Age Estimation via Grouping and Decision Fusion

Kuan-Hsien Liu, *Member, IEEE*, Shuicheng Yan, *Senior Member, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*— We present a novel multistage learning system, called grouping estimation fusion (GEF), for human age estimation via facial images. The GEF consists of three stages: 1) age grouping; 2) age estimation within age groups; and 3) decision fusion for final age estimation. In the first stage, faces are classified into different groups, where each group has a different age range. In the second stage, three methods are adopted to extract global features from the whole face and local features from facial components (e.g., eyes, nose, and mouth). Each global or local feature is individually utilized for age estimation in each group. Thus, several decisions (i.e., estimation results) are derived. In the third stage, we obtain the final estimated age by fusing the diverse decisions from the second stage. To create diverse decisions for fusion, we investigate multiple age grouping systems in the first stage, where each system has a different number of groups and different age ranges. Thus, various decisions can be made from the second stage, and will be delivered to the third stage for fusion. Totally, six fusion schemes (i.e., intra-system fusion, inter-system fusion, intra-inter fusion, inter-intra fusion, maximum-diversity fusion, and composite fusion) are developed and compared. The performance of the GEF system is evaluated on the Face and Gesture Recognition Research Network and the MORPH-II databases, and it outperforms the existing state-of-the-art age estimation methods by a significant margin. That is, the mean absolute errors of age estimation are reduced from 4.48 to 2.81 years on FG-NET and 3.82 to 2.97 years on MORPH-II.**

***Index Terms*— Age estimation, age group classification, decision fusion, feature extraction, feature selection.**

I. INTRODUCTION

IN THE past few years, human facial age estimation has drawn a lot of attention in the computer vision community because of its important applications in age-based image retrieval [1], internet access control, security control and surveillance [2], [3], biometrics [2], [4], [5], human-computer interaction (HCI) [6], [7], and electronic customer relationship management (ECRM) [2].

Estimating human age from a facial image requires a great amount of information from the input image. This kind of information is often called facial aging features. Extraction of these features is important since the performance of an age estimation system will heavily rely on the quality of extracted features [2]. Lots of research on age estimation has been conducted towards aging feature extraction. Examples include: the active appearance model (AAM) [8], age manifold [9], [10], AGing pattern Subspace (AGES) [6], [11], anthropometric model [12], biologically inspired features (BIF) [13], and patch-based appearance model [14], [15].

Another aspect for age estimation is to build a reliable age prediction system (i.e., age estimator) based on extracted features. The age estimator can use a machine learning approach to train a model for extracted features and make age prediction for query faces with the trained model. Generally speaking, age estimation can be viewed as a multi-class classification problem [1], [11], [16], [17], a regression problem [9], [10], [13], [15], [18]–[22], or a composite of these two [16], [23], [24]. From a different perspective, facial aging can also be treated as an ordinal process. For instance, the face of a 2-year-old child should be more closely related to the face of a 3-year-old child than the face of a 15-year-old teenager. Thus, age estimation can also be treated as a ranking problem [25]–[27].

Although many approaches have been presented to deal with age estimation, most of them directly estimate an age from a very wide age range. However, it would be more meaningful to estimate an age from a narrower age range. Some hierarchical methods [28], [29] have shown a good performance for age estimation. For example, estimating an age in the age range of 15 to 20 is easier than estimating an age in the age range of 0 to 60. The task of age grouping (or age group classification) is to classify facial images into different age groups. With higher accuracy in age grouping, the age estimation error in each age group is expected to be lower. Being motivated by the above observation, we present a novel age estimation framework, called Grouping-Estimation-Fusion (GEF) system, which consists of three main stages: 1) age grouping, 2) age estimation within age groups, and 3) fusion of decisions.

The main contributions of this work are: 1) diverse decisions (i.e., different age estimation results) are generated by creating multiple age grouping systems; 2) we extensively demonstrate that the age estimation accuracy is closely dependent on the age grouping; 3) a systematic way of measuring the diversity between decisions for intra-system and inter-system is proposed; and 4) six decision fusion schemes are presented to perform age estimation. The performance of our proposed solution is evaluated on the FG-NET [30] and the

Manuscript received December 5, 2014; revised May 3, 2015 and July 1, 2015; accepted July 6, 2015. Date of publication July 30, 2015; date of current version September 15, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alex C. Kot.

K.-H. Liu and C.-C. J. Kuo are with the Ming Hsieh Department of Electrical Engineering, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA (e-mail: liuk@usc.edu; cckuo@sipi.usc.edu).

S. Yan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: eleyans@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2015.2462732

MORPH-II [31] databases, and it outperforms existing state-of-the-art age estimation methods by a significant margin. That is, the mean absolute errors (MAEs) of age estimation can be reduced from 4.48 to 2.81 years on FG-NET and 3.82 to 2.97 years on MORPH-II.

Fusion is a widely used technology in biometrics, and it is the most common to fuse features from data of multi-modalities (called the data fusion approach). It is also common to fuse multiple decisions to get a more robust one. One famous example is the ensemble learner [32]. We follow the *ensemble learner* methodology in this work. Its main novelty lies in the development of different age grouping schemes so as to allow a diversified set of decisions for decision fusion. Each grouping scheme leads to a weak learner (i.e., an age regressor). The fusion of estimated ages from multiple age regressors results in a more accurate result. The performance improvement is also significant compared with other state-of-the-art methods. The noticeable performance improvement is attributed to ensemble learning. To the best of our knowledge, this is the first work applying ensemble learning to the solution of the age estimation problem. Although the high level concept of ensemble learning is not new, we feel that the realization of the high level concept and a clear demonstration of its superior performance in the context of a certain real world problem are still valuable contributions.

The rest of this paper is organized as follows. Related previous work is briefly reviewed in Section II. An overview of the proposed GEF age estimation scheme is presented in Section III. The age grouping method is introduced in Section IV. Age estimation within each age group is detailed in Section V. Analysis of diversity, designs of fusion schemes, and decision selection algorithms used in our fusion schemes are discussed in Section VI. Experimental results are shown in Section VII. Finally, concluding remarks and future work are given in Section VIII.

II. REVIEW OF PREVIOUS WORK

In recent years, facial age grouping and facial age estimation problems have been extensively studied. A lot of approaches have been proposed for these two research topics. We briefly review age grouping in Section II.A and age estimation in Section II.B.

A. Age Grouping

Age grouping (i.e., age group classification) was first conducted by Kwon and da Vitoria Lobo in [12]. They categorized facial images into three age groups: babies, young adults, and senior adults. They computed six ratios of distances between primary components (e.g., eyes, noses, mouth, etc.) and separated babies from the other two groups. Then, wrinkles on specific areas of a face were located using snakes, and wrinkle indices were used to distinguish senior adults from young adults and babies. There were only 47 images in the experimental dataset, and the correct classification rate for the baby group was below 68%.

Horng *et al.* [33] proposed a system that classifies faces with three steps: primary components detection, feature extraction, and age classification. They classified 230 facial images into four age groups: babies, young, middle-aged and senior adults. They first used the Sobel edge operator [34] and region labeling to locate the positions of eyes, noses, and mouths. Then, two geometric features and three wrinkle features were extracted. Finally, two back-propagation neural networks were constructed for classification. The correct classification rate was 81.58%. The facial age groups were subjectively assigned (i.e., not actual ages) in their experiments.

Thukral *et al.* [35] extracted geometric features from faces and fused the results from five classifiers: ν-SVC [36], partial least squares (PLS) [37], Fisher linear discriminant (FLD), Naïve Bayes, and k-nearest neighbor (KNN) [34], by adopting the majority decision rule. The final rate was 70.04% for three age groups (namely, 0-15, 15-30, and 30+).

Gunay and Nabiyev [38] proposed an automatic age classification system based on local binary patterns (LBP) [39] for face description. Faces were divided into small regions from which the LBP histograms were extracted and concatenated into a feature vector. For every new face presented to the system, spatial LBP histograms were produced and used to classify the image into one of six age groups: 10±5, 20±5, 30±5, 40±5, 50±5, 60±5. The minimum distance, the nearest neighbor and the k-nearest neighbor classifiers were used. Their system gave a classification rate of 80%.

Hajizadeh and Ebrahimnezhad [40] used histograms of oriented gradients (HOG) [41] as the facial feature. HOG features were computed in several regions and these regional features were concatenated to construct a feature vector for each face. A probabilistic neural network (PNN) classifier was used to classify facial images into one of four age groups. The classification rate was 87.25%.

Liu *et al.* [42] proposed a structured fusion method for age group classification by building a region of certainty (ROC) to connect the uncertainty-driven shape features with selected surface features. In the first stage, two shape features are designed to determine the certainty of a face and classify it. In the second stage, the gradient orientation pyramid (GOP) [43] features are selected by a statistical method and then combined with an SVM classifier to perform age grouping. Their method was tested in classifying faces into three age groups, and the classification accuracy of 95.1% was reported.

Sai *et al.* [44] considered age grouping with four pre-defined age groups. The Local Gabor Binary Pattern (LGBP) [45], BIF and Gabor features were first extracted from face images and, then, a machine learning method, called the Extreme Learning Machine (ELM) [46], was adopted for age grouping. They conducted experiments on three aging datasets (called datasets I, II and III) to demonstrate the effectiveness and robustness of their proposed method, where each tested dataset contains only a fraction of the datasets. For example, dataset III was formed by selecting 1,000 images from MORPH-II. The reported accuracy was around 70%.

B. Age Estimation

Lanitis *et al.* [1] used the active appearance models (AAMs) by combining shape and appearance facial features. Age estimation was treated as a classification problem and solved by the shortest distance classifier and neural networks. They differentiated age-specific and appearance-specific estimation problems. Personalized age estimation was introduced to cluster similar faces before classification.

Geng *et al.* [6], [11] proposed an automatic age estimation method named AGES (AGing pattErn Subspace), which modeled the long-term aging process of a person (i.e., a sequence of a person's face images), and estimated the person's age by minimizing the reconstruction error. However, the facial features of the same person could be similar in different ages.

Guo *et al.* [13] extracted BIF for each face, applied the principal component analysis (PCA) [47] for feature dimensionality reduction. They used classification and regression approaches to age estimation.

Yan *et al.* [15] proposed a patch-based regression method for age estimation, where the regression error was minimized by a three-complementary-stage procedure. First, each image was encoded as an ensemble of orderless coordinate patches of GMM (Gaussian Mixture Model) distribution. Then, the patch-kernel was designed for characterizing the Kullback-Leibler divergence between the derived models for any two images, and its discriminating power was further enhanced by a weak learning process, called inter-modality similarity synchronization. Finally, kernel regression was employed for ultimate human age estimation.

Zhang and Yeung [21] proposed a multi-task warped Gaussian process (MTWGP) model for age estimation. Age estimation was formulated as a multi-task regression problem where each learning task was to estimate the age function for each person. Besides modeling common features shared by different tasks (persons), MTWGP also allowed task-specific (person-specific) features to be learned automatically.

Chang *et al.* [25] proposed an ordinal hyperplane ranking algorithm (OHRank) using the relative order information among age labels in a database. Each ordinal hyperplane separated all facial images into two groups by the relative order, and a cost-sensitive property was used to find a better hyperplane by minimizing the classification cost. Human age was then inferred by aggregating a set of preferences from multiple ordinal hyperplanes.

Guo and Mu [18] used the kernel partial least squares (KPLS) regression for age estimation. It has three advantages: 1) the KPLS can reduce feature dimensionality and learn the aging function simultaneously in a single learning framework; 2) the KPLS can find a small number of latent variables (e.g., 20) to project thousands of features into a low-dimensional subspace, which is attractive in real-time applications; and 3) the KPLS has an output vector consisting of multiple labels to solve several related problems (e.g., age estimation, gender classification, and ethnicity estimation) together.

Li *et al.* [48] considered temporally ordinal and continuous characteristics of the aging process and proposed to learn ordinal discriminative facial features. Their method aimed at preserving the local manifold structure of facial images while keeping the ordinal information among aging faces. The two factors were formulated as a unified optimization problem, and a solution was presented.

Several prior age estimation work exploited the idea of "grouping followed by estimation". For example, Guo *et al.* [49] used the biologically inspired features (BIF) with manifold learning for face representation. The gender (female, male) and age groups (young, adult, senior) were classified jointly to result in 6 groups (f-y, f-a, f-s, m-y, m-a, m-s) and age estimation was conducted within each of 6 groups. Guo *et al.* [16] adopted a locally adjusted robust regressor to find the range of ages and, then, used classification to determine the age within a range. Choi *et al.* [29] used fused global and local features to perform classification within 4 age groups, where each group was overlapped with its adjacent groups, and age estimation was conducted within each age group. Dibeklioğlu *et al.* [28] applied dimensionality reduction to appearance features to get combined features, and used them to classify faces into 7 age groups. Then, the age was predicted in each group by trained regressors.

Panis *et al.* [50] presented a survey on facial aging research in the past decade. It discussed the main methodologies, a list of benchmark results and future research trends and requirements. Chang and Chen [51] proposed a cost-sensitive ordinal hyperplanes ranking algorithm for age estimation. They adopted the scattering transform to extract facial features, obtained the age rank by aggregating a series of binary classification results, and conducted an analysis on the cost of each individual binary classifier. Dibeklioglu *et al.* [52] proposed to combine facial dynamics and appearance for age estimation. The dynamic features were derived from facial expressions and fused with the appearance features to train the Support Vector Machine (SVM) classifiers/regressors. They analyzed the discrimination power of smile dynamics for age estimation and showed that smile dynamics can improve the estimation accuracy.

The aforementioned results are categorized in Table I. The difference between these methods and ours is explained below. We first establish multiple age grouping systems (not just one system), where the number of age groups and age ranges can vary. For each grouped system, a trained classifier is used to predict age groups and a regressor is trained to estimate age within each group. This strategy overcomes errors introduced in group's boundary regions and offers the diversity gain. The estimated ages for a face from different systems are fused to get a final decision to exploit this diversity gain. Consequently, our framework is a three-level hierarchical age estimation system, which includes grouping, estimation and fusion. These three levels are cascaded into one full GEF system. In contrast, all other age estimation systems are either one-level (estimation only) or two-level (grouping followed by estimation).

III. Overview of Proposed GEF System

The proposed GEF age estimation scheme consists of three stages. In the first stage, we adopt the age grouping method in [42] to classify face images into different age groups. The entire age range is divided into several non-overlapping ranges and each age group has a different range. Then, the gradient orientation pyramid (GOP) [43] is adopted to represent overall

TABLE I

CATEGORIZATION OF SOME AGE GROUPING AND AGE ESTIMATION METHODS, *HYBRID=SHAPE+SURFACE, **HIERARCHICAL=GROUPING+ESTIMATION

Method		Architecture			Strength for feature
		Age grouping	Age estimation	Hierarchical**	
Feature type	Shape	Thukral 2012 [35]	Geng 2007 [11]	—	More suitable for younger people
	Surface	Gunay 2008 [38] Hajizadeh 2011 [40] Sai 2015 [44]	Guo 2009 [13] Yan 2008 [15] Zhang 2010 [21] Guo 2011 [18] Li 2012 [48] Chang 2015 [51]	Guo 2008 [16] Guo 2009 [49]	More suitable for older people
	Hybrid*	Kwon 1999 [12] Horng 2001 [33] Liu 2014 [42]	Lanitis 2004 [1] Chang 2011 [25] Dibeklioğlu 2015 [52]	Dibeklioğlu 2012 [28] Choi 2011 [29]	Suitable for all ages
Capacity for architecture		Get coarse age ranges	Get ages from entire age range	Get coarse age ranges and exact ages from narrower age ranges	—

facial features. To further increase the discriminating ability of the feature space, the analysis of variance (ANOVA) [53] is employed to select the more discriminative features from the GOP feature vector, which significantly reduces the dimensionality of the GOP feature vector. Then, the linear support vector machine (SVM) [54] is adopted to learn a model and classify faces into age groups.

In the second stage, an exact age for each face is estimated within its group range. Here, both local and global features are used. Local features are obtained by extracting features from local facial areas. A cascaded object detector using the Viola-Jones [55] algorithm is adopted to detect three facial components (eyes, nose and mouth). To compare with other benchmarking methods fairly, we adopt three features commonly used by others for age estimation. They are biologically inspired features (BIF) [13], histograms of oriented gradients (HOG) [41] and local binary pattern (LBP) [39]. The global features are obtained by extracting BIF, HOG, and LBP from the whole face. Every global or local feature (e.g., BIF_eyes or LBP_mouth) is used by the support vector regression (SVR) [54] to predict ages for faces in each group. At the end of this stage, decisions (i.e., estimation results) from the system outputs are produced and they are used as the input features to the third stage.

In the third stage, we fuse decisions obtained from the 2nd stage. To obtain a powerful fusion scheme, it is desired to have rich diversity among decisions. To achieve this goal, multiple age grouping systems are constructed, where each system has a different number of age groups and different age ranges. For example, if the entire age range is from 0 to 70, one system could have 3 age groups: 0-10, 11-30, 31-70 while another may have 5 groups: 0-10, 11-20, 21-30, 31-50, 51-70. With the analysis of diversity in decisions, six efficient fusion schemes are proposed and compared to yield the final age estimation result.

IV. AGE GROUPING

The objective of age grouping (i.e., age group classification) is to classify face images into different groups based on their ages. The entire age range is divided into several non-overlapping ranges, and each range constitutes an age group. As discussed in [42], when the number of groups is small (e.g., 2 or 3), both shape (geometric) and surface (texture) features can be utilized for age grouping. However, when the number of groups is larger (e.g., 4 or above), the shape features do not help much in enhancing classification accuracy. Higher accuracy in age grouping contributes to age estimation in the 2nd stage. However, since existing features such as BIF, HOG, and LBP, do not perform well for age grouping in our experiments as shown in Sec. VII, we follow a recently developed method [42] for age grouping, which is detailed below.

A. Feature Extraction

The gradient orientation (GO) was shown to be robust to illumination change and successfully applied to many areas, such as disparity estimation [56], visual quality assessment [57] and face recognition tasks. The facial aging features were computed based on the gradient orientation pyramid (GOP), which can provide the image gradient information as well as the pyramid information. For a given image, we first build a pyramid of this image and, then, compute the gradients in each layer of the pyramid. Finally, these gradients are combined together as a GOP feature vector.

B. Feature Selection

It is desired to select features with higher discrimination and discard features with lower discrimination. Here, we use a statistics-based method to achieve this goal. Specifically, the analysis of variance (ANOVA) method is used to measure which feature has a higher discriminating power among age groups. The detailed procedure is given in [42]. Simply speaking, we first find the F value for each feature in the GOP feature vector. Next, we find an F_{crit} for each age grouping system becuase F_{crit} depends on the number of groups. Finally, a feature is selected if its F is larger than F_{crit}.

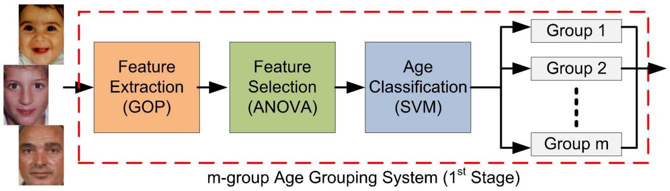


Fig. 1. The age grouping system.

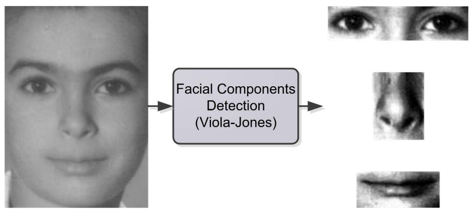


Fig. 2. An example of facial components detection.

In our experiments, only 10% to 20% of features are selected from a GOP feature vector. Thus, feature dimensionality is significantly reduced by the ANOVA method.

C. Age Classification

The support vector machine (SVM) is a widely used machine learning tool for classification, regression, and other learning tasks. In the age grouping stage, we use the multi-class SVM with a linear kernel to train a model and classify faces into different age groups. The entire age grouping algorithm, including feature extraction, feature selection, and age classification, is summarized in Figure 1.

V. Age Estimation Within Age Groups

After the age grouping process, each face is classified into an age group, which has a defined range. The exact age for each classified face is estimated within the defined age range. It includes three steps: (i) facial components detection, (ii) feature extraction from facial components, and (iii) age estimator learning.

A. Facial Components Detection

In addition to the global facial information, we explore the local facial information by detecting facial components and extracting several features from the detected facial components. In image processing, the Viola-Jones [55] algorithm is one of the most efficient and widely used algorithms in object detection, and it demonstrates exceptional competence in face detection. Since the eyes, nose, and mouth are important parts of a face, we intend to extract local aging features from them. In this step, a cascaded object detector using the Viola-Jones [55] algorithm is adopted to detect these three important facial components. To increase detection accuracy, we crop one full face image into its upper half, middle half, and lower half regions for eyes, nose, and mouth detection, respectively. Figure 2 gives an example of detected facial components. The accuracy of the component detectors is given in Table II. One advantage of using these detected facial components is the lower feature dimensionality, since the image sizes of detected facial components are much smaller than that of a whole face. Another advantage is richer diversity due to more features are extracted from multiple representative areas.

TABLE II
ACCURACY OF FACIAL COMPONENTS DETECTION

Database	Eyes	Nose	Mouth
FG-NET	96.1 %	92.6 %	94.3 %
MORPH-II	97.1 %	93.2 %	95.5 %

TABLE III
THE 12 FEATURES USED IN THE 2ND STAGE

Feature	Type
f_1: BIF	Global
f_2: HOG	Global
f_3: LBPu	Global
f_4: BIF_eyes	Local
f_5: BIF_nose	Local
f_6: BIF_mouth	Local
f_7: HOG_eyes	Local
f_8: HOG_nose	Local
f_9: HOG_mouth	Local
f_{10}: LBPu_eyes	Local
f_{11}: LBPu_nose	Local
f_{12}: LBPu_mouth	Local

B. Global and Local Feature Extraction

Recently, the biologically inspired features (BIF) [13], histograms of oriented gradients (HOG) [41], and local binary pattern (LBP) [39] methods are widely used to extract facial aging information. To compare with other benchmarking methods fairly, we adopt those features commonly used by other methods in age estimation. That is, we extract both global and local BIF, HOG, and LBP features. For LBP, we use a modified one, called uniform LBP and denoted by LBPu, which has a much lower dimension than LBP.

To obtain global features, we extract BIF, HOG, and LBPu features from the whole face and denote them with f_1, f_2 and f_3, respectively. Furthermore, we extract BIF, HOG, and LBPu features from 3 facial components to get 9 local features and denote them with f_4-f_{12}. The 12 features are listed in Table III. The same block size (i.e., 32×32 pixels) is used to extract these 12 features, and the window sizes for BIF, HOG, and LBP features are 8×8, 8×8, and 3×3 pixels, respectively.

C. Age Estimator Learning

Each of the 12 features is used to obtain an age estimator (AE). Being similar to the first stage, the SVM method is adopted to train all age estimators (AEs) in the second stage. After the age grouping step, each age group has its own model trained by the SVR with a linear kernel for age prediction. We tried a nonlinear kernel known as the radial basis function (RBF), which is often chosen when the feature dimension is low [58], [59], in the experiment, yet obtained almost the same performance. Thus, the linear kernel is chosen in this stage for its lower complexity. When an input face is

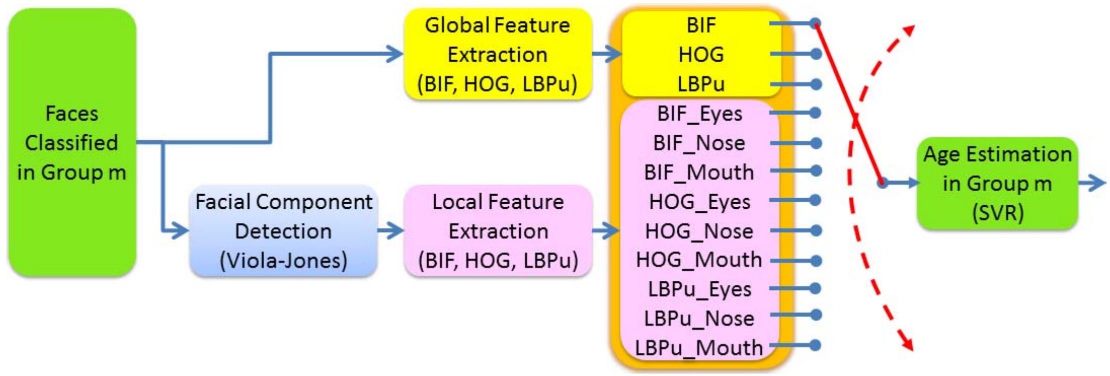


Fig. 3. Age estimation within age groups.

classified into an age group with age interval [a, b], where $a < b$, the SVR of this group will be used to predict its age under the interval constraint. That is, the predicted age will be lower bounded by a and upper bounded by b if the prediction goes outside of the interval. The procedure for using SVR to implement an age estimator is described as follows.

1) *Classified Results:* For an age grouping system with m groups, each face is classified to a group with label i, $i \in \{1, 2, \ldots, m\}$.
2) *Feature Representation:* The feature vector of each face consists of three parameters: group i, classified group i_c, and age a.
3) *Scaling:* Every feature in a feature vector is linearly scaled to range [0, 1] among all faces. This is conducted to avoid the dominance of attributes with a large dynamic range over those with a smaller dynamic range. The linear scaling operation is performed for both training and testing data via
$$x = \frac{r - \min(R)}{\max(R) - \min(R)}, \tag{1}$$
where x is the scaled feature, r is the raw feature, and $\max(R)$ and $\min(R)$ specify the maximum and minimum values of the feature range R, respectively.
4) *Cross Validation:* The leave-one-person-out (LOPO) cross validation technique is used on the FG-NET database. For the MORPH-II database, the same setting as that in [18] and [60] is adopted.
5) *Training:* The training feature vectors are divided into m groups based on their first label i (e.g., a feature vector is in group i if it has a label i). The feature vectors of each group are used to train a model (i.e., age estimator). Totally m age estimators AE_i ($i = 1, \ldots, m$) will be trained.
6) *Testing:* A testing feature vector is first assigned to an age estimator based on its second label i_c ($i_c = 1, \ldots, m$). If $i_c = 1$, then the age of the testing feature vector will be predicted by the age estimator AE_1, and the predicted age will be confined to the age range of group 1.
7) The SVR with a linear kernel is used in training and testing.

The procedure of age estimation within an age group is illustrated in Figure 3.

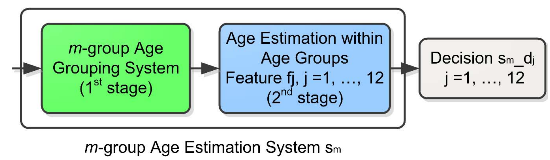


Fig. 4. The m-group age estimation system.

VI. Fusion of Decisions

To further improve the performance, we investigate several fusion schemes based on the decisions (i.e., estimation results from the second stage) in the third stage. Six fusion schemes are compared: intrA-system Fusion (AF), intEr-system Fusion (EF), intrA-intEr Fusion (AEF), intEr-intrA Fusion (EAF), Maximum-Diversity Fusion (MDF), and Composite Fusion (CF). Here, we call the m-group age estimation system simply a system, which is a combination of the first and the second stages as shown in Figure 4.

In FG-NET, we investigate m = 3, 4, 5, 6, 7, 8, 9, 10-group age estimation systems, and each system has 12 decisions (i.e., 12 estimation results from 12 AEs). Totally, 96 decisions can be used for fusion. In MORPH-II, due to lack of age 0 to 15, only m = 2, 3, 4, 5, 6, 7-group systems are investigated. Each system has 12 decisions, and 72 decisions can be used for fusion. In this section, we will use FG-NET as an example to explain the proposed fusion schemes.

In the fusion stage, prediction results obtained from the second stage are treated as input features and used to train another SVR function. Age estimation is realized through a similar procedure in the second stage (Steps 2-7 in an age estimator training). In this stage, we propose two decision selection algorithms for six fusion schemes to find a decision subset, which will demonstrate a competitive result as compared with the optimal subset obtained by the exhaustive search.

A. Diversity Analysis

The goal of analyzing the diversity between different decisions is to find out how to fuse them in a more efficient way for performance improvement. Since each estimator would make different errors on different faces, fusion of these estimators can reduce the total estimation error. For this reason, it is desired to fuse a set of estimators whose decisions

TABLE IV

CORRELATION (p) BETWEEN ANY TWO DECISIONS $d_1 - d_{12}$ IN THE 3-GROUP SYSTEM (s_3), MEAN CORR = 0.9333, ON FG-NET

p	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}
d_1	1.000	0.991	0.991	0.962	0.955	0.956	0.906	0.906	0.897	0.976	0.972	0.969
d_2	0.991	1.000	0.980	0.961	0.961	0.959	0.901	0.911	0.899	0.982	0.979	0.976
d_3	0.991	0.980	1.000	0.964	0.958	0.957	0.907	0.910	0.898	0.979	0.976	0.972
d_4	0.962	0.961	0.964	1.000	0.938	0.939	0.910	0.899	0.883	0.958	0.953	0.950
d_5	0.955	0.961	0.958	0.938	1.000	0.942	0.880	0.905	0.889	0.970	0.974	0.968
d_6	0.956	0.959	0.957	0.939	0.942	1.000	0.887	0.891	0.893	0.958	0.960	0.960
d_7	0.906	0.901	0.907	0.910	0.880	0.887	1.000	0.832	0.831	0.899	0.894	0.893
d_8	0.906	0.911	0.910	0.899	0.905	0.891	0.832	1.000	0.831	0.911	0.914	0.908
d_9	0.897	0.899	0.898	0.883	0.889	0.893	0.831	0.831	1.000	0.901	0.902	0.908
d_{10}	0.976	0.982	0.979	0.958	0.970	0.958	0.899	0.911	0.901	1.000	0.989	0.984
d_{11}	0.972	0.979	0.976	0.953	0.974	0.960	0.894	0.914	0.902	0.989	1.000	0.987
d_{12}	0.969	0.976	0.972	0.950	0.968	0.960	0.893	0.908	0.908	0.984	0.987	1.000

TABLE V

CORRELATION (p) BETWEEN ANY TWO m-GROUP SYSTEMS ($s_3 - s_{10}$) FOR DECISION d_1, MEAN CORR = 0.7777, ON FG-NET

p	$s_3\_d_1$	$s_4\_d_1$	$s_5\_d_1$	$s_6\_d_1$	$s_7\_d_1$	$s_8\_d_1$	$s_9\_d_1$	$s_{10}\_d_1$
$s_3\_d_1$	1.000	0.851	0.829	0.812	0.783	0.750	0.704	0.636
$s_4\_d_1$	0.851	1.000	0.839	0.792	0.786	0.745	0.701	0.632
$s_5\_d_1$	0.829	0.839	1.000	0.841	0.841	0.766	0.726	0.649
$s_6\_d_1$	0.812	0.792	0.841	1.000	0.846	0.829	0.772	0.693
$s_7\_d_1$	0.783	0.786	0.841	0.846	1.000	0.856	0.821	0.748
$s_8\_d_1$	0.750	0.745	0.766	0.829	0.856	1.000	0.899	0.808
$s_9\_d_1$	0.704	0.701	0.726	0.772	0.821	0.899	1.000	0.818
$s_{10}\_d_1$	0.636	0.632	0.649	0.693	0.748	0.808	0.818	1.000

are different from those of others. Some quantitative metric is needed to measure the diversity between pair-wise decisions of estimators. Here, we propose to use Pearson's linear correlation coefficient p to measure the diversity between pair-wise decisions ($d_1, \ldots, d_{12}$), where $0 \leq p \leq 1$. The maximum diversity is observed when $p = 0$, indicating the two decisions are uncorrelated.

Table IV shows the intra correlation p_{intra} between any two decisions ($s_3\_d_i, s_3\_d_j$) for 3-group system s_3. Table V shows the inter correlation p_{inter} between any two systems' decisions ($s_m\_d_1, s_n\_d_1$) for decision d_1. It is clear that the intra-system correlation p_{intra} is higher than the inter-system correlation p_{inter}, which means intra-system diversity Div_{intra} is lower than inter-system diversity Div_{inter}. Thus, it is expected that inter-system fusion will offer greater performance gain than intra-system fusion. It is verified by experimental results that the performance of inter-system fusion (indicated by the MAE in Table XV) is truly better than that of intra-system fusion (indicated by the MAE in Table XII).

B. Intra-System Fusion (AF)

For each m-group age estimation system, the 12 AEs will deliver 12 different decisions $d_1, \ldots, d_{12}$, and there are 2^{12} possible ways of selection for fusion. A systematic algorithm is needed to find a proper subset from 12 decisions $d_1, \ldots, d_{12}$. We propose to use the sequential forward selection (SFS) algorithm to achieve this goal. The intra-system fusion scheme is illustrated in Figure 5.

First, given a decision set $D = \{d_j | j = 1, \ldots, 12\}$, we want to find a subset $D_N = \{d_{i1}, d_{i2}, \ldots, d_{iN}\}$, where $N \leq 12$, to optimize the following objective function

$$J(D_N) = MAE(f(D_N), A_{GT}), \tag{2}$$

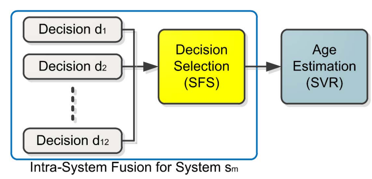


Fig. 5. The intra-system fusion (AF) scheme.

where MAE, which is defined in (3) in Sec. VII-D, is the mean absolute error between the estimated age, $f(D_N)$, and the ground truth age, A_{GT}. The objective function evaluates feature subsets by their estimation accuracy with cross validation to avoid overfitting. The Sequential Forward Selection (SFS) is one of the simplest greedy search algorithms to achieve the above goal. Starting from a decision set D_k (being empty at the start), we sequentially add one decision d^* that results in the lowest objective function $J(D_k + d^*)$ between the ground truth age and the estimated age, $f(D_k + d^*)$, to the set when being combined with the decision set, D_k, that have been selected. Algorithm 1 is given in the next page for clarity.

An example of intra-system fusion is given below. Consider the 3-group age estimation system and we attempt to find a subset from 12 decisions ($s_3\_d_j$, $j = 1, \ldots, 12$) using the SFS algorithm to obtain the best performance (with the smallest MAE). First, each decision $\{d_j\}$ for $j = 1, \ldots, 12$ is selected for MAE performance evaluation. If d_7 provides the smallest MAE (denoted as MAE_1), then the decision subset, DS_3, is updated to $\{d_7\}$. Next, every decision set (d_7, d_j) for $j \neq 7$ is selected for MAE performance evaluation. If (d_7, d_3) provides the smallest MAE (denoted as MAE_2), which is also smaller than MAE_1, then the decision subset is updated to $DS_3 = \{d_7, d_3\}$. After that, every decision set (d_7, d_3, d_j) for $j \neq 3, 7$, is selected for performance evaluation. If (d_7, d_3, d_8)

Algorithm 1 Sequential Forward Selection (SFS)

Input: a decision set $D = \{d_j | j = 1, \ldots, 12\}$.
1) Start with the empty decision set $D_0 = \{\Phi\}$.
2) Select the next best decision.

$$d^* = \arg\min_{d \in D - D_k} J(D_k + d)$$

3) Update

$$D_{k+1} = D_k + d^*;$$
$$k = k + 1.$$

4) Go to 2).
Output: a subset $D_N = \{d_{i1}, d_{i2}, \ldots, d_{iN}\}$.

Algorithm 2 Sequential Backward Selection (SBS)

Input: a system set $S = \{s_j | j = 3, \ldots, 10\}$.
1) Start with the full system set $S_0 = \{\Omega\}$.
2) Remove the worst system.

$$s^* = \arg\min_{s \in S_k} J(S_k - s)$$

3) Update

$$S_{k+1} = S_k - s^*;$$
$$k = k + 1.$$

4) Go to 2).
Output: a subset $S_N = \{s_{i1}, s_{i2}, \ldots, s_{iN}\}$.

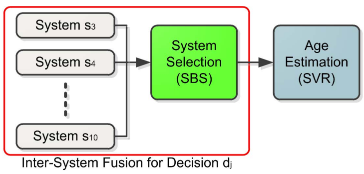


Fig. 6. The inter-system fusion (EF) scheme.

provides the smallest MAE (denoted as MAE_3), which is not smaller than MAE_2, then the decision subset will not be updated and the final decision subset is $DS_3 = \{d_7, d_3\}$.

C. Inter-System Fusion (EF)

To investigate the effectiveness of the inter-system fusion, we focus on one specific decision (e.g., d_1) from 12 decisions and select it from 8 systems ($s_3, s_4, \ldots, s_{10}$) in FG-NET (or 6 systems in MORPH-II) for fusion, where s_m represents the m-group age estimation system. To select good candidates, we could adopt the SFS algorithm to find a good subset from 8 systems s_m ($m = 3, 4, \ldots, 10$). However, different systems have different age ranges for their age groups, and we expect that the same decisions from different systems (e.g., $s_m\_d_1$ for $m = 3, 4, \ldots, 10$) would exhibit higher diversity than different decisions from the same system (e.g., $s_3\_d_j$ for $j = 1, 2, \ldots, 12$). Thus, we propose to use the Sequential Backward Selection (SBS) algorithm to find a subset from 8 systems s_m ($m = 3, 4, \ldots, 10$) to achieve the minimum MAE. The inter-system fusion scheme is illustrated in Figure 6.

SBS works in the opposite direction of SFS. First, SBS starts with a full system set $S = \{s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\}$, and it sequentially removes a system s^* that least reduces the value of the objective function $J(S_k - s^*)$. It stops until a subset $S_N = \{s_{i1}, s_{i2}, \ldots, s_{iN}\}$, where $N \leq 8$, optimizes the objective function. The SBS algorithm (Algorithm 2) is shown at the top of this page.

For given decision d_1, SBS first evaluates the MAE of the fusion of the full system set $S = \{s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\}$, which is denoted by MAE_0. Then, it removes one system and evaluates MAE for the fusion of 7 systems, where the fusion of 7 systems (with s_{10} removed) leads to the smallest MAE as denoted by MAE_1. Then, the system subset is updated to $SS_1 = \{s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$ if MAE_1 is smaller

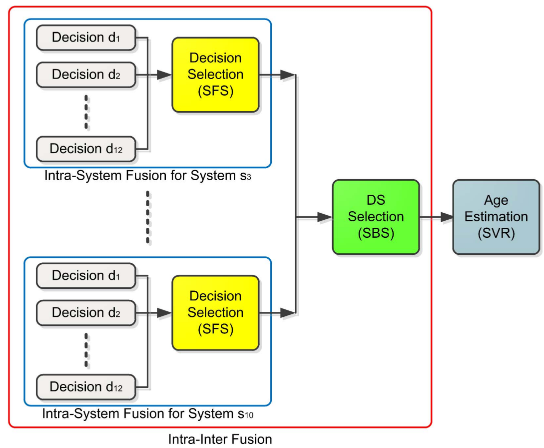


Fig. 7. The intra-inter fusion (AEF) scheme.

than MAE_0. The update stops if MAE_{i+1} is equal or greater than MAE_i.

D. Intra-Inter Fusion (AEF)

Through the intra-system fusion, each system s_m has its best decision subset DS_m selected, where $m = 3, \ldots, 10$. In addition to the intra-system information, we would like to add the inter-system information. The basic idea of intra-inter fusion is described below.

Given decision subsets $\{DS_m, m = 3, \ldots, 10\}$ obtained from intra-system fusion, we perform the fusion on decision subsets from all systems using the SBS algorithm. First, the full set $\{DS_3, \ldots, DS_{10}\}$ is evaluated with MAE_0. Then, it removes one DS_m and evaluates MAE_1 for fusion of the remaining 7 $DS_n (n \neq m)$. If MAE_1 is smaller than MAE_0, the DS subset is updated. If MAE_1 is not smaller than MAE_0, the DS subset is not updated. The same procedure continues until MAE_{i+1} is equal or greater than MAE_i. The intra-inter fusion scheme is illustrated in Figure 7.

E. Inter-Intra Fusion (EAF)

After the inter-system fusion, each decision d_j has its best system subset SS_j selected, where $j = 1, \ldots, 12$. Besides the inter-system information, we would like to add the

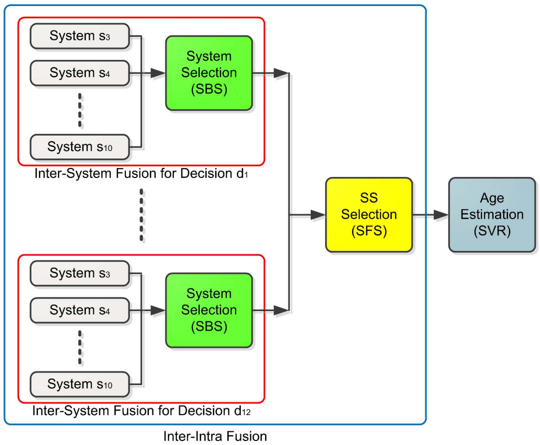


Fig. 8. The inter-intra fusion (EAF) scheme.

intra-system information. The procedure of the inter-intra fusion is described below.

Given system subsets $\{SS_j, j = 1, \ldots, 12\}$ obtained from the inter-system fusion, the SFS algorithm is used for selecting SS_j to cover the intra-system information. First, each SS_j is selected for MAE_1 evaluation. The SS subset is updated to $\{SS_a\}$ if SS_a has the minimum MAE_1. Then, each $\{SS_a, SS_j\}$ (for $j = 1, \ldots, 12$ & $j \neq a$) is evaluated for MAE_2. If $\{SS_a, SS_b\}$ has the minimum MAE_2 and MAE_2 is smaller than MAE_1, the SS subset is updated to $\{SS_a, SS_b\}$. Otherwise, the SS subset is not updated. The same procedure continues until MAE_{i+1} is equal or greater than MAE_i. The inter-intra fusion scheme is illustrated in Figure 8.

F. Maximum-Diversity Fusion (MDF)

Four fusion schemes mentioned in Secs. VI.B-E are mainly structured in one direction and/or the other direction. Besides, we explore two more fusion schemes that consider two directions at the same time. The first two-directional fusion scheme, called the maximum diversity fusion (MDF), is described below.

Given a full set $M = \{s_n\_d_i | n = 3, \ldots, 10, i = 1, \ldots, 12\} = \{m_j | j = 1, \ldots, 96\}$, where m_j represents $s_n\_d_i$, we find a subset $M_N = \{m_{i1}, m_{i2}, \ldots, m_{iN}\}$ with $N \leq 96$ that optimizes the performance based on the diversity. Initially, a decision with the minimum MAE is selected as the first decision set M_1. Among the remaining decisions that are not yet selected, a decision m will be selected and added to the previous updated decision set M_k, if it offers the maximum diversity (i.e., minimum correlation) with the previous decision set, M_k. We use $DIV(M_k, m)$ to denote the diversity between the decision set, M_k, and decision m. The MDF algorithm (Algorithm 3) is shown at the top of this page.

G. Composite Fusion (CF)

The second two-directional fusion scheme is called the composite fusion (CF). To carry out the decision selection in CF, we consider all of 96 decisions together and use the SFS or the SBS algorithm to find the best subset of 96 decisions. Since most of decisions have low diversity with others, it is less efficient to do selection using the SBS algorithm since its begins with the full set. For this reason, we adopt the SFS algorithm for decision selection in CF. For a given full set C $= \{s_n\_d_i | n = 3, \ldots, 10, i = 1, \ldots, 12\} = \{c_j | j = 1, \ldots, 96\}$, where c_j represents $s_n\_d_i$, we find a subset $C_N = \{c_{i1}, c_{i2}, \ldots, c_{iN}\}$ with $N \leq 96$ that optimizes the objective function $J(C_k)$ given in (2). The CF algorithm (Algorithm 4) is shown at the top of this page.

Algorithm 3 Maximum-Diversity Fusion (MDF)

Input: a full decision set $M = \{m_j | j = 1, \ldots, 96\}$.
1) Start with the best decision set $M_1 = \{m_1\}$.
2) Find the decision having the maximum diversity with M_k.

$$m^* = \arg\max_{m \in M - M_k} DIV(M_k, m)$$

3) Update

$$M_{k+1} = M_k + m^*;$$
$$k = k + 1.$$

4) If $J(M_{k+1}) \leq J(M_k)$, then go to 2).
5) If $J(M_{k+1}) > J(M_k)$, then stop.
Output: a subset $M_N = \{m_{i1}, m_{i2}, \ldots, m_{iN}\}$.

Algorithm 4 Composite Fusion (CF)

Input: a full decision set $C = \{c_j | j = 1, \ldots, 96\}$.
1) Start with the empty decision set $C_0 = \{\Phi\}$.
2) Find the next best decision.

$$c^* = \arg\min_{c \in C - C_k} J(C_k + c)$$

3) Update

$$C_{k+1} = C_k + C^*;$$
$$k = k + 1.$$

4) If $J(C_{k+1}) \leq J(C_k)$, then go to 2).
5) If $J(C_{k+1}) > J(C_k)$, then stop.
Output: a subset $C_N = \{c_{i1}, c_{i2}, \ldots, c_{iN}\}$.

The conceptual diagram of six fusion schemes is shown in Figure 9, where AF, EF, MDF, and CF are 1-level fusion schemes while AEF and EAF are 2-level fusion schemes. The main differences among AF, EF, AEF, and EAF are summarized below. The intra-fusion is to fuse decisions for a given age group with different features for final prediction. The inter-fusion is to fuse decisions from different age grouping systems with the same feature for final prediction. The intra-inter fusion is to determine a decision subset for each grouping system (i.e., each system has its needed features) and, then, select some of the decision subsets. The inter-intra fusion is to find a system subset for each feature and, then, select some of the system subsets. The terms, *intra* and *inter*, are used to make a distinction among the proposed fusion methods. For instance, *inter-intra* means that the inter-group fusion is conducted first and the intra-group fusion next. They simply refer

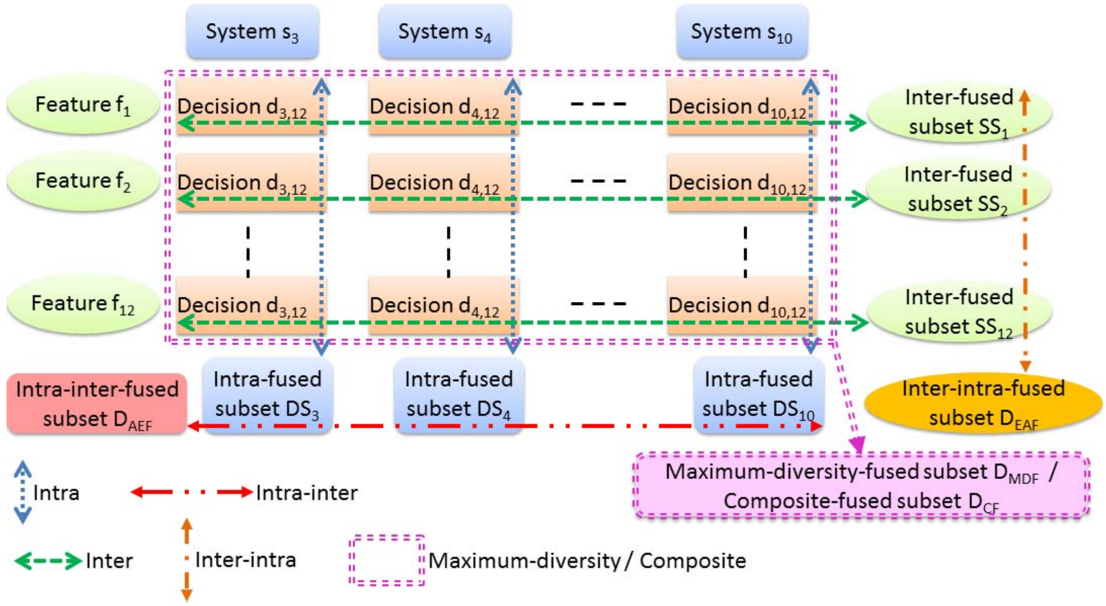


Fig. 9. The conceptual diagram of six fusion schemes.

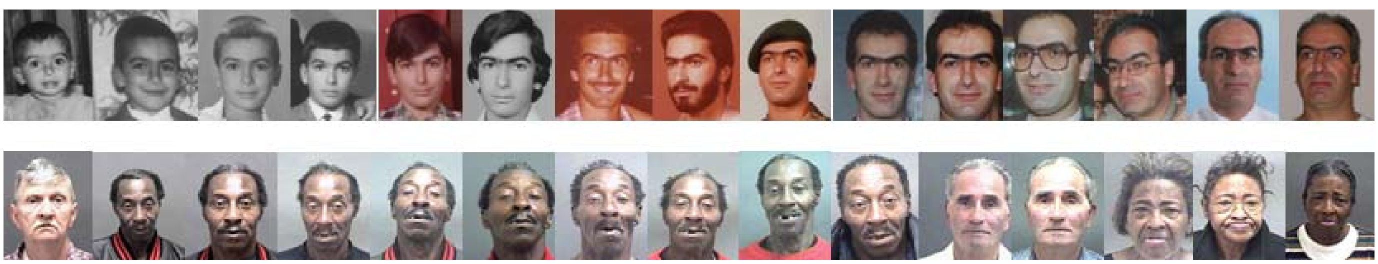


Fig. 10. Some facial images from the FG-NET (top) and MORPH-II (bottom) databases.

to the fusion order, i.e., which fusion scheme is performed first. We will show the performance of these six fusion schemes in Section VII.E.

VII. Experimental Results

A. Database

Two databases used to evaluate the performance of the proposed framework are the FG-NET aging database [30] and MORPH database [31] (MORPH-II is used in our study). The FG-NET database is the most frequently used database for age estimation research since it is publicly available and free. The FG-NET has 1,002 color or gray facial images composed of 82 Europeans with a wide age range from 0 to 69 years old. Each individual has 6-18 images labeled with the ground truth ages. The MORPH-II database contains 55,134 images from 13,618 individuals with ages ranging from 16 to 77. The MORPH-II is a multi-racial database, including African, European, Asian, Hispanic and others. Each individual has about 4 images labeled with the ground truth ages. Some sample images from both databases are shown in Figure 10. The age range distribution of face images is listed in Table VI. The face image is first rotated until the line between two eyes is parallel to the horizontal direction. Then, the facial region is cropped and resized to 180×150 pixels before the age estimation procedure. Only gray level images are used to extract the BIF, HOG and LBPu global and local features.

TABLE VI
Age Range Distribution in the FG-NET and MORPH-II Databases

Age	FG-NET		MORPH-II	
	No. of images	Percentage	No. of images	Percentage
0-9	371	37.03 %	0	0.00 %
10-19	339	33.83 %	7,469	13.55 %
20-29	144	14.37 %	16,325	29.61 %
30-39	70	7.88 %	15,357	27.85 %
40-49	46	4.59 %	12,050	21.85 %
50-59	15	1.50 %	3,599	6.53 %
60-69	8	0.80 %	318	0.58 %
70-77	0	0.00 %	16	0.03 %
Total	1,002	100.00 %	55,134	100.00 %

B. Experimental Setup

To ensure the test data of one stage is not included in the training data of another stage, the same cross validation is carried out at all stages. To compare with other methods, we follow the same experimental settings as others for both databases. The leave-one-person-out (LOPO) cross validation is used on the FG-NET database. The experimental

TABLE VII

AGE RANGES IN THE m-GROUP AGE SYSTEMS FOR THE FG-NET DATABASE AND THE AGE GROUPING RESULTS (m = NO. OF GROUPS)

m	Age range definition for age group										Classification accuracy by			
	AG1	AG2	AG3	AG4	AG5	AG6	AG7	AG8	AG9	AG10	GOP	BIF	HOG	LBPu
3	0-3	4-19	20-69	-	-	-	-	-	-	-	93.5 %	77.5 %	80.8 %	78.8 %
4	0-5	6-12	13-21	22-69	-	-	-	-	-	-	91.3 %	64.5 %	69.0 %	64.3 %
5	0-4	5-10	11-15	16-29	30-77	-	-	-	-	-	87.4 %	59.2 %	62.3 %	58.4 %
6	0-4	5-9	10-14	15-29	30-49	50-77	-	-	-	-	85.5 %	57.9 %	61.2 %	58.0 %
7	0-4	5-9	10-14	15-19	20-25	26-35	36-77	-	-	-	83.6 %	50.4 %	51.4 %	50.9 %
8	0-4	5-9	10-14	15-19	20-29	30-39	40-49	50-77	-	-	79.4 %	51.1 %	50.8 %	49.9 %
9	0-4	5-9	10-14	15-19	20-29	30-35	36-41	42-49	50-77	-	76.0 %	50.7 %	50.2 %	48.9 %
10	0-4	5-9	10-14	15-19	20-29	30-34	35-39	40-44	45-49	50-77	73.5 %	49.7 %	49.5 %	49.1 %

TABLE VIII

AGE RANGES IN THE m-GROUP AGE SYSTEMS FOR THE MORPH-II DATABASE AND THE AGE GROUPING RESULTS (m = NO. OF GROUPS)

m	Age range definition for age group							Classification accuracy by			
	AG1	AG2	AG3	AG4	AG5	AG6	AG7	GOP	BIF	HOG	LBPu
2	16-29	30-77	-	-	-	-	-	92.6 %	73.5 %	75.8 %	80.3 %
3	15-29	30-49	50-77	-	-	-	-	89.8 %	66.6 %	67.6 %	72.3 %
4	15-19	20-25	26-35	36-77	-	-	-	82.8 %	66.7 %	68.6 %	73.6 %
5	15-19	20-29	30-39	40-49	50-77	-	-	78.3 %	54.6 %	57.7 %	60.3 %
6	15-19	20-29	30-35	36-41	42-49	50-77	-	74.9 %	57.5 %	58.7 %	61.0 %
7	15-19	20-29	30-34	35-39	40-44	45-49	50-77	71.4 %	54.4 %	56.3 %	57.0 %

TABLE IX

AGE ESTIMATION RESULTS (IN TERMS OF MAE) OF THE m-GROUP AGE SYSTEMS FOR THE FG-NET DATABASE

Features	MAE (years)								
	$m = 1$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$	$m = 10$
BIF	7.45	4.64	4.03	3.78	3.73	3.69	3.95	4.41	5.13
HOG	7.21	4.51	3.85	3.71	3.65	3.70	3.91	4.36	5.11
LBPu	7.52	4.46	4.23	3.85	3.83	3.71	3.93	4.37	5.14
BIF_EyePair	7.49	4.63	4.11	3.92	3.82	3.74	4.04	4.51	5.19
BIF_Nose	9.28	5.32	4.42	4.22	3.94	3.77	4.14	4.55	5.34
BIF_Mouth	8.25	4.91	4.23	3.99	3.88	3.75	4.09	4.53	5.31
HOG_EyePair	7.89	4.86	4.11	4.02	3.74	3.71	4.01	4.41	5.14
HOG_Nose	8.53	4.97	4.45	4.14	3.98	3.77	4.19	4.55	5.29
HOG_Mouth	8.86	5.11	4.31	3.91	3.70	3.69	4.05	4.49	5.22
LBPu_EyePair	8.15	4.71	4.13	3.79	3.66	3.63	3.95	4.36	5.17
LBPu_Nose	8.76	4.89	4.15	3.88	3.69	3.67	3.99	4.44	5.24
LBPu_Mouth	8.43	4.92	4.04	3.83	3.72	3.69	4.04	4.42	5.23

setting [18], [60] for the MORPH-II database is: the whole MORPH-II database, W, is divided into 3 subsets S1, S2 and S3. The S1 (or S2) is used for training and the remaining W\S1 (or W\S2) is used for testing. The two testing results are then averaged. Since the same cross validation is used throughout all stages, we use the same training data to build a model and the remaining data as the outer example which are to be predicted by the trained model at each stage. Also, we will try to make the source code of our implementations available at [61].

C. Results of Age Grouping

In the age grouping stage, the classification accuracy is used to evaluate our algorithm on FG-NET and MORPH-II databases. To increase diversity between decisions for the fusion stage, multiple age grouping systems are investigated. The age range of each age grouping system is initially defined based on the appearance of the faces. Considering human face has a more visible change during early stage (before 20 years old), we divide the age range in this period into smaller intervals. Since there is no significant change for human face appearances after they become adults ($\geq$20 years old), we choose a larger interval for this age range. However, we need different age grouping systems to increase the diversity. Thus, we have 8 and 6 age grouping systems for FG-NET and MORPH-II, respectively. In addition, we avoid dividing the whole dataset into more than 10 groups since the number of training samples in each group becomes too small, which does affect classification accuracy. The final age range for both databases is determined through repeated experiments. Tables VII and VIII list the age groups for each system in the FG-NET and MORPH-II, respectively. Note that the age ranges for 2-, 3-, ..., 7-group systems in MORPH-II are the same as the age ranges for 5-, 6-, ..., 10-group systems in FG-NET, respectively. The number of groups in MORPH-II is less than that in FG-NET because MORPH-II does not have faces with ages from 0 to 15 year old.

Tables VII and VIII also show the classification accuracy for the proposed age grouping systems. We observe that the overall classification accuracy decreases as the number of groups increases. We created different age grouping systems

TABLE X
AGE ESTIMATION RESULTS (IN TERMS OF MAE) OF m-GROUP AGE SYSTEMS FOR THE MORPH-II DATABASE

Features	MAE (years)					
	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
BIF	4.15	4.47	4.77	4.33	4.08	4.26
HOG	4.22	5.23	4.79	4.33	4.05	4.24
LBPu	4.12	4.65	4.76	4.29	4.04	4.19
BIF_EyePair	4.85	4.92	4.82	4.36	4.09	4.28
BIF_Nose	5.14	5.06	4.86	4.38	4.07	4.29
BIF_Mouth	5.02	4.83	4.84	4.33	4.09	4.28
HOG_EyePair	5.31	5.23	4.76	4.31	4.06	4.23
HOG_Nose	5.63	5.46	4.81	4.34	4.09	4.22
HOG_Mouth	5.66	5.38	4.78	4.35	4.10	4.24
LBPu_EyePair	4.95	4.97	4.73	4.28	4.06	4.21
LBPu_Nose	5.19	5.03	4.78	4.35	4.07	4.22
LBPu_Mouth	4.84	4.81	4.76	4.29	4.07	4.19

TABLE XI
INTRA-FUSION: FINDING DECISION SUBSET BY SFS FOR THE 3-GROUP SYSTEM AGAINST FG-NET

Decision Subset	k-th update	MAE
$D_1=\{d_4\}$	1	4.844
$D_2=\{d_4,d_1\}$	2	4.475
$D_3=\{d_4,d_1,d_2\}$	3	4.349
$D_4=\{d_4,d_1,d_2,d_8\}$	4	4.332
$D_5=\{d_4,d_1,d_2,d_8,d_6\}$	5	4.339

TABLE XII
AGE ESTIMATION RESULTS BY INTRA-FUSION AGAINST FG-NET

System	Decision Subset	No. of updates k	MAE
s_3	$DS_3=\{d_4,d_1,d_2,d_8\}$	4	4.33
s_4	$DS_4=\{d_2,d_1,d_4,d_6\}$	4	3.87
s_5	$DS_5=\{d_2,d_1,d_4\}$	3	3.71
s_6	$DS_6=\{d_2,d_9,d_7,d_1,d_4\}$	5	3.55
s_7	$DS_7=\{d_{11},d_1,d_9,d_{10},d_6,d_4,d_5\}$	7	3.66
s_8	$DS_8=\{d_{10},d_1,d_6,d_7,d_9,d_2\}$	6	3.91
s_9	$DS_9=\{d_{10},d_1,d_7,d_4,d_6,d_9,d_8\}$	7	4.33
s_{10}	$DS_{10}=\{d_{10},d_1,d_2,d_7,d_9\}$	5	5.09

TABLE XIII
AGE ESTIMATION RESULTS BY INTRA-FUSION AGAINST MORPH-II

System	Decision Subset	No. of updates k	MAE
s_2	$DS_2=\{d_4,d_2,d_3,d_7,d_5,d_1\}$	6	3.63
s_3	$DS_3=\{d_1\}$	1	4.72
s_4	$DS_4=\{d_2,d_{10},d_4,d_1,d_5\}$	5	4.38
s_5	$DS_5=\{d_3\}$	1	4.84
s_6	$DS_6=\{d_4,d_{12},d_5\}$	3	4.41
s_7	$DS_7=\{d_{11}\}$	1	4.55

TABLE XIV
INTER-FUSION: FINDING SYSTEM SUBSET FOR d_1 BY SBS AGAINST FG-NET

System Subset	k-th update	MAE
$S_0=\{s_3,s_4,s_5,s_6,s_7,s_8,s_9,s_{10}\}$	0	3.032
$S_1=\{s_3,s_4,s_5,s_6,s_7,s_8,s_{10}\}$	1	3.011
$S_2=\{s_3,s_4,s_5,s_6,s_7,s_8\}$	2	2.983
$S_3=\{s_3,s_4,s_5,s_7,s_8\}$	3	2.989

for the following two reasons: (1) to demonstrate that age estimation accuracy is dependent on the accuracy of age grouping classifiers; and (2) to generate various age estimation results for each test face to result in higher diversity for fusion. The diversity gain is rarely exploited in the context of age estimation. We will demonstrate its effectiveness in the fusion stage, which is the main technical contribution of this work.

TABLE XV
AGE ESTIMATION RESULTS BY INTER-FUSION AGAINST FG-NET

Feature	System Subset	No. of updates k	MAE
f_1	$SS_1=\{s_3,s_4,s_5,s_6,s_7,s_8\}$	2	2.98
f_2	$SS_2=\{s_3,s_4,s_5,s_6,s_7,s_8\}$	2	2.94
f_3	$SS_3=\{s_4,s_5,s_6,s_7,s_8\}$	3	2.99
f_4	$SS_4=\{s_3,s_4,s_5,s_6,s_7,s_9\}$	2	3.01
f_5	$SS_5=\{s_3,s_4,s_5,s_7,s_8,s_9\}$	2	3.11
f_6	$SS_6=\{s_3,s_4,s_5,s_6,s_7,s_8,s_9,s_{10}\}$	0	3.00
f_7	$SS_7=\{s_4,s_5,s_6,s_7,s_8\}$	3	3.02
f_8	$SS_8=\{s_3,s_4,s_5,s_7,s_8,s_9\}$	2	3.18
f_9	$SS_9=\{s_3,s_4,s_5,s_6,s_7,s_8\}$	2	3.09
f_{10}	$SS_{10}=\{s_4,s_5,s_6,s_7,s_8,s_9\}$	2	2.96
f_{11}	$SS_{11}=\{s_3,s_4,s_5,s_7,s_8,s_9\}$	2	2.98
f_{12}	$SS_{12}=\{s_3,s_4,s_5,s_6,s_7,s_8,s_9\}$	1	2.96

TABLE XVI
AGE ESTIMATION RESULTS BY INTER-FUSION AGAINST MORPH-II

Feature	System Subset	No. of updates k	MAE
f_1	$SS_1=\{s_2,s_3,s_4,s_5,s_6,s_7\}$	0	3.23
f_2	$SS_2=\{s_2,s_3,s_4,s_6,s_7\}$	1	3.22
f_3	$SS_3=\{s_2,s_3,s_4,s_6,s_7\}$	1	3.24
f_4	$SS_4=\{s_2,s_3,s_4,s_5,s_7\}$	1	3.51
f_5	$SS_5=\{s_2,s_4,s_6,s_7\}$	2	3.43
f_6	$SS_6=\{s_2,s_3,s_4,s_7\}$	2	3.41
f_7	$SS_7=\{s_2,s_3,s_4,s_6,s_7\}$	1	3.39
f_8	$SS_8=\{s_2,s_3,s_4,s_6,s_7\}$	1	3.47
f_9	$SS_9=\{s_2,s_3,s_4,s_6,s_7\}$	1	3.46
f_{10}	$SS_{10}=\{s_2,s_3,s_4,s_5,s_6,s_7\}$	0	3.31
f_{11}	$SS_{11}=\{s_2,s_3,s_4,s_7\}$	2	3.37
f_{12}	$SS_{12}=\{s_2,s_3,s_4,s_5,s_7\}$	1	3.34

TABLE XVII
AGE ESTIMATION RESULTS BY INTRA-INTER-FUSION (AEF) AGAINST FG-NET AND MORPH-II

Database	DS Subset	MAE
FG-NET	$\{DS_3,DS_4,DS_5,DS_6,DS_8,DS_{10}\}$	2.86
MORPH-II	$\{DS_2,DS_4,DS_6\}$	3.01

TABLE XVIII
AGE ESTIMATION RESULTS BY INTER-INTRA-FUSION (EAF) AGAINST FG-NET AND MORPH-II

Database	SS Subset	MAE
FG-NET	$\{SS_2,SS_4,SS_1,SS_9,SS_7\}$	2.84
MORPH-II	$\{SS_2,SS_{12},SS_3\}$	3.12

D. Results of Age Estimation Within Age Groups

For age estimation, the performance is measured by the mean absolute error (MAE) and the cumulative score (CS) [6], [9]. The MAE is defined as the average of absolute errors between the estimated ages and the ground truth ages as

$$MAE=\sum_{i=1}^{N}\left|a_i^{'}-a_i\right|/N, \tag{3}$$

where a_i is the ground truth age for test image i, $a_i^{'}$ is its estimated age, and N is the total number of test images. The cumulative score (CS) is defined as

$$CS(L)=(n_{e\le L}/N)\times 100\%, \tag{4}$$

where $n_{e\le L}$ denotes the number of test images whose age estimation makes an absolute error e not larger than L years.

TABLE XIX
AGE ESTIMATION RESULTS BY THE COMPOSITE FUSION (CF) AGAINST FG-NET AND MORPH-II

Database	Decision Subset	MAE
FG-NET	$\{s_7\_d_{11}, s_4\_d_2, s_5\_d_4, s_6\_d_9, s_3\_d_1, s_8\_d_6, s_5\_d_6, s_4\_d_7, s_9\_d_{11}, s_5\_d_1, s_8\_d_9\}$	2.81
MORPH-II	$\{s_2\_d_4, s_4\_d_4, s_2\_d_1, s_4\_d_{12}, s_3\_d_{10}, s_3\_d_{11}, s_4\_d_1, s_2\_d_3, s_3\_d_3\}$	2.97

Each feature listed in Table III is used to test our two-stage m-group age estimation system in Figure 4. The MAE results on FG-NET and MORPH-II are shown in Tables IX and X, respectively. Since age estimation should be easier under a narrower age range, we expect lower MAEs for systems with more groups. On the other hand, the classification accuracy decreases when the number of groups increases at the first stage. Thus, there is a trade-off between age grouping accuracy and the number of age groups. For example, 3-group system (the highest age grouping accuracy, see Table VII) and 10-group system (the largest number of groups) do not have the lowest MAEs for FG-NET. Instead, the 7-group system has the lowest MAEs.

One reason of small MAEs is attributed to age grouping. To demonstrate this point, we use one column ($m = 1$ which means no age grouping) in Table IX to list the MAE results without age grouping. We see that MAEs with age grouping are smaller than those without age grouping.

E. Results of Fusion of Decisions

Intra Fusion (AF): We show the MAE results of using the SFS algorithm to determine a decision subset in the 3-group system for FG-NET. In Table XI, the SFS algorithm keeps updating the subset until the MAE starts to increase. After four times of updates, the fusion subset is finalized to be $\{d_4, d_1, d_2, d_8\}$ and it achieves the lowest MAE. Tables XII and XIII show the decision subsets and MAE results for each system against FG-NET and MORPH-II, respectively. Most systems only need to fuse a few decisions to achieve the best performance.

Inter Fusion (EF): We show the MAE results of using the SBS algorithm to find the system subset for decision d_1. As shown in Table XIV, the SBS algorithm finds the system subset after one update and achieves the lowest MAE. Tables XV and XVI show the system subset and MAE results for each decision against FG-NET and MORPH-II, respectively. Most decisions only need about 2 updates to find the optimal system subset for the best results. It means that SBS can find the desired system subset faster than SFS.

Intra-Inter Fusion (AEF): We apply the SBS algorithm to find the DS subset, where each system has its specific decision subset (as shown in Tables XII and XIII). The fusion of the DS subset and the MAE result are shown in Table XVII.

Inter-Intra Fusion (EAF): We use the SFS algorithm to find the SS subset, where each decision has its specific system subset (as shown in Tables XV and XVI). The fusion of the SS subset and the MAE result are shown in Table XVIII.

Maximum-Diversity Fusion (MDF): The numbers of the needed decisions are 22 and 23 to achieve the best MAE performance by applying MDF to all decisions for final age estimation. The minimum MAEs are 2.98 and 3.26 years on FG-NET and MORPH-II, respectively.

TABLE XX
MAE COMPARISON OF THE PROPOSED FUSION METHODS AGAINST SEVERAL BENCHMARKING FUSION METHODS

Method	FG-NET	MORPH-II
Majority voting	4.47	4.65
Equal weighting	3.25	3.42
Best-worst weighted vote	3.12	3.35
Intra-inter fusion [ours]	2.86	3.01
Inter-intra fusion [ours]	2.84	3.12
Maximum diversity fusion [ours]	2.98	3.26
Composite fusion [ours]	2.81	2.97

Composite Fusion (CF): The experimental results of CF against FG-NET and MORPH-II are shown in Table XIX. We list the final decision subset and its corresponding MAE. The numbers of final selected decisions are 11 and 9, and the minimum MAEs are 2.75 and 2.91 years for FG-NET and MORPH-II, respectively. These are the lowest MAEs among six fusion schemes. For example, only 11 out of 96 decisions are selected to achieve the best results for FG-NET. This means that the remaining 85 decisions are not used because decisions from the same grouping system have low diversity and only about 2 decisions are needed from each grouping system.

Comparison With Other Fusion Methods: Some classical fusion methods, such as majority voting, equal weighting, and best-worst weighted vote [62], are tested using all decisions, i.e., 96 and 72 decisions for FG-NET and MORPH-II, respectively. To compare our fusion methods with others fairly, we consider intra-inter fusion, inter-intra fusion, MDF, and CF. The MAEs of our fusion methods and other classical fusion methods are listed in Table XX. It is clear that the proposed fusion schemes are better than other benchmarking methods.

F. Complexity Comparison

Although the composite fusion gives the best performance, discussion on other five fusion methods is meaningful due to the tradeoff between the computational complexity and the age estimation performance. If the complexity is a main concern, we may select the inter fusion over the composite fusion since the former has lower complexity where their performance difference in MAE is small, which is about 0.13 years and 0.25 years for FG-NET and MORPH-II, respectively. By calculating the total number of arithmetic operations required by the MAE computation and the MAE value sorting in the ascending order, we get the complexities of all fusion schemes and their corresponding exhaustive search. They are listed in Tables XXI, XXII and XXIII.

Fusion of all decisions in a straightforward manner still has high complexity. For example, one needs to fuse all 96 decisions for FG-NET. However, our fusion schemes

TABLE XXI
THE NUMBER OF REQUIRED ARITHMETIC OPERATIONS FOR SELECTING N FROM 12 DECISIONS FOR INTRA AND INTER-INTRA FUSIONS

Method	N=3	N=4	N=5	N=6
Intra fusion	36	46	55	63
Inter-intra fusion	36	46	55	63
Exhaustive search	221	496	793	925

TABLE XXII
THE NUMBER OF REQUIRED ARITHMETIC OPERATIONS FOR SELECTING N FROM 8 DECISIONS FOR INTER AND INTRA-INTER FUSIONS

Method	N = 4	N = 5	N = 6
Inter fusion	30	24	17
Intra-inter fusion	30	24	17
Exhaustive search	71	57	29

TABLE XXIII
THE NUMBER OF REQUIRED ARITHMETIC OPERATIONS FOR SELECTING N FROM 96 DECISIONS FOR MAXIMUM-DIVERSITY AND COMPOSITE FUSIONS

Method	N = 9	N = 10	N = 21
Maximum-diversity fusion	837	925	1827
Composite fusion	837	925	1827
Exhaustive search	$\approx 1.3 \times 10^{12}$	$\approx 1.1 \times 10^{13}$	$\approx 7.8 \times 10^{20}$

TABLE XXIV
MAEs OF DIFFERENT AGE ESTIMATION ALGORITHMS FOR THE FG-NET DATABASE

Method	MAE
WAS [11]	8.06
AGES [11]	6.77
KAGES [63]	6.18
QM [1]	6.55
MLPs [1]	6.98
RUN [20]	5.78
RankBoost [27]	5.67
GP [21]	5.39
BM [19]	5.33
RED-SVM [64]	5.24
LARR [16]	5.07
PFA [24]	4.97
RPK [15]	4.95
MHR [65]	4.87
MTWGP [21]	4.83
PLO [48]	4.82
BIF [13]	4.77
NDF [66]	4.67
FLP [67]	4.61
OHRank [25]	4.48
GEF (intra-fusion) [ours]	3.55
GEF (inter-fusion) [ours]	2.94
GEF (intra-inter-fusion) [ours]	2.86
GEF (inter-intra-fusion) [ours]	2.84
GEF (MDF) [ours]	2.98
GEF (CF) [ours]	2.81

exploit two facts for complexity reduction: diversity of decisions and effective selection algorithms (say, forward/backward selection). The use of the exhaustive search in finding the optimal decision subset is costly while the performance gain is not significant. For example, the optimal system subset obtained by exhaustive search for FG-NET turns to be the same as that obtained by inter-fusion denoted by SS_1 as shown in Table XV. The optimal subset is selected in the sense of minimizing MAE for the training dataset. The subset obtained by our proposed forward/backward selection strikes a better balance between the computational cost and the MAE cost as compared with that obtained by the exhaustive search.

TABLE XXV
MAEs OF DIFFERENT AGE ESTIMATION ALGORITHMS FOR THE MORPH-II DATABASE

Method	MAE
BIF [13]	5.09
OHRank [25]	6.07
KPLS [18]	4.04
KCCA [60]	3.98
RED-SVM [64]	6.49
Rank-FFS [68]	4.42
CSOHR [51]	3.82
GEF (intra-fusion) [ours]	3.63
GEF (inter-fusion) [ours]	3.22
GEF (intra-inter-fusion) [ours]	3.01
GEF (inter-intra-fusion) [ours]	3.12
GEF (MDF) [ours]	3.26
GEF (CF) [ours]	2.97

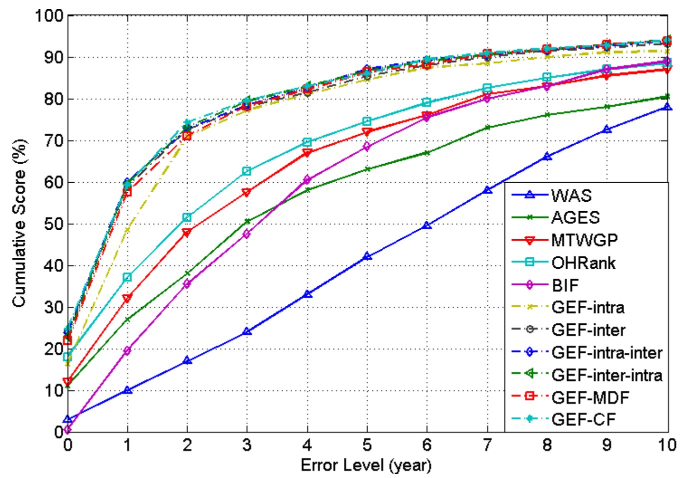


Fig. 11. Cumulative score (CS) curves of error levels from 0 to 10 years of different age estimation algorithms for the FG-NET database.

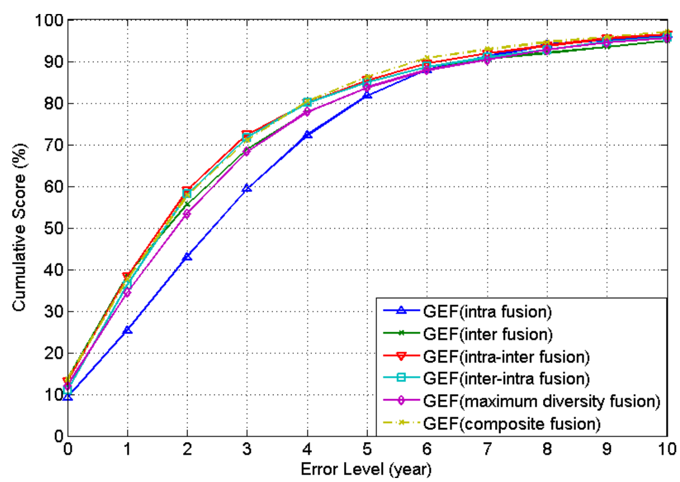


Fig. 12. Cumulative score (CS) curves of error levels from 0 to 10 years of different age estimation algorithms for the MORPH-II database.

G. Performance Comparison

The robustness and effectiveness of the proposed GEF methods are studied in terms of MAEs and cumulative scores (CS) below. First, we list the MAEs of several age estimation methods in Tables XXIV and XXV and the cumulative scores in Figures 11 and 12 for FG-NET and MORPH-II, respectively. It is apparent that the GEF methods outperform other state-of-the-art methods by a significant margin. The best MAEs of the GEF methods are 2.81 and 2.97 years for FG-NET and MORPH-II, respectively.

VIII. Conclusion and Future Work

In this paper, a multistage learning framework called GEF was proposed for age estimation. We present six different fusion schemes to improve the performance. Extensive experiments conducted on two frequently used databases, FG-NET and MORPH-II, demonstrated the effectiveness of the proposed GEF framework. It is possible to improve the performance of the GEF methods by considering the following extensions: 1) to increase the diversity among decisions by including other features or age grouping systems; 2) to explore other decision selection algorithms; and 3) to conduct a theoretical analysis on the diversity gain. It is also worthwhile to verify the robustness of the GEF framework across multiple aging face databases. Moreover, age estimation to variation of facial pose, such as turned or tilted faces, appears to be another interesting problem that can be investigated in our future work.

Acknowledgement

Computation for the work described in this paper was supported by the University of Southern California's Center for High-Performance Computing (http://hpcc.usc.edu).

References

[1] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 621–628, Feb. 2004.
[2] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.
[3] Z. Song, B. Ni, D. Guo, T. Sim, and S. Yan, "Learning universal multi-view age estimator using video context," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 241–248.
[4] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King, "Aspects of age variation in facial morphology affecting biometrics," in *Proc. 1st IEEE Int. Conf. Biometrics, Theory, Appl., Syst. (BTAS)*, Sep. 2007, pp. 1–6.
[5] K. Ricanek, Jr., and E. Boone, "The effect of normal adult aging on standard PCA face recognition accuracy rates," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 4. Jul./Aug. 2005, pp. 2018–2023.
[6] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, 2006, pp. 307–316.
[7] F. Gao and H. Ai, "Face age classification on consumer images with Gabor feature and fuzzy LDA method," in *Advances in Biometrics*. Berlin, Germany: Springer-Verlag, 2009, pp. 132–141.
[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
[9] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 578–584, Jun. 2008.
[10] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 1383–1386.
[11] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.
[12] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Comput. Vis. Image Understand.*, vol. 74, no. 1, pp. 1–21, Apr. 1999.
[13] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 112–119.
[14] S. Yan, M. Liu, and T. S. Huang, "Extracting age information from local spatially flexible patches," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar./Apr. 2008, pp. 737–740.
[15] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
[16] 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 Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.
[17] K. Ueki, T. Hayashida, and T. Kobayashi, "Subspace-based age-group classification using facial images under various lighting conditions," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006.
[18] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 657–664.
[19] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang, "Ranking with uncertain labels," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 96–99.
[20] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.
[21] Y. Zhang and D.-Y. Yeung, "Multi-task warped Gaussian process for personalized age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2622–2629.
[22] S. K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, "Image based regression using boosting method," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Oct. 2005, pp. 541–548.
[23] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer, "Locally adjusted robust regression for human age estimation," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2008, pp. 1–6.
[24] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "A probabilistic fusion approach to human age prediction," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2008, pp. 1–6.
[25] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 585–592.
[26] Y. Ma, T. Xiong, Y. Zou, and K. Wang, "Person-specific age estimation under ranking framework," in *Proc. 1st ACM Int. Conf. Multimedia Retr.*, 2011, Art. ID 38.
[27] P. Yang, L. Zhong, and D. Metaxas, "Ranking model for facial age estimation," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, 2010, pp. 3404–3407.
[28] H. Dibeklioğlu, T. Gevers, A. A. Salah, and R. Valenti, "A smile can reveal your age: Enabling facial dynamics in age estimation," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 209–218.
[29] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognit.*, vol. 44, no. 6, pp. 1262–1281, 2011.
[30] *The FG-NET Aging Database*. [Online]. Available: http://www.fgnet.rsunit.com/, accessed Jul. 24, 2009.
[31] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Proc. 7th IEEE Int. Conf. Autom. Face Gesture Recognit. (FGR)*, Apr. 2006, pp. 341–345.
[32] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, Sep. 2006.
[33] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of age groups based on facial features," *Tamkang J. Sci. Eng.*, vol. 4, no. 3, pp. 183–192, 2001.
[34] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: Wiley, 2012.
[35] P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1529–1532.
[36] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.
[37] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[38] A. Gunay and V. V. Nabiyev, "Automatic age classification with LBP," in *Proc. 23rd Int. Symp. Comput. Inf. Sci. (ISCIS)*, Oct. 2008, pp. 1–4.

[39] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[40] M. A. Hajizadeh and H. Ebrahimnezhad, "Classification of age groups from facial image using histograms of oriented gradients," in *Proc. 7th Iranian Mach. Vis. Image Process. (MVIP)*, 2011, pp. 1–5.

[41] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2005, pp. 886–893.

[42] K.-H. Liu, S. Yan, and C.-C. J. Kuo, "Age group classification via structured fusion of uncertainty-driven shape features and selected surface features," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2014, pp. 445–452.

[43] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face verification across age progression using discriminative methods," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 1, pp. 82–91, Mar. 2010.

[44] P.-K. Sai, J.-G. Wang, and E.-K. Teoh, "Facial age range estimation with extreme learning machines," *Neurocomputing*, vol. 149, pp. 364–372, Feb. 2015.

[45] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Oct. 2005, pp. 786–791.

[46] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.

[47] A. R. Webb, *Statistical Pattern Recognition*. New York, NY, USA: Wiley, 2003.

[48] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2570–2577.

[49] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang, "A study on automatic age estimation using a large database," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1986–1991.

[50] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the FG-NET ageing database," *IET Biometrics*, May 12, 2015.

[51] K.-Y. Chang and C.-S. Chen, "A learning framework for age rank estimation based on face images with scattering transform," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 785–798, Mar. 2015.

[52] H. Dibeklioglu, F. Alnajar, A. A. Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1928–1943, Jun. 2015.

[53] S. A. Glantz, *Primer of Biostatistics*. New York, NY, USA: McGraw-Hill, 2005.

[54] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. ID 27.

[55] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Dec. 2001, pp. I-511–I-518.

[56] K.-H. Liu, T.-J. Liu, and H.-H. Liu, "A sift descriptor based method for global disparity vector estimation in multiview video coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2010, pp. 1214–1218.

[57] T.-J. Liu, K.-H. Liu, and H.-H. Liu, "Temporal information assisted video quality metric for multimedia," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2010, pp. 697–701.

[58] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–1807, May 2013.

[59] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "A multi-metric fusion approach to visual quality assessment," in *Proc. 3rd Int. Workshop Quality Multimedia Exper. (QoMEX)*, 2011, pp. 72–77.

[60] G. Guo and G. Mu, "Joint estimation of age, gender and ethnicity: CCA vs. PLS," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FGR)*, Apr. 2013, pp. 1–6.

[61] [Online]. Available: https://sites.google.com/site/kuanhsienliu23/home/downloads, accessed Dec. 2015.

[62] F. Moreno-Seco, J. M. Inesta, P. J. P. de León, and L. Micó, "Comparison of classifier fusion methods for classification in pattern recognition tasks," in *Structural, Syntactic, and Statistical Pattern Recognition*. Berlin, Germany: Springer-Verlag, 2006, pp. 705–713.

[63] X. Geng, K. Smith-Miles, and Z.-H. Zhou, "Facial age estimation by nonlinear aging pattern subspace," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 721–724.

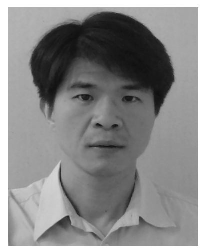
[64] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "A ranking approach for human ages estimation based on face images," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 3396–3399.

[65] T. Qin, X.-D. Zhang, D.-S. Wang, T.-Y. Liu, W. Lai, and H. Li, "Ranking with multiple hyperplanes," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2007, pp. 279–286.

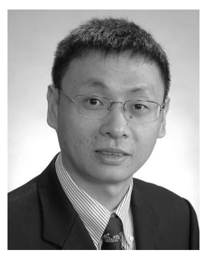
[66] N. Fan, "Learning nonlinear distance functions using neural network for regression with application to robust human age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 249–254.

[67] B. Ni, S. Yan, and A. Kassim, "Learning a propagable graph for semisupervised learning: Classification and regression," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 1, pp. 114–126, Jan. 2012.

[68] Y.-L. Chen and C.-T. Hsu, "Subspace learning for facial age estimation via pairwise age ranking," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 2164–2176, Dec. 2013.



Kuan-Hsien Liu (S'10–M'14) received the B.S. degree in electrical engineering from National Central University, Zhongli, Taiwan, in 1999, the M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 2001, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2014. He is currently a Post-Doctoral Fellow with the Research Center for Information Technology Innovation, Academia Sinica, Taipei. His research interests include digital image analysis and processing, computer vision, and machine learning.



Shuicheng Yan is currently an Associate Professor with the National University of Singapore (NUS), and the Founding Lead of the Learning and Vision Research Group. He has authored or coauthored > 400 technical papers over a wide range of research topics, with Google Scholar > 18,000 times, and H-index of 57. His research areas include machine learning, computer vision, and multimedia. He is an ISI Highly-Cited Researcher in 2014 and a fellow of the International Association for Pattern Recognition in 2014. He received the Best Paper/Demo Awards from ACM MM'13 (Best Paper and Best Student Paper), ACM MM12 (Best Demo), PCM'11, ACM MM10, ICME10, and ICIMCS'09, the Runner-Up Prize of ILSVRC'13, the winner prize of ILSVRC14 detection task, the winner prizes of the classification task in PASCAL VOC 2010–2012, the winner prize of the segmentation task in PASCAL VOC 2012, the honourable mention prize of the detection task in PASCAL VOC'10, the 2010 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Associate Editor Award, the 2010 Young Faculty Research Award, the 2011 Singapore Young Scientist Award, and the 2012 NUS Young Researcher Award. He was/is the Area Chair in CVPR, IJCAI, ACM MM, among others, and the TPC Chair of ACM MM15, ACM MM 2017, among others.

C.-C. Jay Kuo (F'99) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1985 and 1987, respectively, all in electrical engineering. He is currently the Director of the Multimedia Communications Laboratory and the Dean Professor of Electrical Engineering with the University of Southern California, Los Angeles, CA, USA. He has coauthored about 230 journal papers, 870 conference papers, and 13 books. His current research interests include digital image/video analysis and modeling, multimedia data compression, communication and networking, computer vision, and machine learning. He is also a fellow of the American Association for the Advancement of Science and the International Society for Optical Engineers.