserve the ordinal information among aging processes, so they employed ordinal discriminative feature learning. To solve uneven distribution of database, a combination of age group and specific age estimation was adopted. Li et al. [11] firstly divided test samples into two groups: birth to adulthood and adulthood to old age. In the next phase, they were detected within certain groups.

In this paper, we introduce a hierarchical system with two layers, age group and age estimation, to tackle uneven distribution problem and use ordinal information. We categorize samples into several age groups and then conduct age estimation within groups to avoid predicting ages on the whole database. Moreover, we employ ranking-KNN to introduce ordinal information under ranking framework.

The rest of this paper is organized as follows. Section 2 describes our proposed system and related algorithms, including feature extracting, age group classification, and age value estimation. Section 3 presents experimental results on FG-NET database [12] with inner and outer comparisons, and the conclusion is showed in Section 4.

2. The Proposed System and Algorithms

Figure 1 depicts the outline of our system, which contains three main parts: sample preprocessing, feature extraction, and estimation. Age information of a facial image is encoded in both texture and shapes. In feature extraction phase, we use appearance features including shape ratio features, wrinkle features, and ULBP (uniform local binary pattern) features [13]. They are combined to describe facial images, acting as the input of following age estimation stages. As for estimation part, it consists of age group estimation and age value estimation. The result of age group estimation is used to guide the next step: age value estimation.

2.1. Sample Preprocessing. In many practical applications, it is difficult to ask users to keep frontal face postures, which will further affect the accurate localization of local facial regions and the computation of facial geometric measurements. To decrease the influence of non-frontal facial posture, we conduct sample rotation in-plane. First, the rotation angle θ in-plane is computed using the location of two eyes. If the angle θ is larger than a specific threshold, rotation correction with θ will be added for the sample. Considering marking deviations of facial landmarks, we set the threshold as 5°. The computation of rotation angle θ in-plane is shown as follows:

$$\theta = \arctan \frac{y_L - y_R}{x_L - x_R}, \tag{1}$$

where (x_L, y_L) and (x_R, y_R) are the locations of the left and right eyes, respectively.

2.2. Feature Extraction

2.2.1. Geometric Feature. Craniofacial changes during aging process are obvious for the young, distinguishing juvenile group from other adult groups effectively [14]. These changes can cause difference between minors and adults in some facial point distances. Wu et al. [6] pointed that "although facial shape could be affected by many factors, such as expression, pose, and age, it still conveys much information about the age of the subject." In the famous FG-NET aging database, each sample has a corresponding file recording sixty-eight facial feature points. We choose eight distances, which can be calculated by fifteen landmarks from these sixty-eight points. However, sample size normally can influence the distance between certain points, introducing instability to shape description. A good solution to eliminate the dependency on the image scale is to apply shape ratios instead of the distances between facial feature points [15]. We calculate eight distances and then form six ratios as geometric feature.

Let $D(A, B)$ be the Euclidean distance between facial landmark A and landmark B, and let W(face cheek) be the width of cheek between left and right ears. Six shape ratios are defined as

$$R_0 = \frac{D\left(\text{Left Eye, Right Eye}\right)}{D\left(\text{Middle of Eyes, Nose Center}\right)},$$

$$R_1 = \frac{D\left(\text{Middle of Eyes, Mouth Center}\right)}{D\left(\text{Middle of Eyes, Chin}\right)},$$

$$R_2 = \frac{W\left(\text{Face Cheek}\right)}{D\left(\text{Middle of Eyes, Chin}\right)},$$

$$R_3 = \frac{W\left(\text{Lower Jaw}\right)}{D\left(\text{Middle of Eyes, Chin}\right)}, \tag{2}$$

$$R_4 = \frac{D\left(\text{Left Eye, Right Eye}\right)}{W\left(\text{Face Cheek}\right)},$$

$$R_5 = \frac{D\left(\text{Middle of Eyebrows, Chin}\right)}{D\left(\text{Left Ear, Right Ear}\right)}.$$

2.2.2. Texture Feature. Good texture descriptors are very important for image appearance. In this paper, we choose ULBP to describe detailed facial local texture. The algorithm of local binary pattern is initially presented by Ojala et al. [13], which is a simple yet very efficient texture operator. It labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number [16]. A further modification is ULBP, introducing bit transitions (from 1 to 0 or vice versa) to reduce the dimension of traditional LBP. Then, an LBP pattern is recognized as a uniform pattern if it contains the bit transitions no more than twice.

Besides ULBP textures, we also extract wrinkle density (WD) [15], a simple and rough way to present wrinkle texture. This simple feature is used in age group estimation step. Let I be a wrinkle region, let E_I be the pixel amount of canny edge
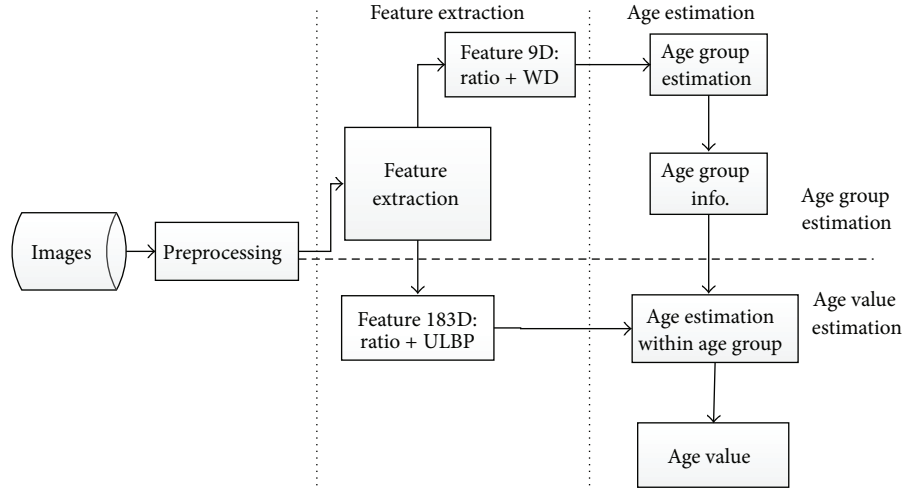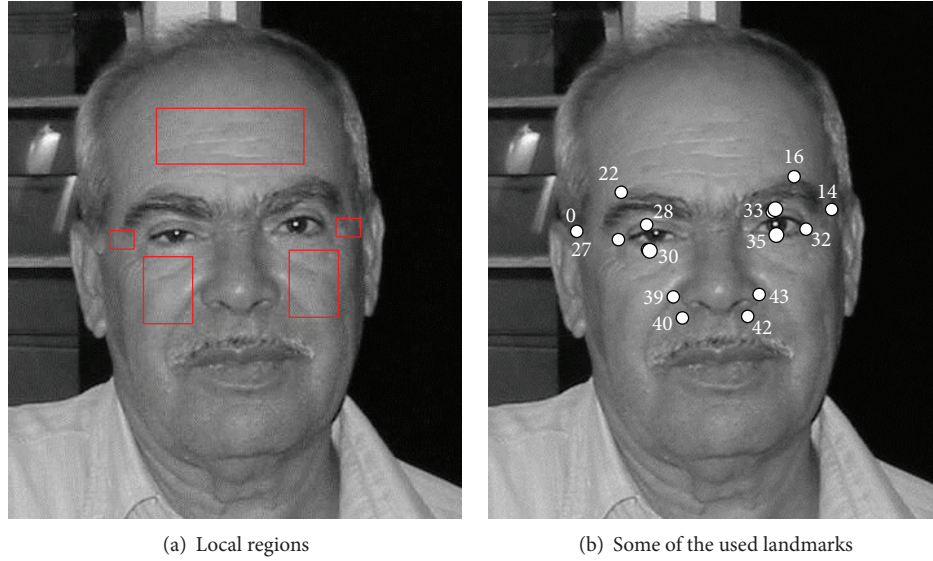
FIGURE 1: Framework of the proposed system.

(a) Local regions

(b) Some of the used landmarks

FIGURE 2: Local regions for texture description.

image of I, and let S_I be the area of I. WD is just the ratio of E_I and S_I listed as follows:

$$\mathrm{WD}_I = \frac{E_I}{S_I}. \tag{3}$$

2.2.3. Combined Feature. Aging is mainly reflected by local wrinkles. When extracting aging features, facial partition is widely used. Considering the influence of moustache, we extract texture features in three local regions: forehead, eye corner, and face cheek. We apply some of 68 facial landmarks shown in Figure 2 to get these regions.

Assume that (x_i, y_i) is the coordinate of the ith facial landmark, $(x_{\mathrm{LT}}, y_{\mathrm{LT}})$ and $(x_{\mathrm{RL}}, y_{\mathrm{RL}})$ are the coordinates of the top-left and low-right points of a rectangle region, respectively, and max() is the function to get the maximum

value of giving data. The detailed localization methods are listed in Table 1.

To reduce the influence of facial deflection in ex-plane, only the larger regions (left or right) of eye corner and face cheek regions are used. So for one image, only three regions are used to extract ULBP texture features and WD features.

In the step of age group estimation, geometric ratios and WD are used. More subtle appearance feature, a combination of ratios and ULBP texture features, is used in the next age value estimation step to get a more accurate estimation within a special age group.

2.3. Age Estimation. Uneven distribution of aging dataset has an important influence on age estimation accuracy [17]. Actually, two issues lie in uneven distribution for aging database. For one thing, some algorithms use the images of

TABLE 1: Localization of local facial regions by facial landmarks.

Local region	x_{LT}	y_{LT}	x_{RL}	y_{RL}
Forehead	$x_{22} + \dfrac{(x_{16} - x_{22})}{8}$	$y_{RL} - \dfrac{(x_{16} - x_{22})}{3}$	$x_{16} - \dfrac{(x_{16} - x_{22})}{8}$	$\max(y_{16}, y_{22}) - \dfrac{(x_{16} - x_{22})}{8}$
Left eye	$x_{RL} - \dfrac{2(x_{27} - x_0)}{3}$	$y_{27} - \dfrac{(y_{30} - y_{28})}{2}$	$x_{27} - 5$	$y_{27} + \dfrac{(y_{30} - y_{28})}{2}$
Right eye	$x_{RL} + \dfrac{2(x_{14} - x_{32})}{3}$	$y_{32} - \dfrac{(y_{35} - y_{33})}{2}$	$x_{32} + 5$	$y_{32} + \dfrac{(y_{35} - y_{33})}{2}$
Left cheek	x_{27}	$y_{30} + \dfrac{(y_{40} - y_{30})}{10}$	x_{39}	$y_{LT} + (y_{40} - y_{30})$
Right cheek	x_{43}	$y_{35} + \dfrac{(y_{42} - y_{35})}{10}$	x_{32}	$y_{LT} + (y_{42} - y_{35})$

TABLE 2: Sample distribution of different age groups on FG-NET.

Age group	0–9	10–19	20–29	30–39	40–49	50–59	60–69
Sample number	372	338	144	79	46	15	8
Scale (%)	37.1	33.7	14.4	7.9	4.6	1.5	0.8

the same person to build a specific aging model. If the samples belonging to this person in the train set are not sufficient to cover a wide age range, reconstruction of the missing ages will be added to assure the integrity. This process may introduce error, because different people may have different aging laws. For another, most samples fall into several age groups, while other age groups comprise much less samples. Unfortunately, sample collection for age estimation is very tough, so effective algorithms are needed to make up this problem. For example, FG-NET is a widely used aging dataset. Sample distribution is displayed in Table 2. Obviously, most of samples fall into 0–29 age groups (85%). Besides, there are no samples for some certain ages.

For age estimation, traditional solutions regard it as a multiclassification or regression problem. However, multiclassification approaches neglect the inherent relationship between labels, assuming that labels are independent. By contrast, regression methods may make use of ordinal information, but, due to difference in personalized aging pattern, kernel functions would be unstable, leading to overfitting problem in learning process easily [18].

In our paper, we treat age estimation as a multiclassification issue and try to utilize ordinal information to construct classifiers. A two-layer estimation system (shown in Figure 3) is employed to conduct age group detection and then age estimation. First, in the age group estimation step, we set all age values into an ordinal category sequence and then map category labels into different age groups by Sequence KNN. For age value estimation step, age value is estimated within a given age group under ranking model with serials of binary KNN classifiers combination.

2.3.1. Sequence KNN.
To reduce the influence of uneven distribution, we bring out an original simple algorithm, SKNN (sequence KNN), to predict age group. Based on the age order of samples, a mapping is built to gain new labels from age value to ordinal category label. The mapping
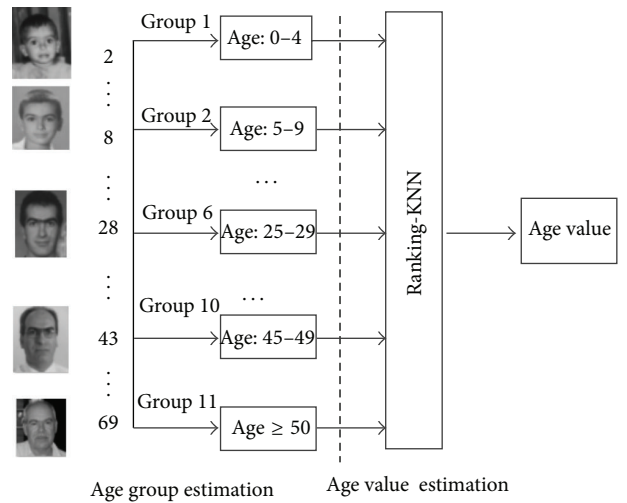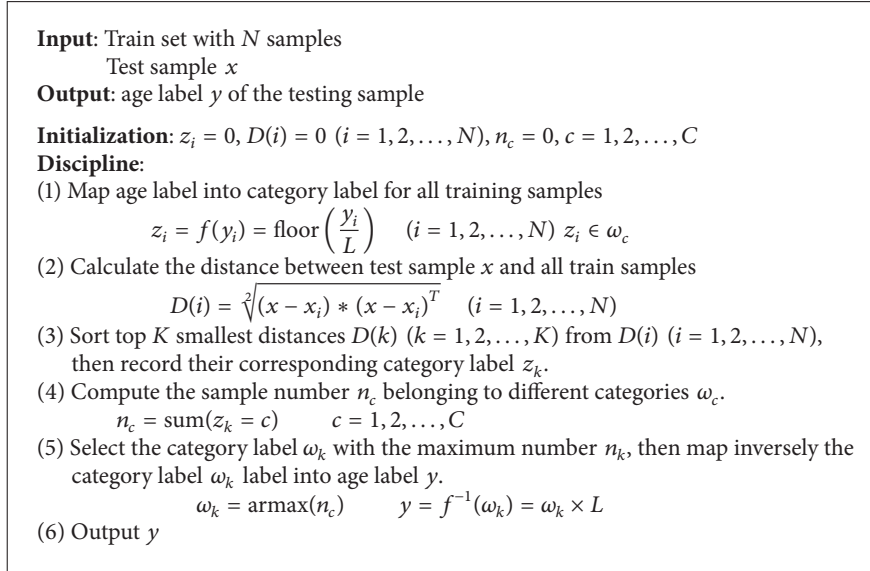

FIGURE 3: Hierarchic framework of age estimation.

makes samples belonging to several age labels classified into one group. By this way, the gap between age labels can be reduced, leading to several better distributed subsets of aging dataset. Then, we can conduct age estimation within certain age groups to avoid the uneven distribution confronted by the whole dataset. The description of SKNN is listed in Algorithm 1.

Given a dataset $\{x_i, y_i\}$ ($i = 1, 2, \ldots, N$), where x_i is a sample, y_i refers to its age label, and N is the number of total samples, we divide all the age labels into C categories ω_c ($c = 1, 2, \ldots, C$); namely, $y_i \in \omega_c$. For each category, the age span is $L = 3$.

After the output of rough age value y, we then map it into a special age group g which is used to get the more precise age value in the following estimation step by ranking-KNN. Suppose there are G age groups divided beforehand; for each

---

Input: Train set with N samples
Test sample x
Output: age label y of the testing sample

Initialization: $z_i = 0, D(i) = 0$ $(i = 1, 2, \ldots, N), n_c = 0, c = 1, 2, \ldots, C$
Discipline:
(1) Map age label into category label for all training samples
$$z_i = f(y_i) = \text{floor}\left(\frac{y_i}{L}\right) \quad (i = 1, 2, \ldots, N)\ z_i \in \omega_c$$
(2) Calculate the distance between test sample x and all train samples
$$D(i) = \sqrt[2]{(x - x_i) * (x - x_i)^T} \quad (i = 1, 2, \ldots, N)$$
(3) Sort top K smallest distances $D(k)$ $(k = 1, 2, \ldots, K)$ from $D(i)$ $(i = 1, 2, \ldots, N)$,
    then record their corresponding category label z_k.
(4) Compute the sample number n_c belonging to different categories ω_c.
$$n_c = \text{sum}(z_k = c) \quad c = 1, 2, \ldots, C$$
(5) Select the category label ω_k with the maximum number n_k, then map inversely the
    category label ω_k label into age label y.
$$\omega_k = \text{armax}(n_c) \qquad y = f^{-1}(\omega_k) = \omega_k \times L$$
(6) Output y

ALGORITHM 1: Algorithm flowchart of SKNN.

group $g(i)$, the boundary age values are $\alpha_{\min}^{(i)}$ and $\alpha_{\max}^{(i)}$; the mapping from an age value to a group is shown as follows:

$$\text{if } \alpha_{\min}^{(k)} \leq y \leq \alpha_{\max}^{(k)} \quad g = g(k), \quad k = 1, 2, \ldots, G. \qquad (4)$$

2.3.2. Ranking-KNN. To utilize the inherent ordinal information between samples, we propose ranking-KNN to predict age value. Under ranking framework, age value estimation is viewed as a series of binary enquiry problems. Each classifier outputs 1 or 0 by checking whether a test sample looks older than the images within a given subset which is formed in age label order. Thus, the ordinal information is used. In the ranking framework, KNN is used as the base classifier.

We consider the nonrepeating age labels as the rank order. Suppose S_c^+ denotes a subdataset with age labels larger than c and S_c^- denotes a subdataset with age labels no more than c. The outline of ranking-KNN is described in Algorithm 2.

3. Experiments and Analysis

We evaluate the effectiveness of the proposed system on the widely used FG-NET aging dataset using leave-one-person-out (LOPO) testing strategy. Two popular evaluation criteria, namely, mean absolute error (MAE) [18, 19] and cumulative score (CS) [1, 2], are adopted. In the following experiments, neighbor number K is set to 30 in age group estimation module and 15 in age value estimation.

3.1. Grouping Evaluation. Group division may influence MAE. Actually, to simplify grouping process, grouping with equal age span is widely used. We also adopt this way to divide all ages into several groups. Considering few ages over 50, to divide ages over 50 into several groups makes no sense. Given group span L_g, we divide samples into $(0 \sim L_g - 1, L_g \sim 2L_g - 1, \ldots, L_g * \text{ceil}(50/l_g) \sim 69)$. Figure 4 presents MAE curves changing with different group divisions.

TABLE 3: Evaluation of the roles of different features.

Feature	WD + ULBP	Ratio + WD	Ratio + ULBP	ULBP	All
MAE	10.37	7.28	9.45	10.83	4.97

From Figure 4, we can see that when group span L_g is more than 10, MAE rises quickly over 6.0. For the group span from 2 to 7, MAE changes slightly. The best performance can be achieved by $L_g = 5$. The reason may be that if group span is too large, more samples with very different age labels will drop into the same group, which will further affect the estimation accuracy of next step.

3.2. Feature Evaluation. In order to test the role of different features used in this paper, we exclude geometric ratio features, wrinkle density features, and ULBP texture features separately in feature extraction stage. Meanwhile, age estimation part is still a hierarchical structure. Thus, we get different fusions of features: WD plus ULBP, ratio plus ULBP, and ratio plus WD. Then, we evaluate MAE under the hierarchical framework with group span $L_g = 5$. Table 3 gives the evaluation results.

From Table 3, it is obvious that different features play different effects, because they depict a face from different views. Exclusion of ratio features from the feature combination will lead to a dramatic rise of MAE. The reason may be that there are a lot of young samples in the database. This is also one of the reasons that the combination of ratio and WD can outperform the combination of ratio and ULBP. ULBP is more suitable to describe adult. Only with the combination of ratio, WD, and ULBP texture under our hierarchical framework can the best system performance be achieved.

```
Input: Training set S with N samples
       Testing sample x
       Group g with boundary α_min and α_max, output by age group estimation
Output: Estimated age of the testing sample

Initialization: S_c^- = ∅, S_c^+ = ∅, z_out = 0, b = a_max − a_min x, l(c) = 0
Discipline:
(1) For c = 1 : b
    (1.1) For i = 1 : N
            If  y(i) ≤ c + a_min − 1        x_i ∈ S_c^-
            Else    x_i ∈ S_c^+
          End for i
    (1.2) Call KNN to decide x belonging to subset S_c^- or S_c^+, using the samples within group g.
            If x ∈ S_c^-, l(c) = 0 else l(c) = 1
    End for c
(2) Deduce the estimated age of testing sample based on l(c):
       For c = 1 : b
       z_out = z_out + l(c)
       End for c
       y = z_out + α_min
(3) Output the estimated age y
```
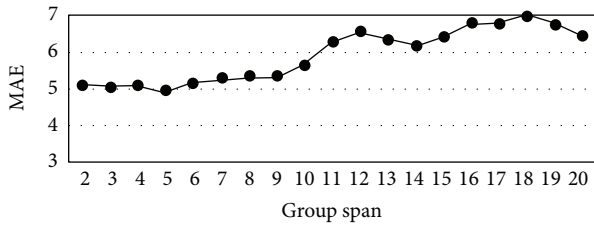
ALGORITHM 2: Algorithm flowchart of ranking-KNN.

FIGURE 4: MAE curves of different group divisions.

TABLE 4: MAE of systems with different components.

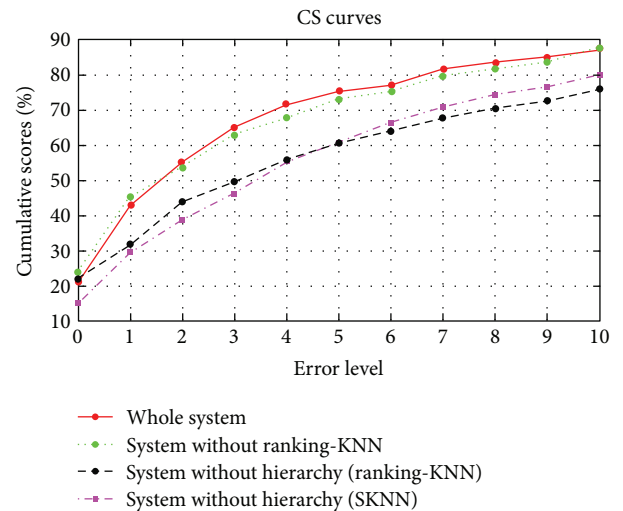| Method	MAE
Our system	4.97
System without ranking-KNN (replaced by KNN)	5.76
System without hierarchy framework (ranking-KNN)	7.65
System without hierarchy framework (SKNN)	8.58

FIGURE 5: Cumulative scores of our system with or without certain component.

3.3. Contribution Evaluation of System Components. Our hierarchy system consists of age group estimation (SKNN) and age value estimation (ranking-KNN). Experiments also are carried out to evaluate their roles. The detailed tests include the following four experiments.

Whole system (our system): system performance is tested by using the overall hierarchy system (shown in Figure 3), consisting both age group and age value.

System without ranking-KNN (replaced by KNN): similar to the whole system, the system is also hierarchical, but in the age value estimation stage ranking-KNN is replaced by traditional KNN.

System without hierarchy system (ranking-KNN): age group estimation component is excluded, so the system just uses ranking-KNN to estimate age. Obviously, this is a nonhierarchical structure.

System without hierarchy system (SKNN): ranking-KNN estimation component is excluded, so the nonhierarchical system just includes SKNN, namely, age group estimation part.

The results are shown in Table 4 and Figure 5. Actually age estimation system with an error more than 10 years is not acceptable, so we just test the CS with error level just within 10 years old.

As is showed in Table 4 and Figure 5, the experiment of "system without ranking-KNN" shows an increase in MAE

Table 5: Results on different age ranges.

Age range	Distribution (%)	MAE
0–9	37.1	1.59
10–19	33.7	2.48
20–29	14.4	2.50
30–39	7.9	3.68
40–69	6.9	17.78

Table 6: MAE comparisons with some of the existing algorithms.

Algorithm	MAE
KNN [20]	8.24
KNN [3]	8.83
Ranking-SVM [18]	6.02
Rank [21]	5.79
Ranking [22]	5.33
AGES [3]	6.52
LARR [4]	5.07
IIS-LLD [23]	5.70
M^2SA [24]	5.36
Our method	4.97

and a slight decrease in CS curves, because we use traditional KNN to replace ranking-KNN. It proves that ranking-KNN performs better than KNN for age estimation. For the system without hierarchy framework (ranking-KNN), MAE drops to 7.65 from 4.97, which proves that age group estimation plays an important role in age value estimation. The reason may be that there exist some "confused" samples with similar aging features but very different age. Limiting age value estimation within a specific group, instead of the whole database, can reduce the influence of these "confused" samples. For the system just with SKNN to get the age value, system performance also drops quickly. Clearly, the two experiments of "system without hierarchy framework" gain apparent MAE decrease and low CS. It well testifies that hierarchic framework can boost estimation accuracy greatly.

3.4. Performance Evaluation of Overall System. Finally, we measure the performance of our proposed system on different age ranges and compare it with some other published systems.

Table 5 reports the results on different age ranges of our proposed system.

The MAE of the whole database is 4.97. From Table 5, it is clear that the sample volume of an age group greatly affects the final prediction result. With the decrease of the sample number within a group subset, MAE increases accordingly. This also can testify that uneven distribution is a key factor that affects the final estimation accuracy. The collection of aging database with wide span and enough higher ages is a very tough and urgent issue in the future.

The performance comparisons between several recently published systems and ours are shown in Table 6. Especially, the four systems listed in the left column are partly similar to ours, either with KNN or ranking model, using LOPO test rule. Table 6 shows that our system performs better than some of the published estimation systems.

4. Conclusions

This paper brings up a two-layer system for age estimation, combining age grouping and age value calculation. First, we use sequence KNN algorithms to classify an unknown sample into a special age group by fusing geometric ration and wrinkle density features. Then, appearance features binding geometric ration and ULBP textures are employed to estimate the age value under ranking-KNN framework. Especially, by translating age label into category label at age group estimation stage, the whole age span is divided into some ordinal subspans, which can help to reduce somewhat the

influence of uneven distribution problem confronted by current face datasets. For ranking model, it takes the age evaluation task as the combination of many ordinal binary classifiers which can utilize the ordinal information hidden among the aging processes.

In the future, we will put more effort on sample preprocessing, considering the great influence from facial occlusion, expression, and pose. At the same time, we will try to refine our age group estimation algorithm, since its result has a crucial effect on the whole system's performance.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[2] X. Wang, L. Liang, Z. Wang, and S. Hu, "Age estimation by facial images: a survey," *China Journal of Image and Graphics*, vol. 17, no. 6, pp. 603–622, 2012.

[3] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.

[4] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

[5] G. Guo and X. Wang, "A study on human age estimation under facial expression changes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2547–2553, 2012.

[6] T. Wu, P. Turaga, and R. Chellappa, "Age estimation and face verification across aging using landmarks," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1780–1788, 2012.

[7] J.-M. Guo, Y.-M. Liou, and H.-S. Nguyen, "Human face age estimation with adaptive hybrid features," in *Proceedings of the International Conference on System Science and Engineering (ICSSE '11)*, pp. 55–58, Macau, China, June 2011.

[8] W. Chao, J. Liu, and J. Ding, "Facial age estimation based on label-sensitive learning and age-specific local regression," in *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1941–1944, 2012.

[9] J. Lu and Y.-P. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 2, pp. 249–258, 2013.

[10] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2570–2577, Providence, RI, USA, June 2012.

[11] W. Li, Y. Wang, and Z. Zhang, "A hierarchical framework for image-based human age estimation by weighted and OHRanked Sparse Representation-based classification," in *Proceedings of the 5th IAPR International Conference on Biometrics (ICB '12)*, pp. 19–25, New Delhi, India, April 2012.

[12] "The FG-NET Aging Database," http://www.fgnet.rsunit.com/.

[13] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 782–786, 2002.

[14] Y. H. Kwon and L. N. da Vitoria, "Age classification from facial images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 762–767, June 1994.

[15] M. M. Dehshibi and A. Bastanfard, "A new algorithm for age recognition from facial images," *Signal Processing*, vol. 90, no. 8, pp. 2431–2444, 2010.

[16] M. Pietikainen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision with Local Binary Patterns*, Springer, 2011.

[17] L. Yingdong and N. Huimin, "KNN classification algorithm based on k-nearest neighbor graph for small sample," *Computer Engineering*, vol. 37, no. 9, pp. 198–200, 2011.

[18] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "A ranking approach for human age estimation based on face images," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 3396–3399, Istanbul, Turkey, August 2010.

[19] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.

[20] E. A. Patrick and F. P. Fischer III, "A generalized k-nearest neighbor rule," *Information and Control*, vol. 16, no. 2, pp. 128–152, 1970.

[21] Y. Ma, T. Xiong, Y. Zou, and K. Wang, "Person-specific age estimation under ranking framework," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR '11)*, Trento, Italy, April 2011.

[22] Y. Shuicheng, W. Huan, T. S. Huang, Y. Qiong, and T. Xiaoou, "Ranking with uncertain labels," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '07)*, pp. 96–99, Beijing, China, July 2007.

[23] X. Geng, K. Smith-Miles, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," in *Proceedings of the 24th AAAI Conference on Artificial Intelligence and the 22nd Innovative Applications of Artificial Intelligence Conference (AAAI '10/IAAI '10)*, pp. 451–456, Atlanta, Ga, USA, July 2010.

[24] X. Geng, K. Smith-Miles, Z.-H. Zhou, and L. Wang, "Face image modeling by multilinear subspace analysis with missing values," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, vol. 41, no. 3, pp. 881–892, 2010.