Pre-print of article that will appear at IJCB 2011.

Fusing with Context: a Bayesian Approach to Combining Descriptive Attributes

*Walter J. Scheirer[1] Neeraj Kumar[2] Karl Ricanek[3] Peter N. Belhumeur[2] Terrance E. Boult[1]

[1]University of Colorado at Colorado Springs & Securics, Inc., Colorado Springs, CO, USA
[2]Columbia University, New York, NY, USA
[3]University of North Carolina Wilmington, Wilmington, NC, USA

Abstract

For identity related problems, descriptive attributes can take the form of any information that helps represent an individual, including age data, describable visual attributes, and contextual data. With a rich set of descriptive attributes, it is possible to enhance the base matching accuracy of a traditional face identification system through intelligent score weighting. If we can factor any attribute differences between people into our match score calculation, we can deemphasize incorrect results, and ideally lift the correct matching record to a higher rank position. Naturally, the presence of all descriptive attributes during a match instance cannot be expected, especially when considering non-biometric context. Thus, in this paper, we examine the application of Bayesian Attribute Networks to combine descriptive attributes and produce accurate weighting factors to apply to match scores from face recognition systems based on incomplete observations made at match time. We also examine the pragmatic concerns of attribute network creation, and introduce a Noisy-OR formulation for streamlined truth value assignment and more accurate weighting. Experimental results show that incorporating descriptive attributes into the matching process significantly enhances face identification over the baseline by up to 32.8%.

1. Introduction

The growing demand for highly accurate surveillance, intelligence, and forensics systems has propelled the unconstrained face identification problem to the forefront of biometric research. Over the past decade, excellent progress has been made towards the constrained and unconstrained face verification problems, but only small incremental advances have been achieved for unconstrained identification. Verification is a fundamentally easier problem than identification, as it only considers discrete pairs of samples for matching, with a claimed identity choosing a comparison identity that is known to the matching system. Identification is made more difficult by the need to identify an unknown identity out of the entire set of (often numerous) known identities. Compounding this structural consideration is the overall environment of the unconstrained scenario, where any number of effects (pose, illumination, expression, sensor noise, focus, weather, etc.) can impact accuracy.

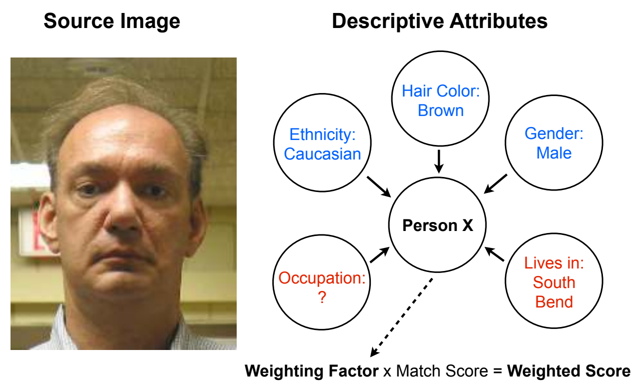

Figure 1. An overview of the attribute fusion approach. By constructing a Bayesian *attribute network* of descriptive attributes during enrollment, it is possible to incorporate additional information into the matching process, while allowing for some unknown attributes. Attributes in blue can be automatically extracted from a face image, while attributes in red reflect non-biometric context.

In the presence of such a challenge, it is advantageous to exploit other information beyond the features used by a face recognition algorithm to improve matching accuracy. For identity related problems, descriptive attributes can take the form of any information that helps represent an individual, including age data [6], describable visual attributes [12], and contextual data. All, when combined, provide a stronger or weaker evidentiary claim for identity. Contextual data is of particular interest, as it brings information to the matching process that is outside of the biometric domain, such as locality, professional descriptors, personal

*This work was supported by ONR SBIR Award Number N00014-11-C-0243, and is an outgrowth of prior work supported by BIMA.

descriptors and action assignment. In a Bayesian framework [4], *attribute networks* of descriptive attributes can be built with appropriate probability assignments for each attribute, leading to a probabilistic "weight" when the network is solved for a set of observations.

For example, consider a case where the enrollment record for a rank-1 match indicates that the person associated with the record in Fig. 1 is a male Caucasian with brown hair, lives in South Bend, and works as a university professor, but the attributes of the probe subject indicate Asian ethnicity, black hair and bushy eyebrows, and residence in Boston. In an attribute network context, the mis-matched attributes produce a low probability solution, which, when applied as a weighting factor to the rank-1 match score produced by the face recognition algorithm, deemphasizes the incorrect result, and ideally lifts the correct matching record to a higher rank position.

While we most often think of soft biometrics as descriptive attributes, contextual attributes can actually be more useful in a fusion framework. In many situations, they are also easier to obtain. Imagine a commercial banking application designed to detect fraud, where a user's face is captured by the camera at the teller's station or in the ATM, and branch locality information is recorded for each transaction. If a particular user banks at a single branch 95% of the time, and in a defined geographic region 100% of the time, then if their account is accessed in a different geographic region and the recorded face biometric does not match their enrollment record, it is likely that an act of fraud is being perpetrated. No added effort is required on the bank's part to record locality information, which is simply being exploited in a new way to add benefit.

Our contributions in this work seek to extend prior work in soft biometric fusion. Here we introduce a Bayesian formulation that incorporates information beyond soft biometrics, including non-biometric contextual data. We also introduce a Noisy-OR formulation for streamlined truth value assignment and more accurate weighting. Finally, we examine the accuracy of Bayesian weighting in the presence of unknown attributes. The experiments presented in this paper incorporate the best robust age estimation and describable visual attribute approaches that have been reported in the literature to date. We show that by incorporating additional information into the matching process, we can significantly enhance the accuracy of a leading face recognition algorithm on an identification problem.

We organize the rest of this paper as follows. In Section 2, we review the relevant prior work related to decision-level fusion for biometric systems, soft biometrics, describable visual attributes, and Bayesian networks for intelligence decisions. In Secs. 3 & 4, we describe our Bayesian approach to combining descriptive attributes, including a Noisy-OR formulation to improve usability and accuracy. To demonstrate the feasibility of our approach, we present a series of experiments for the MBGC data set in Sec. 5. We conclude and sketch some ideas for future work in Sec. 6.

2. Related Work

There is significant interest in several problems related to this work, including decision-level fusion for biometric systems, soft biometrics, describable visual attributes, and Bayesian networks for intelligence decisions. As of this writing, no proposed solution for the above problems, either individually or in combination, fuses biometric data, contextual data and traditional recognition score data in a pragmatic fashion. Thus, our contribution here is the logical next step of combining all of these elements.

Decision-level fusion is defined as data processing by independent classifiers, followed by the fusion of decisions (based upon the calculated results) of each classifier. This idea can be thought of as n different inputs to n different classifiers, producing n decisions that are combined together to produce a final decision that the system will act upon. Several strategies exist for decision-level fusion, including the very basic AND and OR rules, majority voting [14], behavior knowledge space [13], and weighted voting based on the Dempster-Shafer theory of evidence [33]. Similarly, rank level fusion [29] can help decide a more accurate rank configuration between classifiers. However, the strategies for combining traditional recognition score data are not always appropriate for combining descriptive attributes, which are typically weaker (even when combined) as a holistic representation of an identity than a biometric recognition algorithm. Moreover, when we combine observations for attributes that exist over large populations, we'd like to do so in a probabilistic manner, leading us to the Bayesian formulation.

For our attributes, we make the distinction between traditional soft biometrics and descriptive attributes. Soft biometrics, defined by Jain et al. in [11], are "characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals." Examples of soft biometrics are height, weight, ethnicity, gender, etc. Many different approaches to identifying these types of traits have been proposed in the literature, such as LDA [16], full-face SVMs using brightness normalized pixel values [19], Adaboost using Haar-like features [27], and Adaboost using pixel comparison features [2]. The downside to much of the work in soft biometrics has been an emphasis on algorithms for specific soft biometric traits, as opposed to a general classification framework. This is where describable visual attributes come into play.

Describable visual attributes are labels given to an image to describe any aspect of its appearance. The approach suggested by Kumar et al. in [12] has several important advantages over previous work in soft biometrics, including: a general framework to represent objects in the images beyond human features, representation at various levels of specificity, the ability to generalize based on a common feature set, and efficiency at a feature and learning level. Describable visual attributes have been shown to be quite useful for image based search and face verification. Other recent work by IBM in this area has explored hierarchical ranking of attributes computed from texture, shape and color features [7], as well as cross-attribute correlation modeling [28]. In this work, we make use of the robust age estimation of Chen et al. [6], which is explained in detail in Sec. 5.1., as well as the general describable visual attribute methodology of Kumar et al., explained in Sec. 5.2.

Beyond describable visual attributes, we are also interested in contextual attributes. A Bayesian network approach is a natural fit for combining contextual attributes – especially when probabilistic behaviors are of interest, and knowledge of the world is incomplete. The use of Bayesian networks for traditional intelligence and security purposes is well established, with current research focusing on knowledge representation and interaction. Wright et al. [32] introduced techniques to encode military domain and doctrinal expertise in reusable knowledge chunks using Bayesian networks, leading to anomaly detection for military force protection. In a similar approach, Laskey et al. [15] explored Bayesian networks for access pattern analysis, with the goal of detecting abnormal (and hence possibly illegal) activity related to document control.

Finally, we note that our work is not the first to combine a Bayesian framework and biometrics. Early work [3] [31] sought to apply Bayesian classifiers for decision fusion to choose amongst some set of match scores in a multibiometric context, but did not consider soft biometrics or contextual attributes. The well known work of Jain et al. [11] popularized the use of soft biometrics as a complement to traditional biometric recognition through a Bayesian weighting process. While that work established the feasibility of soft biometric fusion, it did not present the makings of an effective and scalable identification system. Specifically, a framework for only soft biometrics is provided, with no clear strategy for handling unknown variables and variable probability assignments that grow exponentially. In this present work, we seek to improve upon the approaches of these first attempts at intelligent Bayesian weighting for recognition systems.

3. Bayesian Weighting Approach

Bayesian networks [4] offer several advantages for descriptive attribute oriented fusion, including a convenient graph based representation (Fig. 2) that expresses relationships, as well as a powerful probabilistic framework for producing weighting factors that can be applied to a conventional biometric recognition system. In essence, a Bayesian network is a directed acyclic graph (DAG), where each vertex possesses a probability density function or (more commonly) a conditional probability table (CPT; shown as a truth table for each vertex in Fig. 2) that expresses the vertex's dependence upon its parents. The probabilities in the table can be derived from collections of statistical sources. For example, Vertex 3's CPT in Fig. 2 is derived from the statistical distribution of gender (~50%), and the error of that particular attribute classifier. The entire network represents the joint probability distribution over all of the variables.

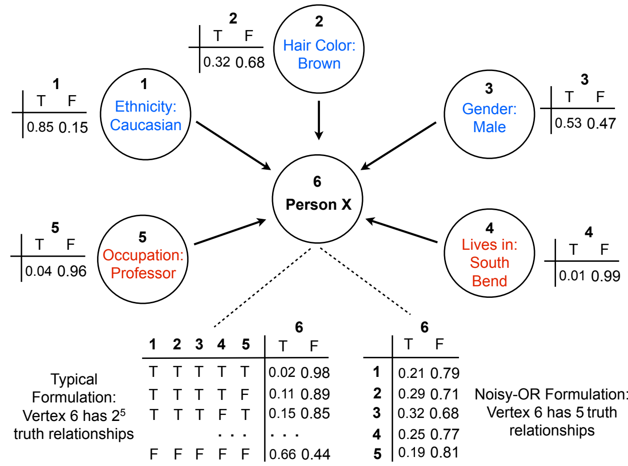

Figure 2. Example Bayesian attribute network in a typical and Noisy-OR formulation for the example of Fig. 1. The Noisy-OR formulation requires exponentially fewer truth relationships.

More formally, let $G = (V, E)$ be a DAG where V is the set of vertices and E is the set of edges. A set of random variables $\mathbf{x}$ is a Bayesian network with respect to G if its joint probability distribution function can be expressed as a product over all vertices in the graph conditioned on the variables corresponding to the parents of a particular vertex:

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i|\mathrm{pa}_i) \qquad (1)$$

where pa_i is the set of parents and $\mathbf{x} = \{x_1, \ldots, x_N\}$.

For our particular problem, we need to combine some set of observed attributes, while accounting for possibly unobserved attributes. This can be done using a conditional probability formula that takes into consideration the values of a CPT at each vertex. Let $\mathbf{x}_k$ be a set of variables representing descriptive attributes observed from a probe that exist in a Bayesian network, and $\mathbf{x}_u$ be a set of variables representing unobserved descriptive attributes that exist in the same network. To compute the probability that a probe is user $\mathcal{U}$ given all $\mathbf{x}_k$ are true, the following formula can be solved:

$$p(\mathcal{U}|\mathbf{x}_k) = \frac{p(\mathbf{x}_k = T, \mathcal{U} = T)}{\sum_{\mathbf{x}_u, \mathcal{U} \in \{T,F\}} p(\mathbf{x}_k = T, \mathbf{x}_u, \mathcal{U})} \qquad (2)$$

where T & F denote truth assignments.

Extending this to our application, assuming the CPTs of an attribute network are an accurate reflection of the interaction between attributes for an individual user, the probability $p(\mathcal{U}|\mathbf{x}_k)$ becomes a useful quantification of confidence that can be directly factored into the biometric matching process. For an identification problem, consider a set of match scores $\{s_1, s_2, \ldots, s_n\}$ produced by a recognition algorithm for a single probe matching against a gallery of n enrollment entries. Assume these scores are normalized to some positive range, where higher scores indicate stronger similarity. Also assume that each enrollment entry also has a corresponding attribute network capturing individual attributes and conditional probabilities. A probability $p(\mathcal{U}_i|\mathbf{x}_k)$ can then be computed for each gallery entry based on its attribute network and the set of descriptive attributes observed from a probe. If we assign each $p(\mathcal{U}_i|\mathbf{x}_k)$ to a weighting factor w_i, we can adjust all of the scores using a multiplicative weighting:

$$\{(s_1 \times w_1), (s_2 \times w_2), \ldots, (s_n \times w_n)\} \tag{3}$$

This approach penalizes gallery entries with many mismatching attributes, while preserving the scores associated with many matching attributes.

4. Noisy-OR Approximation

For many applications, the use of a CPT at each vertex in the attribute network is a convenient representation of variable interaction. However, there are two primary reasons why this representation in its usual formulation is problematic for descriptive attributes. First, for even modestly sized networks, it is not practical to assign all of the truth relationships contained in the CPT by hand [10]. For any vertex, the number of truth relationships grows exponentially as a function of the number of its parent vertices. Considering the example in Fig. 2, if Vertex 6 has five parents, then the number of truth relationships that must be established is 2^5 – a daunting task for an operator of the system, and most likely infeasible given limited sampling or incomplete knowledge of the world. Second, in most cases for descriptive attributes, a model of causal interaction between attributes is illogical [30]: possessing brown hair is completely independent of employment as a university professor. Decoupling attributes from one another is a more correct strategy, and it increases the accuracy of the weights.

With these two issues in mind, we can consider an alternative model for our CPTs called the Noisy-OR function [21]. The Noisy-OR function is a compact representation that serves as a good approximation of the full binary CPTs. Again, let $\mathbf{x}_k$ be a set of variables representing descriptive attributes observed from a probe that exist in a Bayesian network, and $\mathbf{x}_u$ be a set of variables representing unobserved descriptive attributes that exist in the same network. Assume we have n probability values p_i, where p_i is the probability that $\mathcal{U} = T$ conditioned on only $x_i \in \mathbf{x}_k = T$. More specifically, $p_i = p(\mathcal{U} = T|x_i = T, \{x_j = F\}_{j=1,j\neq i}^{n})$, where $x_j \in \mathbf{x}_k$ or $x_j \in \mathbf{x}_u$. To compute the probability that a probe is user $\mathcal{U}$ given all $\mathbf{x}_k$ are true, the following formula can be applied:

$$p(\mathcal{U}|\mathbf{x}_k) = 1 - \prod_{i=1}^{n}(1 - p_i) \tag{4}$$

The above product of probabilities will give us an indication of the chance that all observed attributes occur simultaneously as independent variables. Thus, a very low probability actually indicates a high probability that a series of matching attributes are correct for a specific person, and are not occurring by chance. Recontextualized, we want these low probabilities to be strong weighting factors, hence the subtraction from one in the formula. In the Noisy-OR formulation of the network in Fig. 2, only five different truth relationships must be assigned to Vertex 6. Given the $p(\mathcal{U}|\mathbf{x}_k)$ of Eq. 4, the weighting procedure for the biometric match scores proceeds directly from the description given at the end of Sec. 3.

5. Experimental Evaluation

To evaluate our proposed Bayesian approach to combining descriptive attributes, we examine the following scenario: identification of an unknown individual from a single face image and some amount of non-visual contextual information. From the face images, we estimate the age of the person, and then extract a set of other descriptive facial attributes to be used, along with the contextual information, as observations for an attribute network. Using the weighting approach of Secs. 3 & 4, we readjust the match scores produced by a face recognition algorithm based on our attribute observations.

All data considered here for the biometric identification experiments is drawn from the NIST MBGC set [22]. For this work, we have chosen to use the LRPCA[1] face recognition algorithm [23], which is an established benchmark for NIST testing and a top performer on MBGC data. To emphasize the difficulty of large-scale face identification, we selected 466 unique individuals from MBGC, with one image each for enrollment and testing. Image preprocessing for LRPCA matching took the form of automatic face detection and eye localization (independent of the alignment described below for attribute extraction) using the built-in capabilities of the LRPCA reference implementation. No ground-truth data was used for the recognition process. Match scores were generated via an all-vs-all comparison, where every test image was used as a probe

[1] http://www.cs.colostate.edu/facerec/algorithms/lrpca2010.php

matching across the entire gallery of 466 individuals, leading to a total of 217,156 scores for consideration. The baseline rank-1 recognition rate for this data is 65.9%.

In the rest of this section, we provide exact details of our age estimation and facial attribute extraction approaches, and detail our experimental results. We show that by fusing descriptive attributes with these baseline match scores through intelligent weighting, we are able to significantly lift the baseline recognition rates of LRPCA.

5.1. Robust Age Estimation

Determining a person's age from the face is a difficult task even for humans, let alone a computing device. The difficulty arises from the many variables that affect the outward appearance of the face which range from ethnic-factors to genetic pre-disposition to lifestyle choices. However, it has been established in work by Albert et al. [1] that there are general patterns of adult aging, which in combination with a general understanding of the facial regions that exhibit aging can be used to formulate a system for automatic age-estimation.

The early work of Ricanek et al. [25] laid the foundation for the approach developed by Chen et al. [6], which was used here. Features for age determination are extracted from the face using statistical modeling techniques from detected fiducial points that describe the face (shape, weight, sagging) and texture (most notably wrinkles, lines, ptosis). A combination of Stacked Active Shape Models (STASM) and Active Appearance Models (AAM) are used for automatic landmarking of fiducial points. These models are trained with a set of hand marked exemplars from the MORPH [24] and PAL [18] datasets.

The age-estimation framework used requires the detection and landmarking of the face (see above) for the generation of face features. A method of feature selection is used to determine the best feature combination and weighting for accurate age-estimation via regression. The framework is achieved by the extraction of face features from a set of n aligned face images represented by $\aleph = \{\mathbb{x}_i : \mathbb{x}_i \in R^D\}_{i=1}^n$. where D is the dimension of the face features. The landmarked faces are used to obtain the face features, a set of shape and texture parameters, denoted by $\mathfrak{B} = \{\mathbb{b}_i : \mathbb{b}_i \in R^D\}_{i=1}^n$. Hereafter, an input face image is then represented by the set of fitted parameters. A method for feature selection is used such that coordinates $\mathfrak{B}_1 = \{b_i : b_i \in R^d\}_{i=1}^n$, where $d \ll D$ (See Chen et al. [6] for details on feature selection). With the new vector $\mathfrak{B}_1$ and corresponding age $\{y_i\}_{i=1}^n$, a regression is performed to obtain the model for the future age estimation.

The age-estimator incorporated in this work was built on a set of more than 600 adult faces from the MORPH and PAL face databases. These two databases were used to build the system due to their ethnic and gender diversity and the availability of ground-truth age. The regression technique used is Support Vector Regression (SVR) implemented in the LIBSVM package [5]. The approach adopted from Chen et al. [6] has demonstrated the best performance (as measured by mean absolute error, MAE) against FG-NET [8] of 4.04 years with the next two best performances from Luu et al. [17] of 4.37 years and Gou et al. [9] of 4.77 years. Fig. 3 illustrates the age estimates for three images from MBGC, and Table 2 notes the overall accuracy for the data considered in this paper.

5.2. Facial Attribute Extraction

For facial attribute extraction, we make use of the visual attribute approach of Kumar et al. [12]. As noted in Sec. 2, this approach is highly accurate, with published results on tens of thousands of unique facial images. To achieve this level of performance, a combination of low-level simple features and machine learning is used to build generalized classifiers on a per attribute basis from large amounts of hand labeled exemplars. The input for such a process consists of facial images warped by an affine alignment procedure that makes use of detected fiducial points (eyes, nose, mouth, etc.) that represent the general structure of the face.

From an aligned image I, a set of k low level feature extractors $\mathbf{f}_j$ are applied to form the feature set $\mathcal{F}(I)$:

$$\mathcal{F}(I) = \{\mathbf{f}_1(I), \ldots, \mathbf{f}_k(I)\} \tag{5}$$

Each feature extractor $\mathbf{f}_j$ is composed of four different elements: pixels from a particular region of the face, a choice of pixel value type, normalization, and aggregation (choices for the last three are listed in Table 1). The facial regions capture dominant physical features including the forehead, nose, eyes, mouth, cheeks, eyebrows, and chin.

Pixel Value Type	**Normalization**	**Aggregation**
RGB	None	None
HSV	Mean Norm.	Histogram
Image Intensity	Energy Norm.	Mean/Variance
Edge Magnitude		
Edge Orientation		

Table 1. Feature types for computing describable visual attributes are constructed by applying a pixel conversion from column 1, normalizing via an option from column 2, and aggregating via a strategy from column 3.

Following pixel feature generation, attribute classifiers C_i can be built using supervised learning. In a binary Support Vector Machine approach, data labeled positive and negative (ground-truth data was gathered using Amazon's Mechanical Turk service) for each attribute is used for training. The goal of the machine learning is to build a classifier that generalizes by choosing a subset of the feature set

Figure 3. Example age & face attributes automatically extracted by the algorithms described in Secs. 5.1 & 5.2. The complete list of attributes and their accuracies is provided in Table 2.

$\mathcal{F}(I)$. Using iterative forward feature selection, several individual classifiers on the current set of features in the output set are trained and then concatenated in a region-free combination at each iteration. The performance of each classifier is evaluated using cross-validation, with the features used to train the classifier with the highest accuracy subsequently added to the output set. Features are added until the accuracy stops improving, up to a maximum of 6 low-level features.

The binary SVMs make use of a Radial Basis Function kernel, with all functionality provided by the LIBSVM package [5]. Each classifier utilizes between 500 and 2000 positive and negative examples from the Columbia Face Database [12]. To set the C and γ parameters of the SVM, a grid search is used to find values that maximize classification accuracy. In total, we have access to 73 different attribute classifiers. For the experiments of this paper, we make use of nine of these. Some examples of extracted visual attributes are given in Fig. 3. Overall accuracies for the attributes used in this paper are given in Table 2.

5.3. Results

Based on our list of visual attributes, and a list of contextual attributes (both are found in Table 2), we defined attribute network enrollment records for each person in the gallery. Networks of 5 & 6 vertices were chosen, which contain 20 & 37 total CPT entries respectively for the typical Bayesian formulation, and 9 & 10 total entries respectively for the Noisy-OR formulation. CPT entries for each attribute were assigned in a probabilistic manner specific to each person in the gallery. As can be seen from the drastic reduction in table entries, there was less of a burden building CPTs for the Noisy-OR networks. In total, 10 different combinations of the attributes from Table 2 were chosen for each experiment (each attribute is used at least once), with consistency between the typical and Noisy-OR formulations for a direct comparison.

Visual Attrs. & Accuracy	**Contextual Attrs.**
Age (+/- 7 years); 89.9%	Lives in city X
Gender; 86.7%	Works as X
Eyeglasses; 96.6%	Works at X
Weight: Chubby; 87.8%	Has n children
Eyebrows: Bushy; 88.2%	Is the mother of X
Hair Color: Black; 92.3%	Is the brother of X
Hair Color: Brown; 86.5%	Frequents bank X
Ethnicity: Asian; 94.6%	Owns a car
Ethnicity: African; 97.4%	Attends school X
Ethnicity: European; 87.1%	Graduated in X

Table 2. A listing of descriptive attributes considered for experimentation. All visual attributes are computed from the original source images, while the contextual attributes represent simulated data for each image. Reported accuracies are computed over the MBGC test set used for all experiments in this paper.

Since we do not have actual context for any of the people in the MBGC set, we created a set of simulated context for each identity (simulated data is commonly used for the evaluation of Bayesian frameworks [11, 32, 15]). All visual attributes (age, gender, ethnicity, weight, hair color, etc.), however, are automatically extracted from the source images, and are not simulated in any way. During matching, observations from the probe image are matched to each stored attribute network for each gallery entry, producing a unique weighting factor when each network is solved. This weighting factor is then applied to the original match score from LRPCA for the corresponding gallery entry. The set of observations from the probe is left incomplete, where one variable in the enrollment network is always unknown.

The curves in Fig. 4 summarize our experiments over the typical and Noisy-OR formulations of attribute networks. The rank-n recognition rate for biometric identification is calculated by dividing the number of correct matches by the number of incorrect matches at a particular rank, leading to a rate $\mathcal{R}_n$. Since we have 10 different trials for each experiment, we take the mean of all recognition rates at a particular rank, leading to a rate $\overline{\mathcal{R}}_n$. Mean percentage improvement is then calculated as:

$$\overline{\%\mathcal{I}_n} = (\text{Weighted } \overline{\mathcal{R}}_n - \text{Original } \overline{\mathcal{R}}_n)/\text{Original } \overline{\mathcal{R}}_n \quad (6)$$

We also report a summary of rank-1 accuracies for all experiments in Table 3.

From Fig. 4, it can be seen that a significant amount of accuracy is gained when applying the Bayesian weighting to the LRPCA match scores at every rank shown. As expected, there is an increase in accuracy when we go from five to six vertices in the typical formulation of the Bayesian attribute network (blue curves), reflecting the advantage of more information in the fusion process. Of course, with

Experiment	Vertex Count	CPT Entries	$\overline{\mathcal{R}}_1$	$\overline{\%\mathcal{I}_1}$
Baseline LRPCA	–	–	65.9%	–
Bayesian Weighting	5	20	71.7%	8.9%
Bayesian Weighting	6	37	77.0%	18.8%
Bayesian Noisy-OR Weighting	5	9	77.9%	18.3%
Bayesian Noisy-OR Weighting	6	10	87.5%	32.8%

Table 3. A summary of mean rank-1 accuracies ($\overline{\mathcal{R}}_1$) and mean percentage improvement ($\overline{\mathcal{I}}_1$) for the LRPCA recognition score weighting experiments. Baseline LRPCA accuracy at rank-1 is provided for comparison.

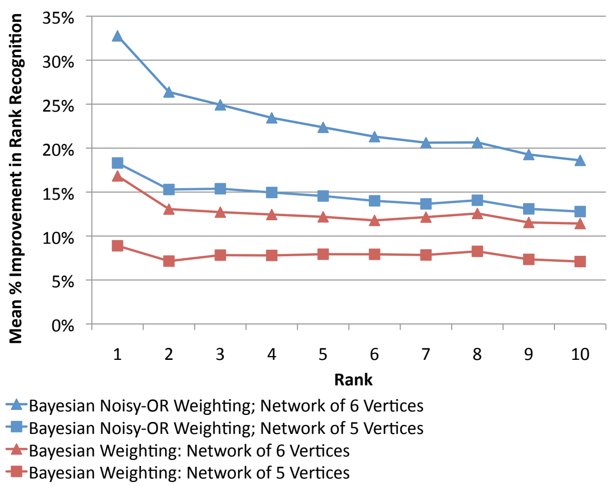

Figure 4. A summary of all experiments performed for LRPCA recognition score weighting over a gallery of 466 people using both the typical (blue) and Noisy-OR (red) formulations for Bayesian attribute networks. Each point represents mean percentage improvement in recognition accuracy at a particular rank ($\overline{\%\mathcal{I}_n}$). Note the increase in accuracy from five to six vertices in the Bayesian network, as well as the significant increase in accuracy of the Noisy-OR formulation over the typical Bayesian networks.

more information, we are faced with increased network complexity. Thus, turning to the Noisy-OR formulation (red curves) in Fig. 4, we see that we do not compromise any accuracy for convenience – in fact, we observe quite the opposite. There is a significant increase in accuracy over the typical Bayesian attribute networks, where the best case of a 6 vertex Noisy-OR network results in a mean rank-1 accuracy of 87.5%. This finding is consistent with previously reported results in other domains [20], where removing variable dependency has been shown to produce more accurate models of the data. Thus, in summary, the Noisy-OR formulation should be used for combining descriptive attributes where little to no dependency exists between them.

6. Conclusion

Descriptive attributes represent an important advance over traditional soft biometrics, with the ability to introduce visual characteristics and non-visual context about an individual as discriminating features for the identification process. The overall goal of applying descriptive attributes to biometric identification is to increase the accuracy of a face recognition algorithm by augmenting the algorithm's feature set with information outside of its direct operation. To combine these descriptive attributes, we require a method that is more sophisticated than typical decision fusion, but still allows us to incorporate a fused representation of the attributes into the biometric decision process. We have considered a Bayesian network framework to do this. Although Bayesian weighting has been previously considered in the biometrics literature, prior work had several limitations, including: a strict focus on just soft biometrics; an assumption of complete knowledge of soft biometrics; and no strategy to handle truth assignments that grow exponentially.

In this paper, we have addressed these issues. We introduced a Bayesian *attribute network* formulation that incorporates descriptive attributes that are well beyond the confines of soft biometrics, including non-biometric contextual data. We also introduced a Noisy-OR formulation for streamlined truth value assignment and more accurate weighting. Finally, we examined the accuracy of Bayesian weighting in the presence of unknown attributes. The experiments presented in this paper incorporate the best robust age estimation and describable visual attribute approaches reported in the literature to date, giving us a state-of-the-art indication of our fusion approach's potential. By using a Noisy-OR formulation to decouple unnecessary attribute dependence, increased accuracy can be achieved over a baseline rank-n identification rate, while simultaneously reducing the computational complexity of the network.

While we have established the groundwork for a more intelligent Bayesian weighting approach in this work, there are several different aspects that can be expanded upon. Most obviously, work to enhance and expand our describable visual attributes is of particular value, and this is ongoing. A more fundamental issue that requires further consideration, however, is related to the Bayesian weighting itself. We made use of a multiplicative score weighting ($s \times w$), which produced promising results, but may not be the optimal strategy. Alternative strategies include applying the weight at the recognition algorithm's sensor or feature level.

Since our weights are, in essence, probabilities, we can also move to more advanced statistical modeling by combining a probabilistic normalization approach [26] with the Bayesian approach described here, giving us a more formal sense of probabilistic certainty.

References

[1] M. Albert, K. Ricanek, and E. Patterson. A Review of the Literature on the Aging Adult Skull and Face: Implications for Forensic Science Research and Applications. *Journal of Forensic Science International*, 172(1):1–9, 2007. 5

[2] S. Baluja and H. Rowley. Boosting Sex Identification Performance. *IJCV*, 71(1), 2007. 2

[3] E. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics. In *IAPR AVBPA*, 1997. 3

[4] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st edition, 2006. 2, 3

[5] C.-C. Chang and C.-J. Lin. *LIBSVM: a Library for Support Vector Machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`. 5, 6

[6] C. Chen, Y. Chang, K. Ricanek, and Y. Wang. Face Age Estimation Using Model Selection. In *IEEE AMFG*, 2010. 1, 3, 5

[7] A. Datta, R. Feris, and D. Vaquero. Hierarchical Ranking of Facial Attributes. In *IEEE AFGR*, March 2011. 3

[8] FG-Net. FG-Net Aging Database. Face and Gesture Recognition Research Network. `http://www.fgnet.rsunit.com/`. Accessed 22 May 2011. 5

[9] G. Guo, G. Mu, F. Fu, and T. Huang. Human Age Estimation Using Bio-Inspired Features. In *IEEE CVPR*, 2009. 5

[10] M. Henrion. Practical Issues in Constructing a Bayes' Belief Network. In *Proc. of the Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-87)*, pages 132–139, New York, NY, 1987. Elsevier Science. 4

[11] A. Jain, S. Dass, and K. Nandakumar. Soft Biometric Traits for Personal Recognition Systems. In *Proc. of the Intl. Conf. on Biometric Authentication*, June 2004. 2, 3, 6

[12] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Describable Visual Attributes for Face Verification and Image Search. *IEEE TPAMI Pre-Print*, 2011. `http://dx.doi.org/10.1109/TPAMI.2011.48`. 1, 2, 5, 6

[13] L. Lam and C. Suen. Optimal Combination of Pattern Classifiers. *Pattern Recognition Letters*, 16(9):945–954, 1995. 2

[14] L. Lam and C. Suen. Application of Majority Voting to Pattern Recognition: Analysis is of its Behavior and Performance. *IEEE TSMC Part A: Systems and Humans*, 27(5):553–568, 1997. 2

[15] K. Laskey, G. Alghamdi, X. Wang, D. Barbara, T. Shackelford, E. Wright, and J. Fitzgerald. Detecting Threatening Behavior Using Bayesian Networks. In *Proc. of the Conference on Behavioral Representation in Modeling and Simulation*, 2004. 3, 6

[16] X. Lu and A. Jain. Ethnicity Identification from Face Images. In *Proc. of SPIE International Symposium on Defense and Security : Biometric Technology for Human Identification*, pages 114–123, 2004. 2

[17] K. Luu, K. Ricanek, , T. Bui, and C. Suen. Age Estimation Using Active Appearance Models and Support Vector Machine Regression. In *IEEE BTAS*, 2009. 5

[18] M. Minear and D. Park. A Lifespan Database of Adult Facial Stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004. 5

[19] B. Moghaddam and M.-H. Yang. Decision Combination in Multiple Classifier Systems. *IEEE TPAMI*, 24(5):707–711, 2002. 2

[20] A. Onisko, M. Druzdzel, and H. Wasyluk. Learning Bayesian Network Parameters From Small Data Sets: Application of Noisy-OR Gates. In *Proc. of the European Conference on Artificial Intelligence (ECAI)*, 2000. 7

[21] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988. 4

[22] P. Phillips, J. Beveridge, W. Scruggs, A. O'Toole, D. Bolme, K. Bowyer, B. Draper, G. Givens, Y. M. Lui, H. Sahibzada, J. S. III, , and S. Weimer. Overview of the Multiple Biometrics Grand Challenge. In *Proc. of ICB*, 2009. 4

[23] P. Phillips, J. R. Beveridge, B. Draper, G. Givens, A. O'Toole, D. Bolme, J. Dunlop, Y. Lui, H. Sahibzada, and S. Weimer. An Introduction to the Good, the Bad, & the Ugly Face Recognition Challenge Problem. In *IEEE AFGR*, March 2011. 4

[24] K. Ricanek and T. Tesafaye. MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. In *IEEE AFGR*, April 2006. 5

[25] K. Ricanek, Y. Wang, C. Chen, and S. Simmons. Generalized Multi-Ethnic Age Estimation. In *IEEE BTAS*, 2009. 5

[26] W. Scheirer, A. Rocha, R. Micheals, and T. Boult. Robust Fusion: Extreme Value Theory for Recognition Score Normalization. In *ECCV*, pages 481–495, 2010. 8

[27] G. Shakhnarovich, P. Viola, and B. Moghaddam. A Unified Learning Framework for Real Time Face Detection and Classification,. In *IEEE AFGR*, 2002. 2

[28] B. Siddiquie, R. Feris, and L. Davis. Image Ranking and Retrieval Based on Multi-Attribute Queries. In *IEEE CVPR*, June 2011. 3

[29] S. S. T.K. Ho, J.J. Hull. Decision Combination in Multiple Classifier Systems. *IEEE TPAMI*, 16(1):66–75, 1994. 2

[30] C. Twardy and K. Korb. Measuring Causal Interaction in Bayesian Networks. In *Proc. of Neural Information Processing Systems*, December 2002. 4

[31] P. Verlinde, P. Druyts, G. Cholet, and M. Acheroy. Applying Bayes Based Classifiers for Decision Fusion in a Multimodal Identity Verification System. In *Intl. Symposium. on Pattern Recognition*, 1999. 3

[32] E. Wright, S. Mahoney, K. Laskey, M. Takikawa, and T. Levitt. Multi-entity Bayesian Networks for Situation Assessment. In *Proc. of the 5th Intl. Conf. on Information Fusion*, November 2002. 3, 6

[33] L. Xu, A. Krzyzak, and C. Suen. Methods for Combining Multiple Classifiers and their Applications to Handwriting Recognition. *IEEE TSMC*, 22(3):418–435, 1992. 2