Face Verification with Aging Using AdaBoost and Local Binary Patterns

Gayathri Mahalingam
Video/Image Modeling and Synthesis (VIMS)
Laboratory
Dept. of Computer Science
University of Delaware
Newark, DE 19711
mahaling@udel.edu

Chandra Kambhamettu
Video/Image Modeling and Synthesis (VIMS)
Laboratory
Dept. of Computer Science
University of Delaware
Newark, DE 19711
chandra@cis.udel.edu

ABSTRACT
In this paper, we study the face verification task across age by constructing a simple but powerful representation of the face which uses Local Binary Pattern (LBP) histograms. The spatial information is incorporated by constructing a hierarchical representation of the face image and computing the LBP histogram at each level. A set of most discriminative LBP features of the face are extracted using the AdaBoost learning algorithm. A strong classifier is built using a set of weak classifiers extracted and is used for classification purposes. Several experiments on the FGnet and the MORPH database were performed and the results indicate a significant improvement in the performance when compared with other discriminative approaches. Performance improvement is achieved with smaller age gaps between image pairs and it stabilizes as the age gap increases. Also, the facial hair, glasses, etc. provide discriminative cues to the system in face verification.

Keywords
Face Verification, Aging, Local Binary Patterns, AdaBoost.

1. INTRODUCTION
Face verification has been an active research area with potential application in law enforcement, crowd surveillance, etc. According to Zhao et al. [31], age related face verification is one of the challenging problems which has gained much attention in recent years. The effect of aging has been studied in the problem of estimating age from facial images ([10], [7], [5], [6], [15]). A detailed survey of contributions from both psychologists and computer scientists on facial aging is given in [24], [31]. Facial age simulation has been studied in [26], [28], [27], [12]. Age simulation techniques involved generating age models which use the shape and texture information from the face images. Lanitis et al. [11] proposed a statistical model for simulating aging effects and used it for face recognition tasks. Tiddeman et al. [29] proposed an aging model for age simulation in which the model is generated using 2D shape vectors.

Face verification tasks under the influence of aging effects have been studied recently. A detailed survey of the effects of aging on face verification tasks is studied by Lanitis [11]. Ramanathan and Chellappa [22] proposed a face growing model and used it for face verification tasks to recognize people under the age of eighteen. Geng et al. [6] proposed a method in which an aging pattern is developed which is defined as the sequence of face images sorted in time order, by constructing a subspace. The aging pattern of the query image is determined by projecting it in the subspace. The position of the face image in aging pattern which is sorted in the time order indicates the age. The above method requires prior information like the actual age of the individual, the facial feature points to model the aging pattern in the facial images. Singh et al. [25] and Park et al. [19] proposed methods that use age transformation techniques for face verification.

Ramanathan and Chellappa [23] use a discriminative approach for face verification across age progression. The authors use probabilistic eigenspace technique and Bayesian model for face identification across age progression. To avoid the challenges caused by illumination variations, they use the half face that has better illumination (termed as Point-Five faces) and use the symmetry property of the face to construct a face image. Ling et al. [13] also use a discriminative approach for face verification across age progression. They propose a face representation called gradient orientation pyramid, in which a Gaussian pyramid is generated for each face image and the gradient orientation is computed for each pixel at all levels of the pyramid. SVM is then used to classify the image pairs as intra-personal or extra-personal. Zhang et al. proposed a boosted classifier that uses Local Binary Patterns as features for verification across same age. The work most relevant to ours is [30], where individual bins have been used instead of using the histograms of Local Binary Pattern (LBP) as features for learning, and the features are classified using a boosted Multi-Task Learning Framework. We vary from the way the features are extracted from the face images. In our approach, we extract the features...
in a hierarchical way by constructing an image pyramid for each face image and computing the LBP at each level and then concatenate them. This gives us an effective feature representation for verification tasks across age.

In this paper, we study the task of face verification across age progression. Our work differs in both the representation of the face images and also the classification framework. We use the Local Binary Pattern (LBP) operator, a powerful descriptor for representing face images and the AdaBoost learning algorithm for classification framework. The results of our experiments are compared with the results from various approaches.

The goal of our study is to present an effective representation of the facial image and an algorithm for face verification with aging effects in human faces. We use the FGnet Aging Database [1] and MORPH Database [9] which are publicly available and are being widely used for image based face aging analysis. The task of face verification across age progression is challenged by several factors. Changes in facial features such as shape (due to weight gain/loss), texture (due to wrinkles, scar, etc.), facial hair have major impact on the task of face verification across age. Besides the biometric changes in the face, the other factors that influence the face verification task are illumination conditions when the image was taken, the image resolution, etc.

2. PROBLEM FORMULATION

In this section, we define our framework for solving the problem of face verification and explain our face description using Local Binary Patterns (LBP) [16], [18], [17]. We then explain the procedure of LBP feature extraction using AdaBoost algorithm.

2.1 Framework

Face verification is the task of classifying a pair of images as intra-personal or extra-personal, i.e., to classify whether the two images belong to the same subject or two different subjects. The advantage of face verification task is that the verification can be performed even when there are limited number of images for each subject, which is most common in datasets across aging. Face verification can be treated as a two-class classification problem and has been used for various face analysis tasks. Phillips [20] used SVM for face recognition tasks. Ramanathan and Chellappa [22], and Moghaddam and Pentland [14] used a Bayesian framework for classifying the image pairs as intra-personal and extra-personal. Ling et al. [13] used a SVM based classification technique combined with Gradient Orientation Pyramid (GOP) for intra-personal and extra-personal classifications. The image is represented using its intensity in all the above methods. In our approach, we use the hierarchical Local Binary Pattern (LBP) operator for representation and apply the framework for face verification problem involving age gaps. In this paper, LBP operator is applied hierarchically to extract the features of a pair of images by mapping it to the LBP feature space which are used to construct a set of weak classifiers. Then we apply the AdaBoost learning algorithm proposed by Yoav et al. [4] to obtain the most discriminant features to represent the image pair. The final strong classifier constructed using a few hundreds of weak classifiers can evaluate similarity between the two images. The entire method is represented in Figure 1.

2.2 Face Description with Local Binary Pattern

The original LBP operator proposed by Ojala et al. [16] is a simple but very efficient and powerful operator for texture description. The operator labels the pixels of an image by thresholding the \( n \times n \) neighborhood of each pixel with the value of the center pixel and considering the result value as a binary number. Figure 2 shows an example of the basic LBP operator. Figure 3 shows the \((4, 1)\) and \((8, 2)\) neighborhood circular LBP operator. The calculation of the LBP labels can be easily done in a single scan of the image. The histogram of the labels of the pixels of the image can be used as a texture descriptor. The grey-scale invariance is achieved by considering a local neighborhood for each pixel, and invariance with respect to the scaling of the grey scale is achieved by considering just the signs of the differences in the pixel values instead of their exact values. The LBP operator was then extended by Ojala et al. [18] which consider different neighborhood sizes. The labels of each pixel are obtained using circular neighborhoods. The bi-linear interpolation of the pixel values from circular neighborhood allows the usage of any radius and number of pixels in the neighborhood. The LBP operator with \( P \) sampling points on a circle of radius \( R \) is given by,

\[
LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p
\]

where

\[
s(x) = \begin{cases} 
1 & \text{if } x \geq 0 \\
0 & \text{if } x < 0
\end{cases}
\]

and \( g_p \) corresponds to the grey value of the center pixel of the local neighborhood pixels with grey values \( g_p \), \( p = \)
Ojala et al. [17] also introduced another extension to the original operator which uses the property called uniform patterns according to which a LBP is called uniform if there exist at most two bitwise transitions from 0 to 1 or vice versa. Uniform patterns represent local micro-patterns of the image such as edges, spots and flat areas. In addition to this, uniform patterns can reduce the dimension of the LBP significantly which is advantageous for face verification. Inverse rotation invariance proposed by Ojala et al. [16], [17] as an extension to the LBP operator. The idea is to rotate the grey values of the neighboring pixels of an image pixel so as to obtain the least binary value for the operator. In our experiments, we use the LBP$^2_{p,k}$, which is the uniform LBP operator with a window size of 5 × 5 around each pixel. In addition, we collect the LBP features in a hierarchical way, which has been shown to retain the most visual information as in [2] and [8].

Given an image $I(x,y)$, where $(x,y)$ indicates pixel locations, we first define the pyramid of $I$ as

$$G_k(I) = I(x,y,k) : k = 0, ..., s$$

with

$$G_0(I) = I(x,y,0)$$

$$G_k(I) = [I(x,y,k-1) \otimes \Phi(x,y)]$$

where $\Phi(x,y)$ is the Guassian kernel and $s$ is the number of pyramid levels. The image at level $k$ of the pyramid is obtained by convolving the Gaussian kernel with the image at level $k-1$, and are scaled down from one level to another. The image at each level of the pyramid is divided into blocks of size 8 and the LBP operator is applied to each block of the image at each level to obtain a LBP histogram for each block. The cumulative LBP histogram is obtained by concatenating the LBP histograms of each block of the image. The LBP operator is applied to all the images in the pyramid and a cumulative LBP histogram is obtained for the images at each level of the pyramid. The LBP pyramid of the image $I$ is defined as follows;

$$L(I) = [LBP(G_0(I)); LBP(G_1(I)); ...; LBP(G_s(I))]$$

where $L(I) \in \mathbb{R}^{d \times s}$ which maps the image $I$ into a $d \times s$ representation, where $d$ is the length of the cumulative LBP histogram obtained from the image at each level of the pyramid. Figure 4 illustrates the computation of a LBP pyramid from an Image.

Given and image pair $(I_1, I_j)$ and corresponding LBP pyramids $L(I_1)$ and $L(I_j)$, the feature vector $x$ is given by

$$x = S(I_i, I_j)$$

where $S$ is defined as the dot product between the LBP histograms of all the image blocks at all the levels of the pyramid and given by

$$x = S(I_i, I_j) = (L(I_i) \ast L(I_j)) \begin{bmatrix} 1 \vdots 1 \end{bmatrix}_{d \times 1}$$

where $\ast$ is the element wise product.

### 2.3 LBP Feature Extraction Using AdaBoost

The face verification is a two-class problem in which image pairs are classified either as intra-personal (both the images in the image pair belong to the same subject) and extra-personal (the images in the image pair belong to different subjects). Given two images $I_i$ and $I_j$, the task is reduced to classify this image pair as either intra-personal or extra-personal. The image pairs are first mapped onto the feature space using equation 7, where $x \in \mathbb{R}^d$ is the feature vector from the d-dimensional feature space and is obtained using the feature extraction function $S : I \times I \rightarrow \mathbb{R}^d$ with the set of all images $I$. However, not all the features extracted are effective. Selecting the most discriminative features improves the performance of a face verification system. We use AdaBoost algorithm for effective feature selection.

AdaBoost introduced by Yoav et al. [4] is a strong tool to solve a two-class classification problem. We use AdaBoost to select a set of discriminative LBP features to form a set of weak classifiers which are used to form a final strong classifier. The algorithm maintains a probability distribution of weights $\omega_t$ over the training set. The initial values for these weights is based on the proportion of intra-personal and extra-personal data. At iteration $t$, the weight $\omega_{t+1,i}$ is decreased by a factor $\beta^{1-e_i}$, where $e_i$ is 0 if the training example $x_i$ is classified correctly and $\omega_{t+1,i}$ remains constant if the example $x_i$ is misclassified. At each iteration, a weak classifier is obtained whose error value $e_t$ is minimum. The final strong classifier obtained from the algorithm is a combination of weighted weak classifiers $h_t$. The AdaBoost algorithm is explained as follows.

- Given the training set $(x_1, y_1), ..., (x_n, y_n)$, where $x_i$ is the data of the $i^{th}$ example, and $y_i = -1, 1$ for extra-personal and intra-personal respectively.
• Initialize weights $\omega_{l,i} = 1/2m, 1/l$ for $y_i = -1, 1$ respectively, where $m$ and $l$ are the number of extra-personal and intra-personal data respectively.

• For $t = 1, \ldots, T$:
  - Normalize the weights $\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{i=1}^{n} \omega_{t,i}}$.
  - For each feature $j$, train a classifier $h_j$ which uses a single feature. The error is evaluated with respect to $\omega_t, \varepsilon_j = \sum_{i} \omega_t |h_j(x_i) - y_i|^2$.
  - Choose the classifier $h_t$ with the lowest error $\varepsilon_t$.
  - update the weights:
    $$\omega_{t+1,i} = \omega_{t,i} \beta_{t}^{1-\varepsilon_t}$$

• The final strong classifier is given as
  $$H(x) = \text{sign} \left( \sum_{t} \alpha_t h_t(x) \right)$$
  where $\alpha_t = \log \frac{1}{\beta_t}$.

The strong classifier thus obtained is a combination of distinctive features. GMLAdaBoost library [3] is used in our experiments.

3. FACE VERIFICATION EXPERIMENTS

3.1 Experimental Setup

Face verification experiments were conducted on the FGnet aging database [1] and MORPH database [9]. The FGnet database consists of 1002 images from 82 subjects with an average of 12 images per subject. The dataset includes a large range of age gaps for each subject, with the largest age gap being 45 years. The MORPH database consists of two sets of images, "Album1" and "Album2". Album1 contains digital scans of photographs of individuals of various ages, with the largest age gap being 29 years. Album2 contains more than 21,000 images of more than 4000 subjects of various ages. We use "Album1" for our experiments. The images in MORPH database are collected under controlled environment, and non-age related variations are relatively small, but are sufficient to significantly affect the performance of the system. The two datasets include images from individuals of different ethnic region, age, etc. which makes these two datasets suitable for our experiments in terms of fair comparison of the performance of the proposed approach. The images from these datasets are preprocessed by manually cropping the face region from the images by aligning the eye locations. The images were resized to 120 x 120 pixels for computational efficiency. The images are then normalized using the histogram equalization technique.

To evaluate the performance of our approach named as LBPH+AdaBoost, we performed experiments on different subsets of FGnet and MORPH databases having different age ranges. We implemented and tested the performance of other discriminative approaches. The approach proposed in [13] uses Gradient Orientation Pyramid (GOP) for representation and SVM for classification. Ramanathan and Chellappa [23] proposed a Bayesian+PFF approach which uses a Bayesian framework and Point Five Face (PFF) for face verification purposes. We also compare the performance with a variant of our approach (LBP+AdaBoost) which does not involve the spatial information which is embedded in the system using the Gaussian pyramid construction. This variant was designed to study the effect of spatial information in the face verification task.

The performance of the algorithms are evaluated using the True Positive Rate (TPR) - False Positive Rate (FPR) curves. The confidence parameter of the AdaBoost algorithm is varied to compute the points on the TPR-FPR curve. The parameters TPR and FPR are defined as,

$$\text{TPR} = \frac{\text{#truly accepted intra-personal pairs}}{\text{#total intra-personal pairs}}$$

$$\text{FPR} = \frac{\text{#falsely accepted extra-personal pairs}}{\text{#total extra-personal pairs}}$$

where an image pair is truly accepted as intra-personal if the images are from the same subject and the image pair is said to be falsely accepted extra-personal pair if the images are from different subjects, but has been classified as an intra-personal pair. The equal error rate (EER), defined as the error rate when a solution has the same TRR and TPR. The EER is also used to measure the performance in addition to TPR and FPR. TRR is defined as,

$$\text{TRR} = \frac{\text{#truly rejected extra-personal pairs}}{\text{#total extra-personal pairs}}$$

Figure 5: TPR-FPR curve for the experiment on the FGnet subset including images of subjects with age above 18.

Figure 6: TPR-FPR curve for the experiment on the FGnet dataset to show the effects of Aging in children.
3.2 Experiments on the FGnet Database

To study the effects of aging in face verification, we performed two experiments using various subsets of the FGnet database. The first subset included images of subjects whose age was above 18 (including 18). The images were chosen such that the faces in the images were frontal or nearly frontal. This subset includes 348 images of 67 subjects. We used 600 intra-personal pairs and 800 extra-personal pairs. These pairs were randomly chosen to avoid bias between the intra and extra personal pairs. Three-fold cross validation was performed on the image pairs generated. The training set and the testing set included image pairs with different age gaps. The training and testing were done on mutually exclusive subjects. The experimental results are shown in Figure 5 and Figure 7. The equal error rates are also given in Table 1 (row 1). There are several observations from the experimental results.

First, it can be seen that the hierarchical LBP feature representation provides a better performance than other methods. The results show that LBP is a simple but powerful feature descriptor and when combined with hierarchical spatial information, the performance is greatly improved. The second observation is that the discriminative way of representing faces will improve the performance of such a system. Since the image pairs chosen for training and testing set mimic the scenario of passport verification task, the discriminative way of representing faces will improve the performance of such a system. Figure 9 shows samples of correctly identified intra-personal image pairs from the FGnet database. The third observation is that the number of intra-personal and extra-personal image pairs and their ratio in the training set significantly affected the performance of the system. Experiments with various numbers of intra and extra-personal pairs were tested and it is observed that the performance is improved with the increase in size of the training set. The increase in the number of extra-personal pairs effectively discriminate the intra-personal pairs of a subject during training than training with equal number of intra and extra-personal pairs.

According to Pittenger and Shaw [21], human faces undergo appearance changes as they get transformed from a child to an adult. It is shown in [21] that the profile of the face of an adult remains stable above age 18. The changes in human faces of an adult are caused due to wrinkles, facial hair, glasses, etc. It is interesting to observe the effect of large facial shape changes of an individual in the performance of the system. As faces of children below 18 years mainly undergo large shape changes, which contrasts aging in adults which mainly involves textural changes, it would be useful to evaluate the performance of our approach when both sets of images are used for training and testing. Thus, the second experiment focused on the entire FGnet dataset where the training and testing set included image pairs of children. For verification tasks, we generated 1270 intra-personal pairs and 1320 randomly generated extra-personal pairs for training and testing purposes. The aim of this experiment was to perform face verification tasks under the influence of child images. The extra-personal pairs are randomly generated to avoid bias in training. The TPR-FPR curve of Figure 6 shows the experimental results. The TPR-TRR curve is shown in Figure 8. The results indicate that the face verification task is harder in the presence of children than the face verification task for adults. The equal error rates (computed using the TPR and True Rejection Rate (TRR)) are shown in Table 1 (row 2). It can be seen that the performance of all the approaches is much lower than the performance of the approaches when classifying adults.

3.3 Experiments on the MORPH Dataset

The Album1 of the MORPH database was used in our experiments. Each subject in the album has about 3 images on an average. The images were preprocessed as explained in section 3.1. We used 800 intra-personal pairs and 1200 randomly chosen extra-personal pairs for training purposes. The TPR-FPR curves are shown in Figure 10, and the TPR-TRR curves are shown in Figure 13. The equal error rates are shown in Table 1 (row 3). From the results it can be seen that the hierarchical representation is robust in face verification. The system provides a better performance than other approaches. This is due to the effective texture representation using the hierarchical LBP.

The MORPH database included images with minimal variations in pose, expressions, hair style, glasses, scar, facial hair, etc., besides age variations. Hence the MORPH database is suitable to study the effect of these influential factors (with a minimal degree of variation in such factors) that affect the
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Bayesian + PFF</th>
<th>SVM + GOP</th>
<th>AdaBoost + LBP</th>
<th>AdaBoost + LBPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGnet (Exp1)</td>
<td>24.3%</td>
<td>22.6%</td>
<td>21.2%</td>
<td>19.4%</td>
</tr>
<tr>
<td>FGnet (Exp2)</td>
<td>28.5%</td>
<td>26.4%</td>
<td>25.8%</td>
<td>24.2%</td>
</tr>
<tr>
<td>MORPH</td>
<td>26.3%</td>
<td>25.3%</td>
<td>24.1%</td>
<td>22.6%</td>
</tr>
</tbody>
</table>

Table 1: Equal Error Rates (EER) from various approaches for FGnet and MORPH dataset

3.4 Effect of Age Gaps

To study the effect of age gaps between image pairs, the image pairs from the FGnet dataset were grouped into four classes based on the age gap between them. The five classes of age gaps used in our experiments are 0 to 2 years, 3 to 5 years, 6 to 8 years, 9 to 11 years, and greater than 11 years. Figure 14 shows the distribution of the number of image pairs across age gaps from the FGnet dataset. The image pairs were classified according to their age gaps, and 10 fold cross validation was performed on all the intra-personal image pairs of each class. Equal number of extra-personal pairs were chosen at random. The verification performance for our approach (AdaBoost+LBPH) and its variant (AdaBoost+LBP) is shown in Figure 15. The results were recorded in terms of the mean of the equal error rates from the cross validation experiment (each fold is run with different parameter setting to obtain the ROC curve, and the EER is the average of the EERs obtained from each fold). From the experiments, we found that the error rates
are increasing as the age gap increases between an image pair. However, the rate at which it increases is drastically reduced as the age gap between an image pair increases. The error rate saturates as the age gap between an image pair increases.

4. CONCLUSION AND DISCUSSION

In this paper, we presented a feature based representation to study the problem of face verification under the influence of aging effects. The Local Binary Pattern (LBP) is used as a simple, but powerful feature descriptor to represent the face images. The spatial information is incorporated by combining the LBP at each level of the Gaussian pyramid constructed for each face image. The experimental results show that the LBP provides a powerful and effective feature representation. However, the hierarchical information does not prove to be useful in the case of aging effects in children. An age model that captures the textural and appearance changes of children can be used to improve the performance of the system. From our experiments, we observed that the age gaps between the image pairs proved significant in the face verification process. We also performed experiments on the MORPH database which included images with slight variations in pose, expressions, glasses, facial hair, etc. The pose and expressions of the face images had a significant effect in the performance of the system due to lack of discriminative features. As a future work, we would like to investigate the effects of disguise, facial hair, scar, etc. in face verification. We also intend to use this approach for face verification tasks on videos.

5. REFERENCES


