Hierarchical Multi-Label Framework for Robust Face Recognition

Lingfeng Zhang Pengfei Dou Shishir K. Shah Ioannis A. Kakadiaris
Computational Biomedicine Lab
Department of Computer Science, University of Houston, Houston, TX, USA
lzhang34@uh.edu, bensondou@gmail.com, sshah@central.uh.edu, ioannisk@uh.edu

Abstract

In this paper, we propose a patch based face recognition framework. First, a face image is iteratively divided into multi-level patches and assigned hierarchical labels. Second, local classifiers are built to learn the local prediction of each patch. Third, the hierarchical relationships defined between local patches are used to obtain the global prediction of each patch. We develop three ways to learn the global prediction: majority voting, ℓ_1-regularized weighting, and decision rule. Last, the global predictions of different levels are combined as the final prediction. Experimental results on different face recognition tasks demonstrate the effectiveness of our method.

1. Introduction

Face recognition is an active topic for researchers in the fields of biometrics, computer vision, image processing and machine learning. In the past decades, both global and local methods have been developed. Global methods learn discriminative information from the whole face image, such as subspace methods [21, 3], Sparse Representation based Classification (SRC) [23, 24] and Collaborative Representation based Classification (CRC) [31, 26]. Although global methods have achieved great success in controlled environments, they are sensitive to the variations of facial expression, illumination and occlusion in uncontrolled real-world scenarios. Proven to be more robust, local methods extract features from local regions. The classic local features include Local Binary Patterns (LBP) [1, 11], Gabor features [28, 19], Scale-Invariant Feature Transform (SIFT) [12, 4] and gray values.

In local methods, more and more efforts focus on patch (block) based methods, which usually involve steps of local patch partition, local feature extraction, and local prediction combination. With intelligent combination, these methods weaken the influence of variant-prone or occluded patches and ensemble the prediction of invariant or unoccluded patches. Martinez [14] proposed to divide face images into several local patches and model each patch with a Gaussian distribution. The final prediction is reached by summing the Mahalanobis distance of each patch. Wright *et al*. [23] extended SRC into a patch version that achieves better performance by a voting ensemble. Taking into account the global holistic features, Su *et al*. [19] developed a hierarchical method that ensembles both global and local classifiers. Fisher linear discriminant classifiers are applied to global Fourier transform features and local Gabor wavelet features. A two-layer ensemble is proposed to obtain the final prediction. To overcome the impact of patch scale, multi-scale patch based methods were proposed. Yuk *et al*. [25] proposed the Multi-Level Supporting scheme (MLS). First, Fisherface based classifiers are built on multi-scale patches. Then, a criteria-based class candidate selection technique is designed to fuse local prediction. Zhu *et al*. [31] developed Patch-based CRC (PCRC) and Multi-scale PCRC (MPCRC). Constrained ℓ_1-regularization is applied to combine each patch's local prediction.

Although previous patch based methods have achieved great performance, they still suffer from two drawbacks: First, since each patch is handled individually, the dependence between different patches is ignored. Second, the performance is much affected by patch size, which is assigned by experimental experience and varies in different databases. To overcome these problems, we propose a Hierarchical Multi-Label (HML) framework for face recognition. First, each face image is hierarchically divided into multi-level patches and assigned multi-labels. Second, a local classifier is constructed on each patch to obtain the local prediction. Then, the global prediction of each patch is learned based on different types of hierarchical relationships. Last, the global predictions of different levels are combined as the final prediction.

The contributions of this paper include: First, we introduce HML framework into face recognition. To the best of our knowledge, this is the first attempt in the domain. Second, we illustrate two ways to construct hierarchical patches. Third, the correlations between different patches are learned based on their hierarchical relationships. This step is usually neglected by previous methods.

The rest of this paper is organized as follows: in Section 2 we discuss related work. Section 3 describes the proposed method. The experimental design, results and analysis are presented in Section 4. Section 5 concludes the paper.

2. Related work

Local features computed from small patches of the face image are less likely to be corrupted than global features. Applying local features starts from the component based method, where local features are extracted and combined first. Then, classifiers are built on the combined local features. Heisele *et al*. [9] introduced the component based Support Vector Machines (SVM) to avoid pose changes. Subspace models are also extended to component based methods, for example Principal Component Analysis (PCA) and Fisher Linear Discriminant (FLD) [6, 10].

For face recognition in an occlusion scenario, researchers also developed methods to detect and eliminate the occluded patches [2]. Oh *et al*. [15] introduced the Selective Local Non-Negative Matrix Factorization (S-LNMF) method. First, PCA and the Nearest Neighbor (NN) classifier are applied to detect the occluded patches. Then, LNMF-based recognition is performed on the occlusion-free patches. Zhao *et al*. [29] proposed to partition the face image into two layers and use the difference of sparsity to detect the occluded patches. The final prediction is also obtained based on the unoccluded patches. The problem with these methods is they are sensitive to the performance of occlusion detection.

The proposed framework is mainly inspired by the Hierarchical Multi-label Classification (HMC) problem, where each sample has more than one label and all these labels are organized hierarchically in a tree or Direct Acyclic Graph (DAG) [18]. Hierarchical information in tree and DAG structures is used to improve classification performance [22, 27]. In this paper, we introduce the hierarchical multi-label framework into patch based face recognition and make use of the hierarchical relationships between different patches to improve their local predictions.

3. Hierarchical multi-label framework

We start with building a non-overlapping tree-structured hierarchical multi-label framework. Let $\mathbb{D} = \{1, 2, \cdots, D\}$ represent a set of hierarchical levels. Given a face image $X \in \mathbb{R}^{u \times v}$, we assume that it is on level 1. First we partition the level 1 image and obtain the level 2 patches. Then we divide each patch on level 2, and obtain the patches on the next level. By repeating the process $D - 1$ times, we have all the hierarchical patches. Let $X_{i,j} \in \mathbb{R}^{u_{i,j} \times v_{i,j}}$ denote the j^{th} patch on level i. Let N and N_i denote the total number of patches and the number of patches on level i, respectively. So we have $N = \sum_{i=1}^{D} N_i$. Meanwhile, we define a hierarchical label set $\mathbb{L} = \{l_{i,j}\}$, where $l_{i,j}$ represents the label of patch $X_{i,j}$. Note that $l_{i,j} \in \{1, 2, \cdots, C\}$, and C represents the total number of classes in the training data. By now, we have partitioned the face image into multi-level patches and assigned a hierarchical label for each patch. Figure 1 depicts an example of a 3-level partition and its corresponding tree-structured patch and label hierarchies. In practice, to emphasize local information, we can also start level 1 with divided patches rather than the original face image. Thus, the corresponding label hierarchy becomes a free tree without any root.

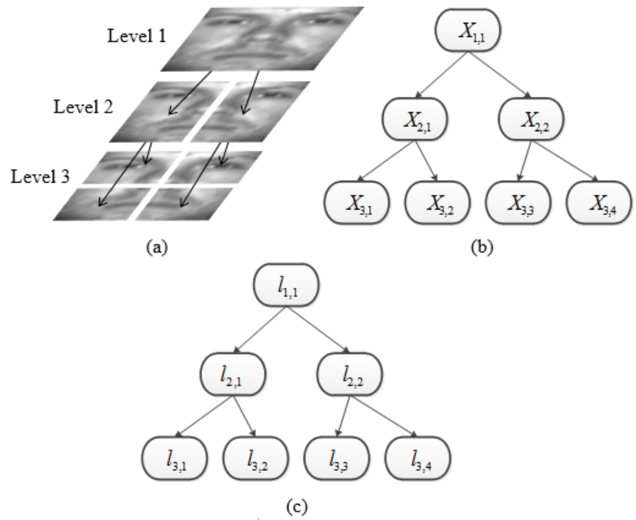

Figure 1: An example of 3-level tree-structured HML framework. (a) Hierarchical face partition. (b) Patch hierarchy. (c) Label hierarchy.

By organizing multi-level patches hierarchically, we can explore the dependence between different patches based on the relationships between their labels. Following the definitions in HMC [7, 18], we define five patch notations and their three hierarchical patch relationships: “parent-child”, “ancestor-descendant” and “siblings”. The notations and examples are shown in Table 1. Note that a patch and its parent patch in the hierarchy have an ”IS-A” relationship, which means that a parent patch always has all the valid pixels of its child patches.

To handle more challenging face recognition tasks, such as 2D-3D face recognition [20, 30], we can also create an overlapping DAG-structured hierarchical multi-label framework, where each patch can have more than one parent patch, and sibling patches can have overlapping regions. One way to build a DAG-structured framework is to remove sub-patches from the parent patch iteratively, which will be explained in Section 4.2. In general, if the number of training samples is sufficient, the tree-structured framework is

Notations	Meanings	Patch Relationship Examples	Label Relationship Examples
$\uparrow (X_{i,j})$	Parent patches of $X_{i,j}$	$X_{1,1} = \uparrow (X_{2,1})$	$l_{1,1} = \uparrow (l_{2,1})$
$\downarrow (X_{i,j})$	Child patches of $X_{i,j}$	$X_{2,1} = \downarrow (X_{1,1})$	$l_{2,1} = \downarrow (l_{1,1})$
$\Uparrow (X_{i,j})$	Ancestor patches of $X_{i,j}$	$X_{1,1} = \Uparrow (X_{3,1})$	$l_{1,1} = \Uparrow (l_{3,1})$
$\Downarrow (X_{i,j})$	Descendant patches of $X_{i,j}$	$X_{3,1} = \Downarrow (X_{1,1})$	$l_{3,1} = \Downarrow (l_{1,1})$
$\Longleftrightarrow (X_{i,j})$	Sibling patches of $X_{i,j}$	$X_{2,2} = \Longleftrightarrow (X_{2,1})$	$l_{2,2} = \Longleftrightarrow (l_{2,1})$

Table 1: The HML patch notations with examples based on Figure 1.

the first choice. It is easy to create and less complex. On the other hand, when we have limited samples in the training data, the DAG-structured framework becomes a better choice. There are two reasons: First, in the tree-structured framework, the discriminative information of each patch is limited by its patch size. Second, more hierarchical information can be extracted from the DAG-structured framework.

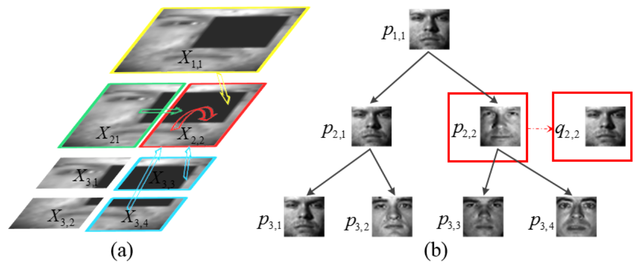

Figure 2: The intuition behind our method in an occlusion scenario. (a) Tree-structured HMC framework on an occluded image. (b) Global prediction correction. The occluded region is marked by a black rectangle. For the occluded patch $X_{2,2}$ in (a), red, yellow, green, cyan arrows represent the contributions of itself, parent, sibling and child patches, respectively. In (b), we can observe that the global prediction of patch $X_{2,2}$ gives a more robust prediction.

3.1. Local prediction

Given a face image set $\mathbb{X} = \{X^1, X^2, \cdots, X^M\}$ and its class label set $\mathbb{Y} = \{y^1, y^2, \cdots, y^M\}$, where $y^m \in \{1, 2, \cdots, C\}$ and $m \in \{1, 2, \cdots, M\}$, the corresponding hierarchical patch sets are denoted by $\mathbb{X}_{i,j} = \{X^1_{i,j},\ X^2_{i,j}, \cdots, X^M_{i,j}\}$. As a patch based method, we build local classifiers $\mathbb{F} = \{f_{i,j}(X_{i,j})\}$ for each patch separately. Let $\mathbb{P} = \{p^m_{i,j}\}$ denote the local prediction set, where $p^m_{i,j}$ represents the predicted labels of $X^m_{i,j}$, and $p^m_{i,j} \in \{1, 2, \cdots, C\}$. We have:

$$p^m_{i,j} = f_{i,j}(X^m_{i,j}). \tag{1}$$

The local prediction here is both feature-free and classifier-free. Any the classifiers (*e.g.*, NN, SRC and CRC) can be used based on any features (*e.g.*, LBP, Gabor features and gray values).

3.2. Hierarchical global prediction

There are several reasons why local prediction is not accurate and only some patches return promising results. First, variations from facial expressions, illumination and occlusion affect different patches differently. Second, some patches are less discriminative than other patches. Third, human faces exhibit distinct structures and characteristics on different-scale patches [31]. Previous methods usually relied on different patch sizes and ensemble methods to address these challenges. However, the correlations between different patches are neglected. In this paper, we use the hierarchical relationships between locally related patches to improve each patch's local prediction and get its new label prediction, which we call "global prediction". Let $\mathbb{Q} = \{q^m_{i,j}\}$ denote the global prediction set, where $q^m_{i,j}$ represents the global prediction of $X^m_{i,j}$, and $q^m_{i,j} \in \{1, 2, \cdots, C\}$. We define a global classifier for each patch to learn the correlation between its global prediction and the local predictions of itself and its parent patches, sibling patches and child patches (if any). For patch $X_{i,j}$, we define its hierarchical prediction matrix as $H_{i,j} = [f(X_{i,j}), f(\uparrow (X_{i,j})), f(\downarrow (X_{i,j})), f(\Longleftrightarrow (X_{i,j}))] \in \mathbb{R}^{M \times S_{i,j}}$, where each element $h^{m,s}_{i,j}$ represents the local prediction of the s^{th} hierarchically related patch for the m^{th} sample and $S_{i,j}$ represents the total number of hierarchically related patches. Let $\mathbb{G} = \{g_{i,j}(X_{i,j}, H_{i,j})\}$ represent the learned global classifier set, so we have:

$$q^m_{i,j} = g_{i,j}(X^m_{i,j}, H^m_{i,j}). \tag{2}$$

Take occlusion for example. Figure 2 depicts the intuition of our method. As we can observe, for the occluded patch $X_{2,2}$, three types of hierarchically related patches (parent patch $X_{1,1}$, sibling patch $X_{2,1}$ and child patches $X_{3,3}$ and $X_{3,4}$) will contribute to correct the erroneous local prediction and obtain the robust global prediction. We introduce three different ways to build the global classifier.

3.2.1 Majority voting (V-HML)

Voting is the most popular ensemble technique in patch based methods. It is easy and training-free. The global prediction of a testing patch is simply given by the majority local prediction of all the hierarchically related patches. For these patches that have more than one majority prediction candidate, the one that gives higher average similarity is selected. The drawback is that majority voting ignores the important differences between different hierarchically related patches. As we can observe in the example of patch $X_{2,2}$ in Figure 2, compared with the child patches $X_{3,3}$ and $X_{3,4}$, the sibling patch $X_{2,1}$ and the parent patch $X_{1,1}$ provide more discriminative information.

3.2.2 ℓ_1-regularized weighting (W-HML)

Based on the above analysis, we introduce different weights to the hierarchically related patches. Let $\mathbf{w}_{i,j} = \{w_{i,j}^1, w_{i,j}^2, \cdots, w_{i,j}^{S_{i,j}}\}^T$ represent the weight vector of the patches related to $X_{i,j}$, and $\sum_{s=1}^{S_{i,j}} w_{i,j}^s = 1$. Following [31], we define a decision matrix $Z_{i,j} = \{z_{i,j}^{m,s}\} \in \mathbb{R}^{M \times S_{i,j}}$ as:

$$z_{i,j}^{m,s} = \begin{cases} +1, & \textit{if } y^m = h_{i,j}^{m,s} \\ -1, & \textit{if } y^m = h_{i,j}^{m,s} \end{cases} . \tag{3}$$

Note that $z_{i,j}^{m,s} = 1$ means that $h_{i,j}^{m,s}$ gives a correct prediction, otherwise it gives a wrong prediction. To measure the misclassification of all the hierarchically related patches, the ensemble margin of the m^{th} sample can be defined as:

$$\varepsilon\left(X_{i,j}^m\right) = \sum_{s=1}^{S_{i,j}} w_{i,j}^s z_{i,j}^{m,s}. \tag{4}$$

For the sample set $\mathbb{X}$, the ensemble loss under square loss can be defined as:

$$\begin{aligned} Loss\left(\mathbb{X}_{i,j}\right) &= \sum_{m=1}^{M} \left[1 - \varepsilon\left(X_{i,j}^m\right)\right]^2 \\ &= \sum_{m=1}^{M} \left(1 - \sum_{s=1}^{S_{i,j}} w_{i,j}^s z_{i,j}^{m,s}\right)^2 \\ &= \|\mathbf{e} - Z_{i,j}\mathbf{w}_{i,j}\|_2^2, \end{aligned} \tag{5}$$

where $\mathbf{e} = [1, 1, \cdots, 1]^T$, and $dim(\mathbf{e}) = S_{i,j}$. Considering that some hierarchically related patches do not make much contribution, like patch $X_{2,2}$'s sibling patch $X_{3,3}$ in Figure 2, we enforce sparsity of $\mathbf{w}_{i,j}$ with ℓ_1-norm. Also the learned weights should be positive. With these constraints, the optimization problem becomes:

$$\begin{gathered} \|\mathbf{e} - Z_{i,j}\mathbf{w}_{i,j}\|_2^2 + \lambda \|\mathbf{w}_{i,j}\|_1 \\ s.t.\ \sum_{s=1}^{S_{i,j}} w_{i,j}^s = 1, w_{i,j}^s > 0,\ s = 1, 2, \cdots, S_{i,j}. \end{gathered} \tag{6}$$

Using the same strategy as [31], converting the weight constraint to $\mathbf{e}\mathbf{w}_{i,j} = 1$, and adding to the objective function, we have:

$$\begin{gathered} \mathbf{w}_{i,j}^* = argmin_{\mathbf{w}_{i,j}} \{\left\|\mathbf{e}' - Z'_{i,j}\mathbf{w}_{i,j}\right\|_2^2 + \lambda \|\mathbf{w}_{i,j}\|_1\} \\ s.t.\ \mathbf{w}_{i,j}^s > 0,\ s = 1,\ 2,\ \cdots, S_{i,j} \end{gathered} , \tag{7}$$

where $\mathbf{e}' = [\mathbf{e}; 1]$, $Z'_{i,j} = \left[Z_{i,j}; \mathbf{e}^T\right]$. The function can be solved using popular ℓ_1-minimization methods. After weight learning, for a testing patch $\hat{X}_{i,j}^k$, the global prediction is $\hat{q}_{i,j}^k = \arg max_c\{\sum w_{i,j}^s | \hat{h}^{k,s} = c\}$.

3.2.3 Decision rule (R-HML)

To find the hidden relationships between different hierarchically related patches, another good method is to use rule-based classifiers [17, 5]. The advantages include: easy to interpret and fast to generate. For the example of patch $X_{2,2}$ in Figure 2, a simple decision rule is:

$$\begin{aligned} &(f\left(\Longleftrightarrow X_{i,j}\right) = c)\ \wedge (f\left(\uparrow X_{i,j}\right) = c) \\ &\longmapsto (g(X_{i,j}, H_{i,j}) = c). \end{aligned} \tag{8}$$

Considering the variety of hierarchical relationships, we use Random Forest [5] to learn the global prediction of each patch. Figure 3 depicts an example of the comparison between local prediction and global prediction in the Extended Yale B database [8] under the Random Forest based global classifier. A 5-level non-overlapping tree-structured HML framework is constructed on each face image. The local prediction of each patch is obtained based on the simplest choices of classifier and features (NN and gray values).

3.3. Hierarchical ensemble

Combining multi-level global prediction is an ensemble learning problem. Every method introduced above for single patch global prediction learning can be applied to combine the global predictions of different levels. Our framework focuses on hierarchically global learning, so we simply use majority voting to get the final prediction. The proposed method is summarized in Algorithms 1 and 2.

4. Experiments

4.1. 2D face recognition in occlusion scenario

The HML framework is first evaluated on the 2D face recognition task in an occlusion scenario with the Extended Yale B database [8] and the AR database [13]. The Extended Yale B database contains 38 subjects under 9 poses and 64 illumination conditions. The image number of each subject ranges from 59 to 64. The classic mandrill image is used to create synthetic square block occlusions on all the images in random positions. The occlusion takes up 25% of the whole face image. The AR database contains over 4,000

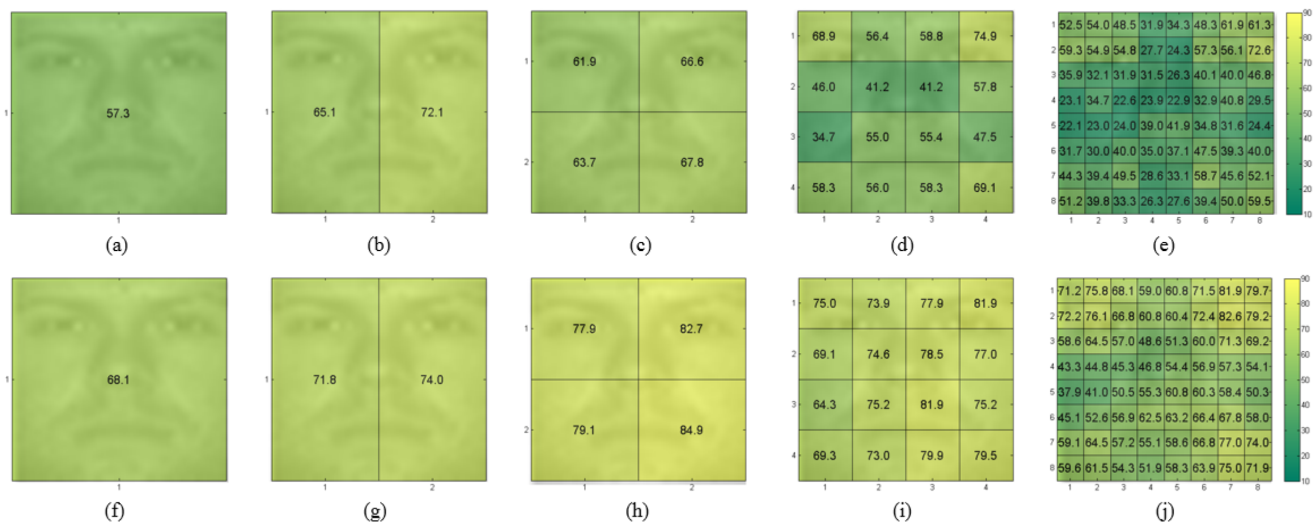

Figure 3: The per patch accuracy comparison between local prediction and global prediction (%). (a) Local level 1. (b) Local level 2. (c) Local level 3. (d) Local level 4. (e) Local level 5. (f) Global level 1. (g) Global level 2. (h) Global level 3. (i) Global level 4. (j) Global level 5. We can observe that, after hierarchical global learning, per patch accuracy is improved significantly on different levels.

Algorithm 1: The HML Framework: Training

Input: Training set $\mathbb{X} = \left\{X^1, X^2, \cdots, X^M\right\}$ and class label set $\mathbb{Y} = \left\{y^1, y^2, \cdots, y^M\right\}$

Output: Local classifier set $\{f_{i,j}(X_{i,j})\}$, global classifier set $\{g_{i,j}(X_{i,j}, H_{i,j})\}$ and final prediction rule $O(\{q_{i,j}\})$

Partition training images hierarchically to $\{X_{i,j}^m\}$

Build local classifier $f_{i,j}(X_{i,j})$ for each patch $X_{i,j}$

for $i \leftarrow 1$ ***to*** D **do**

  for $j \leftarrow 1$ ***to*** N_i **do**

    Construct hierarchical prediction matrix $H_{i,j}$

    Build global classifier $g_{i,j}(X_{i,j}, H_{i,j})$

    Learn global prediction set $\{q_{i,j}\}$

Learn final prediction rule $O(\{q_{i,j}\})$

return $\{\{f_{i,j}(X_{i,j})\}, \{g_{i,j}(X_{i,j}, H_{i,j})\}, O(\{q_{i,j}\})\}$;

Algorithm 2: The HML Framework: Testing

Input: Testing set $\hat{\mathbb{X}} = \left\{\hat{X}^1, \hat{X}^2, \cdots, \hat{X}^K\right\}$, local classifier set $\{f_{i,j}(X_{i,j})\}$, global classifier set $\{g_{i,j}(X_{i,j}, H_{i,j})\}$ and final prediction rule $O(\{q_{i,j}\})$

Output: Predicted label set $\hat{\mathbb{Y}} = \left\{\hat{y}^1, \hat{y}^2, \cdots, \hat{y}^K\right\}$

Partition testing images hierarchically to $\{\hat{X}_{i,j}^k\}$

Compute local prediction $f_{i,j}(\hat{X}_{i,j})$ for each patch $\hat{X}_{i,j}$

for $i \leftarrow 1$ ***to*** D **do**

  for $j \leftarrow 1$ ***to*** N_i **do**

    Construct hierarchical prediction matrix $\hat{H}_{i,j}$

    Learn global prediction set $\{\hat{q}_{i,j}\}$

Apply final prediction rule $O(\{\hat{q}_{i,j}\})$

return $\{\hat{\mathbb{Y}}\}$;

color face images of 126 subjects with real occlusion and different facial expressions. As in [26, 31], we use a subset with both illumination and expression changes that contains 50 male subjects and 50 female subjects. Each subject has 26 images. All the face images are resized to 32×32.

Based on the size of face image, we build a 5-level HML framework for each image; the numbers of patches on each level are: 1×1, 1×2, 2×2, 4×4, 8×8, respectively. According to the best result from [31], CRC is chosen as the local classifier. In CRC, the regularization parameter is set to 0.001. In W-HML, the parameter λ is set to 0.1. In R-HML, the number of trees is set to 100. The baseline methods are MLS and MPCRC. In MLS, the parameter ω is set to 0.1. We use gray values as the original feature. PCA is applied to reduce dimensionality of each patch to 100 (the smaller size patches with less than 100 pixels keep their original features). To be fair for each subject, we first randomly select the greatest common number of images for each subject. Then different proportions of the selected images are used for training and testing. All the experiments were run 10 times. We first test the influence of D with 20% of training data. The performance is depicted in Figures 4

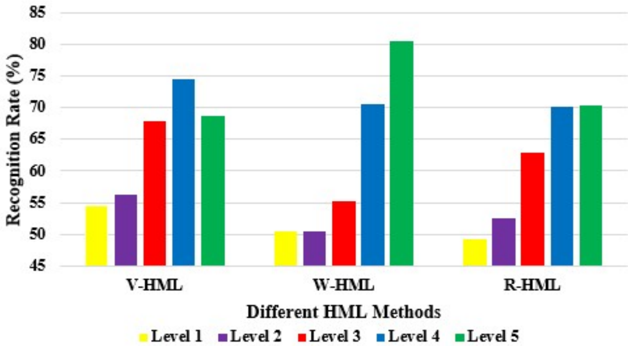

Figure 4: The performance on the Extended Yale B database with different depths of tree-structured HML frameworks.

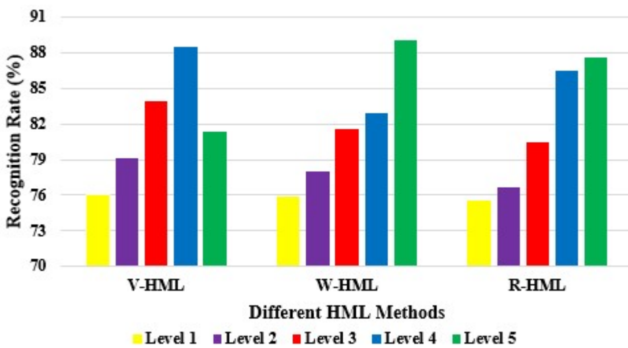

Figure 5: The performance on the AR database with different depths of tree-structured HML frameworks.

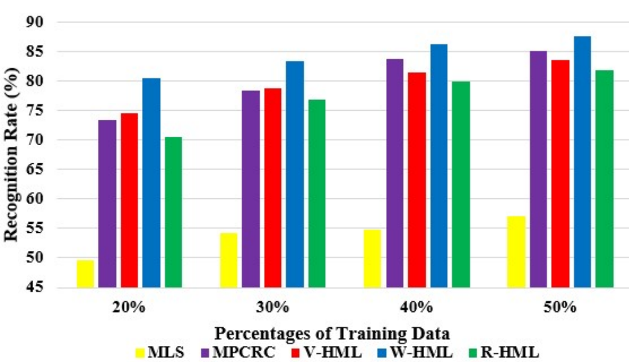

Figure 6: The results on the Extended Yale B database with 25% block occlusions.

and 5. The results of different methods are presented in Figures 6 and 7 (D is set to 4 in V-HML and 5 in W-HML and R-HML).

We can observe in Figures 4 and 5 that different HML methods achieve the best performance at different levels. V-HML performs best with 4-level framework, while W-HML

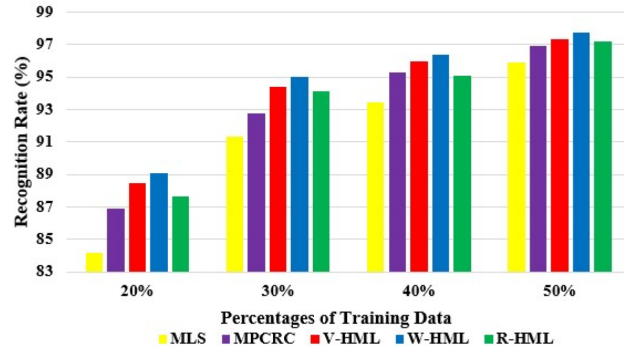

Figure 7: The results on the AR database with real occlusions.

and R-HML perform best with 5-level framework. In Figure 6 we can observe that, in the Yale B database, W-HML performs better than other methods under different percentages of training data. V-HML achieves better or comparable results compared to R-HML, MPCRC and MLS. The results of V-HML under 40 % and 50 % of training data (81.37 %, 83.47%) are a little worse than that of MPCRC (83.81%, 85.03%). In Figure 7, we can observe similar results in the AR database. W-HML achieves the best results under different percentages of training data. V-HML performs better than R-HML and other methods.

4.2. 2D-3D face recognition

We also evaluate the HML framework on the 2D-3D face recognition task with the UHDB11 database [20] and the FRGC v2.0 [16] database, which both contain 3D gallery and 2D probe. In UHDB11, the 3D gallery contains frontal 3D models of 23 subjects while the 2D probe contains 1,602 images from 6 illumination conditions with four yaw rotations and three roll rotations. In FRGC v2.0, we use a subset with 83 subjects in the 3D gallery and 300 images in the 2D probe. Following the pipeline of [30], we apply our method to the unlit texture images (Figure 8). To overcome the limited number of gallery images, we develop an overlapping DAG-structure HML partition: six 100×100 overlapping patches are defined on the most discriminative area of the frontal image, shown in the first image of Figure 8. The child images are obtained by removing one patch from the parent image iteratively. A partition example and the corresponding DAG label hierarchy are shown in Figure 9. In this training-limited 3D-gallery 2D-probe pipeline, only V-MHL is applicable to learn the global prediction. The ensemble score is computed by averaging the local scores that generate consistent results with the global prediction. We compare the results of the 2-level V-HML and the 3-level V-HML with the Albedo Estimation using Lighting Maps (AELM) [30]. The ROC curves of different methods are

Figure 8: The unlit texture image samples from the UHDB11 database.

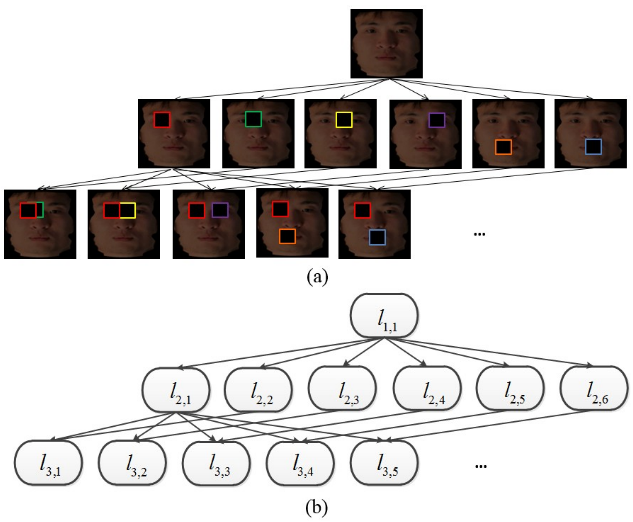

Figure 9: The overlapping DAG-structure HML framework. (a) Hierarchical face partition. (b) Label hierarchy.

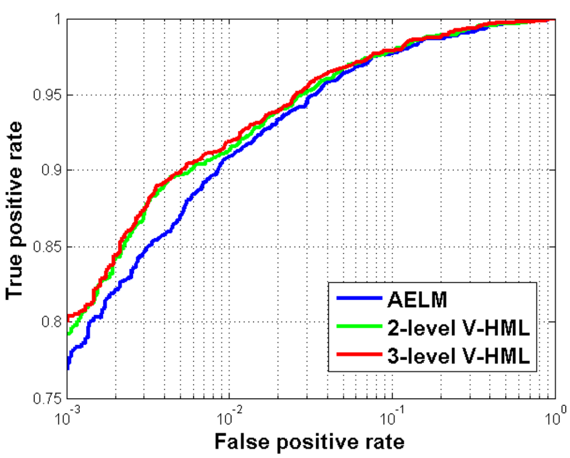

Figure 10: The results on the 2D-3D UHDB11 database.

depicted in Figures 10 and 11.

From Figure 10 and Figure 11, we can observe that the proposed HML methods perform better than the baseline. In UHDB11, the verification rates at 10^{-3} FAR are 76.89% (AELM), 79.20% (2-level V-HML) and 80.01% (3-level V-HML). The 3-level V-HML framework achieves the best result. In FRGC v2.0, the verification rates at 10^{-3} FAR are 37.30% (AELM), 39.00% (2-level V-HML) and 37.67% (3-level V-HML). The 2-level V-HML framework achieves the best result. These improvements can be attributed to two main reasons. First, patch based methods are more robust to the distortions in texture image. Second, hierarchical multi-level learning methods perform better than the original single level method.

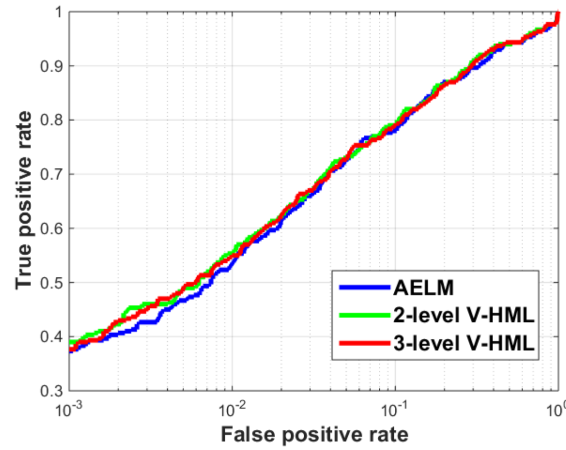

Figure 11: The results on the FRGC v2.0 database.

5. Conclusion

This paper presents a patch based framework for robust face recognition. We build multi-level patches hierarchically and use the hierarchical relationships to improve the local prediction of each patch. The proposed method achieves better results compared to previous methods. Future work will focus on how to design a data-driven HML partition and choose a better global ensemble technique.

Acknowledgements

This research was funded in part by the US Army Research Lab (W911NF-13-1-0127) and the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[2] A. Azeem, M. Sharif, M. Raza, and M. Murtaza. A survey: Face recognition techniques under partial occlusion. *Interna-*

tional Arab Journal of Information Technology, 11(1):1–10, 2014.

[3] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[4] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of SIFT features for face authentication. In *Proc. Computer Vision and Pattern Recognition Workshop*, New York City, NY, June 17-22 2006.

[5] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[6] S. Chen and Y. Zhu. Subpattern-based principle component analysis. *Pattern Recognition*, 37(5):1081–1083, 2004.

[7] T. Fagni and F. Sebastiani. On the selection of negative examples for hierarchical text categorization. In *Proc. Language and Technology Conference*, pages 24–28, Poznań, Poland, Oct. 5-7 2007.

[8] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[9] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: Global versus component-based approach. In *Proc. International Conference on Computer Vision*, volume 2, pages 688–694, Vancouver, British Columbia, Canada, July 7-14 2001.

[10] T. K. Kim, H. Kim, W. Hwang, and J. Kittler. Component-based LDA face description for image retrieval and MPEG-7 standardisation. *Image and Vision Computing*, 23(7):631–642, 2005.

[11] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *Advances in Biometrics*, volume 4642, pages 828–837. 2007.

[12] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. Lu. Person-specific SIFT features for face recognition. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 593–596, Honolulu, HI, Apr. 15-20 2007.

[13] A. Martınez and R. Benavente. The AR face database. *Computer Vision Center, Universitat Autónoma de Barcelona, CVC Technical Report*, 24, 1998.

[14] A. M. Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.

[15] H. J. Oh, K. M. Lee, S. U. Lee, and C. H. Yim. Occlusion invariant face recognition using selective LNMF basis images. In *Proc. Asian Conference on Computer Vision*, pages 120–129, Hyderabad, India, Jan. 13-16 2006.

[16] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):831–846, 2010.

[17] S. R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):660–674, 1991.

[18] C. N. Silla Jr and A. A. Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22(1-2):31–72, 2011.

[19] Y. Su, S. Shan, X. Chen, and W. Gao. Hierarchical ensemble of global and local classifiers for face recognition. *IEEE Transactions on Image Processing*, 18(8):1885–1896, 2009.

[20] G. Toderici, G. Evangelopoulos, T. Fang, T. Theoharis, and I. Kakadiaris. UHDB11 database for 3D-2D face recognition. In *Image and Video Technology*, volume 8333, pages 73–86. 2014.

[21] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[22] G. Valentini. True path rule hierarchical ensembles for genome-wide gene function prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(3):832–847, 2011.

[23] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[24] M. Yang, D. Zhang, and J. Yang. Robust sparse coding for face recognition. In *Proc. Computer Vision and Pattern Recognition*, pages 625–632, Colorado Springs, CO, June 20-25 2011.

[25] J. S. Yuk, K. K. Wong, and R. H. Chung. A multi-level supporting scheme for face recognition under partial occlusions and disguise. In *Proc. Asian Conference on Computer Vision*, pages 690–701, Queenstown, New Zealand, Nov. 8-12 2010.

[26] D. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *Proc. International Conference on Computer Vision*, pages 471–478, Barcelona, Spain, Nov. 6-13 2011.

[27] L. Zhang, S. K. Shah, and I. A. Kakadiaris. Fully associative ensemble learning for hierarchical multi-label classification. In *Proc. British Machine Vision Conference*, Nottingham, UK, Sept. 1-5 2014.

[28] 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. International Conference on Computer Vision*, pages 786–791, Beijing, China, Oct. 17-21 2005.

[29] S. Zhao and Z. Hu. Occluded face recognition based on double layers module sparsity difference. *Advances in Electronics*, 2014, 2014.

[30] X. Zhao, G. Evangelopoulos, D. Chu, S. Shah, and I. Kakadiaris. Minimizing illumination differences for 3D to 2D face recognition using lighting maps. *IEEE Transactions on Cybernetics*, 44(5):725–736, 2014.

[31] P. Zhu, L. Zhang, Q. Hu, and S. C. Shiu. Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In *Proc. European Conference on Computer Vision*, pages 822–835, Florence, Italy, Oct. 7-13 2012.