Face Verification using LBP Feature and Clustering

Chenqi Wang, Kevin Lin and Yi-Ping Hung
National Taiwan University, Taipei, Taiwan

Keywords: Face Recognition, Local Binary Pattern (LBP), Unsupervised Learning.

Abstract: In this paper, we present a mechanism to extract certain special faces—LBP-Faces, which are designed to represent different kinds of faces around the world, and utilize them as the basis to verify other faces. In particular, we show how our idea can integrate with Local Binary Pattern (LBP) and improve its performance. Other than most of the previous LBP-variant approaches, which, no matter try to improve coding mechanism or optimize the neighbourhood sizes, first divide a face into patch-level regions (e.g. 7×7 patches), concatenating histograms calculated in each patch to derive a rather long dimension vector, and then apply PCA to implement dimension reduction, our work use original LBP histograms, trying to retain the major properties such as discriminability and invariance, but in a much bigger component-level region (we divide faces into 7 components). In each component, we cluster LBP descriptors—in the form of histograms to derive N clustering centroids, which we define as LBP-Faces. Then, to any input face, we calculate its similarities with all these N LBP-Faces and use the similarities as final features to verify the face. It looks like we project the faces image into a new feature space—LBP-Faces space. The intuition within it is that when we depict an unknown face, we are prone to use description such as how likely the face's eye or nose is to an known one. Result of our experiment on the Labeled Face in Wild (LFW) database shows that our method outperforms LBP in face verification.

1 INTRODUCTION

Face recognition has been an important issue in the field of pattern recognition. This problem has been addressed on giving two images of face, and verifying that these two images were captured from the same person or different people, so-called the face verification. This task has been widely applied on the intelligent surveillance system and become more and more popular for commercial use. However, most of the face images captured from the surveillance system are not ideal. It still has several challenges, such as the changes of illumination, the variety of head poses, partially occlusion by addressing the accessories and so on. Since the face verification is a binary classification problem, which classifies the given two faces into same or different people, the most important portion become the feature extraction. Carefully design a robust and discriminative feature can improve the performance of face recognition. Thus, the extracted feature—descriptor is required to be not only discriminative but also robust to some noise. Among all existing technologies, local binary pattern(LBP) (Ahonen et al., 2004) has been demonstrated that it can successfully represent the structure of the faces by exploiting the distribution of such pixel neighbourhood. However, those methods suffer from several limitations, such as the fixed quantization and the redundant feature dimension. It has been argued that a large proportion of the 256 codes in original LBP occur with a very low frequency, which may cause the code histogram less informative and more redundant. Thus, a lot of LBP varieties have been proposed to improve the original version.

Figure 1: Several examples face pairs of the same person from the Labeled Faces in the Wild data set. Pairs on the top and bottom are correctly and incorrectly classified with our method respectively.

A famous extension to the original one is known as Uniform LBP (Ojala et al., 2002), where 256 codes

Wang C., Lin K. and Hung Y..
Face Verification using LBP Feature and Clustering.
DOI: 10.5220/0004736905720578
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 572-578
ISBN: 978-989-758-003-1

in LBP reduce to 59 codes by merging 198 non-uniform codes into one bin. The idea stems from the observation that the codes of non-uniform pattern—a binary pattern that contains more than two bitwise transitions from 0 to 1 or vice versa, have very low frequency to emerge. However, this method is entirely based on the empirical statistics, which seems too heuristic and has no cogent theory to explain the reasonableness of mapping each non-uniform code into a single bin. Additionally, Cao et al. (Cao et al., 2010) also testify that even in Uniform LBP, some codes still appear rarely in real-life face images. On the other hand, when using LBP, most works firstly divide a face image into dozens patches and calculate LBP histograms in each patch before concatenating all the obtained histograms, which lead to an oversize-dimension feature vector. To reduce the feature size, dimension reduction technique such as principle component analysis(PCA) should be used to avoid computation load and over-fitting. Despite preserving the most energy and the largest variation after projection, PCA may ignore some key discriminative factors.

Instead, our work look at LBP in a totally different way—from the histogram's point of view. That is, we preserve the intact histograms of original LBP and try to find out if there exist some unforeseen yet useful relations and laws within them. In particular, while the circular neighbouring pixels can be clustered into fixed groups in LBP(e.g. if circle size is 8, we can get 2^8 codes), we want to know whether histograms can also be clustered. The intuition is based on the following observation, that is, when we depict a person, we prefer to use the description such as 'he has a hooked nose' or 'she has a big eye'. Therefore, considering LBP histograms are proven to be strong descriptors for face recognition, we divide a image into 7 component, namely eyes, nose, mouth and so on, and use LBP histogram in each component to represent its character. Then, we use unsupervised learning on component basis to get N clustering centroids—called LBP-Faces, in each component, which we believe can represent N different kinds of component around the world if we collect enough data for learning. Finally, we use the dissimilarities between these LBP-Faces and each input face image to verify the people. Figure 1 shows some sample face pairs used for evaluation in our work. The top pairs are correctly classified while the bottom ones not. But we can see those pairs that we didn't make right decision are really difficult to be distinguished, even by human perceptions. Section 4 illustrates the details of the dataset—LFW dataset we used in the work. Result on the dataset shows we have better performance over LBP. The contribution of this paper includes:

- This paper proposes an innovative idea that inspired by the natural recognition procedure of the human beings to improve the performance of LBP in face verification.
- To our best knowledge, we are the first to use unsupervised learning in the histogram's point of view, which may give a new thought in face verification and related recognition field.
- Our result on the restricted data set of LFW—a challenging and authoritative dataset, outperforms several state-of-the-art face verification method, which prove the rationality and feasibility of our idea. We believe not only LBP histograms can be applied on our framework, but some other descriptors may also be applicable.

The rest of this paper is organized as follows: Section 2 discusses the related work in face recognition. We then describes overview of our proposed method in Section 3. In Section 4 we introduces the dataset used in this paper. Section 5 elaborates on how to derive LBP-Faces and how to use these LBP-Faces to verify people. Our experiment and results are shown in Section 6 and we conclude our work in Section 7.

2 RELATED WORK

There are a lot of existing approaches to extract features in face recognition. Turk et al. (Turk and Pentland, 1991) proposed descriptor called EIGENFACE, where each images is presented as an n-dimension vector. The input face is projected into the weight space and the nearest-neighbor method is performed to find the best matched face in the database. FISHERFACE (Belhumeur et al., 1997) is an alternative method, whose idea is to project the image into a subspace in the manner which discounts those regions of the face with large deviation. Other famous descriptors include discrete cosine transform (DCT) (Rod et al., 2000) and Gabor (Liu and Wechsler, 2002). In addition, Ahonen et al. (Ahonen et al., 2004) proposed the Local Binary Pattern (LBP) to represent features of face and get a reasonable performance. In detail, a face image is divided into several regions, and in each region we can derive the LBP histogram. Finally, the face descriptor is completed by concatenating all the histograms into an enhanced feature vector. However, G. Sharma et al. (Sharma et al., 2012) pointed out that LBP still has several nonnegligible limitations, such as the baseless heuristic encoding mechanisms to reduce the dimension of feature space, the hard quantization of the feature space, and the histogram-based feature representation which

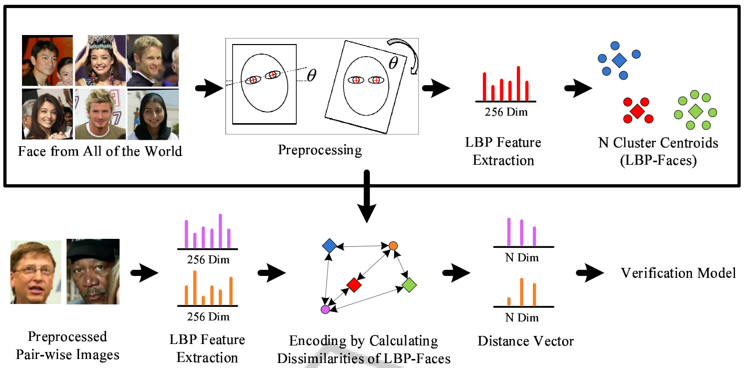


Figure 2: Pipeline of the proposed unsupervised LBP-Face algorithm.

impeded us to apply higher-order statistics.

Uniform LBP (Ojala et al., 2002) is a common extension of original one, but Zhimin Cao et al. (Cao et al., 2010) also argued that even in Uniform LBP, there are a lot of codes which may rarely appear in real-life face images. Several methods have been proposed to tackled this problem. Zhimin Cao et al. introduced the Learning-based (LE) encoder which projects the ring-based sampled pattern into another refined feature space. The encoder is trained using a set of training face images with unsupervised algorithm, such as K-means and random-projection tree. G. Sharma et al. (Sharma et al., 2012) further proposed the Local Higher-order Statistic (LHS) model, which improve the LBP feature by modifying the LBP feature extraction procedure. They described the local pattern of pixel neighbourhoods with their proposed differential vectors, which records the difference value between the center of the LBP operator and the circular pixel neighbourhoods. The Gaussian Mixture Model is then applied to describe all of the differential vectors. Other famous feature-based methods include local texture pattern(LTP) (Tan and Triggs, 2010).

Some of the works (Ahonen et al., 2004; Turk and Pentland, 1991; Belhumeur et al., 1997) mentioned above focus on how to extract the most discriminative feature for face recognition. In another aspect, several previous works (Rod et al., 2000; Cao et al., 2010) tried to reduce the feature dimension by projecting the feature into the refined new feature space. We can easily find that there is a trade-off between the dimension of the feature space and the representative ability of the feature. The goal of the feature dimension reduction is to reduce the dimension of the feature space and decrease the complexity of the computation. However, the more dimension of the feature we reduce, the more useful information we might lost. Since there is no theory that can prove which codes within the histogram are useless for sure, in our work, we preserve all the histogram codes of LBP to retain as much representative ability of the initial feature as possible. Simultaneously, we project the features into a low-dimension feature space —LBP-Face space red by using unsupervised learning to fulfil dimension reduction.

3 OVERVIEW OF FRAMEWORK

In this paper, we propose a novel idea which is inspired by the natural recognition procedure of the human beings, that is, we usually depict a person by describing how likely he/she is to an known person. Therefore, we present a mechanism by red finding out these ‘well-known’ faces—which we call LBP-Faces, calculating the dissimilarities between given image face and all these LBP-Faces, and utilizing the new distance vectors to verify other unknown faces.

Figure 2 demonstrates the pipeline of our approach. There are following main steps in our work.

- Firstly, we need to derive LBP-Faces, which is the most important process in our work since it results—LBP-Faces would directly influence our final performance. The upper frame box shows this process. Specifically, 13232 face images from all of world are implemented preprocessing before extracting their LBP histograms, then we use unsupervised learning technology to obtain final N centroids, which are denoted as diamond icons in Figure 2 and can be seen as the most representative N different faces in the word. The N centroids are exactly what we call LBP-Faces. Additionally, in order to preserve local information, all the process is carried out in component-level(there are 7 components in each face). For easy description, we still use LBP-Faces to denote centroids in each component.
- After acquiring the LBP-Faces, we are ready to

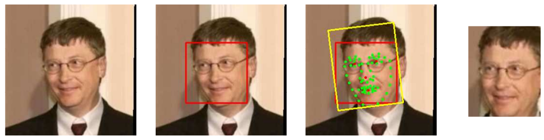
do face verification. Thus, the second step is to obtain final feature for face matching. To any given face image, we calculate the dissimilarities between it and all the LBP-Faces to get an N-dimension distance vector—this vector is our final feature vector.

- Thus, in face matching stage, we just need to compare these distance vectors among each image. The smaller the distance of two distance vectors is, the more chance the two face images are from the same person. Particularly, since the feature dimension reduction is simultaneously conducted within the above process, we don't need do dimension reduction such as PCA like traditional LBP-variants do, which also makes our method competitive in computation efficiency.

4 DATA SET

Before going further, it is essential to elaborate on the data set we use in the paper. Since the derivation of LBP-Faces and our experiment are all based on the dataset, we use a separate section to illustrate it—the publicly available dataset Labeled Faces in the Wild (Huang et al., 2007). Figure 1 in the first page shows some samples of the dataset, which are used in our experiment in Section 6. This well-known dataset is specifically designed for the study of face recognition. This database contains 13233 face images and 5749 people collected form the web. There are 1680 people have two or more face images for the purpose of verification. All the face images are labeled with the name of the person. Since the LFW data is collected from the Internet, the images are not always ideal and become more challenge for face recognition. That is, there is a large variety in head poses, different illumination situations, and so on. In order to eliminate the avoidable variation in the LFW original version images, the stage of image preprocessing is performed and is described in detail in subsection 5.1.

In general, the LFW dataset can be separated into two parts. The first part—View 1 is designed for algorithm development, such as model selection or validation. The other part—View 2 is provided for the performance comparison, which is used for the final evaluation of a proposed algorithm and comparison of performance of different algorithms. For View 1, there are 1100 matched pairs and 1100 mismatched pairs of face images, which can be seen as the positive and negative samples for the stage of model training. As the same setting, there are 500 pairs of matched and mismatched pairs images in View 1, which is given for the stage of validation. In the configuration of View 2, the database is split into 10 sets randomly. There are 300 matched pairs and 300 mismatched pairs images randomly selected within each set. The final evaluation of performance comparison is given using the 10 folder cross validation, and can be found in Section 6.



(a) (b) (c) (d)

Figure 3: (a) input image, (b) face detection result, (c) face alignment using AAM, and (d) Preprocessing result.

The evaluation results are plotted in the standard ROC curve. The ROC curve is constructed by plotting the fraction of True Positive Rate (TPR) and False Positive Rate (FPR), which illustrates the characteristic of the binary classifier at various threshold setting. In Section 6, we compare our proposed method to some state-of-the-art recognition algorithms. The ROC curves of state-of-the-art are available from LFW.

5 METHODOLOGY

In this section, we will elaborate on the entire process of our work. We firstly describe some necessary preprocessing before getting LBP-Faces in subsection 5.1. Then subsection 5.2 depict in detail how to derive LBP-Faces. The comparison between two faces images is described in subsection 5.3.

5.1 Preprocessing

First of all, the haar feature-based cascade classifier proposed by Viola and Jones (Viola and Jones, 2001) is applied to detect human faces in the given image, and extract the region of interest for further processing. Before feature extraction, we need to do some preprocessing to eliminate avoidable variations and noise such as illumination, face rotation, and so on. As to face rotation, since the face image may have different poses, e.g. the frontal face or left face, we will perform the component alignment by using Active Appearance Model (AAM) (Cootes et al., 2001), which is proven to be a reliable method in face alignment. Then we can resolve the face rotation based on the line of two detected eyes like Fig 3 shows.

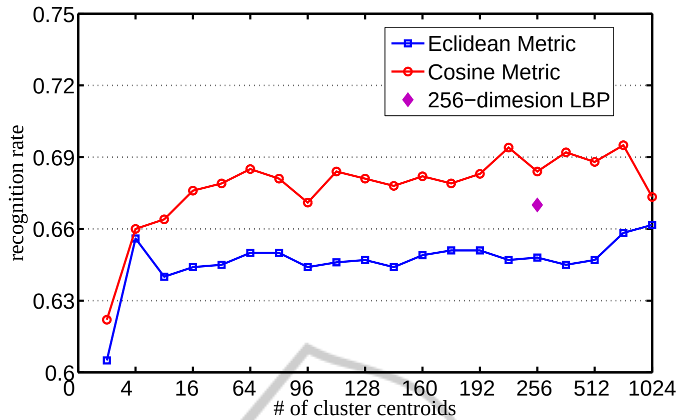


Figure 4: Performance under different cluster numbers with Euclidean Metric and Cosine Metric.

5.2 Derive LBP-Faces

Our approach is mainly combined with two methods, LBP and learning-based descriptor. LBP has been proven to be a both computation efficient and discriminative descriptor. However, it has been argued about the rationality of its too manual encoding mechanism and the quite low frequent emergence of a large proportion of codes. We propose to use learning method to tackle the problem. Unlike most previous works trying to use learning method in encoding, we think in a totally different way—we're learning in histogram's point of view. Particularly, we enlarge the region where a histogram is obtained. The intuition about this is we think in a larger area, the histogram has stronger anti-interference ability against the noise. In other word, the histograms of images from the same people have more chance to be similar. Therefore, we can group those similar histograms together and use a most representative one to stand for this cluster—a group of similar faces. These representatives are what we define LBP-Faces. We use unsupervised learning techniques to find out these LBP-Faces. K-means is a classic method for exploring clusters by choosing K centroids to minimize the total distance between them and their nearest neighbourhoods, whose theory is in accordance with our idea, so we use it as our default unsupervised learning algorithm. The centroids are exactly the LBP-Faces we want to derive.

In unsupervised learning, the choice of the number of centroids is very important. So we vary the value of cluster number from 2 to 1024, using default parameters and initial vectors in K-means, to compare the recognition rate(Face matching will be described in next subsection). Also, we compare two common distance metric, namely Euclidean and Cosine distance in unsupervised learning stage. Figure 4 shows the recognition rate under different cluster numbers using the two distance metric respectively. From the figure, we can notice that the cosine distance metric outperforms the euclidean one, which corresponds to the point of view proposed by (Nguyen and Bai, 2011). Thus, we set it as our default distance metric. It's worth noting that the accuracy doesn't have swift growth as the centroids' number increases—when number greater than 16, the curve fluctuate at a relatively stable level. Therefore, considering efficiency and accuracy, we choose 200 as our default number of clustering. Moreover, the purple diamond represents the accuracy of LBP in the same context, that is, directly extract LBP histograms in the each component and then concatenate the 7 histograms to form a $1792(7 \times 256)$-dimension vector for face matching. We use it as a benchmark to evaluate the power of our method. Under the same condition, our initial result achieves higher performance than LBP, which can prove the reasonableness of our approach.

5.3 Face Matching using LBP-Faces

After deriving LBP-Faces, we can do face verification. Figure 5 shows the process of face matching. When a pair of face are input, we first divide them into 7 components. It should be necessary to advert that the histogram icons in the Figure 5 does not represent LBP histograms but the distances between the component's histograms and the N LBP-Faces. So each histogram icon denotes a N-dimension distance vector, which can stand for how similar the component is to the N LBP-Faces. In face matching, we simply calculate the distance between the two faces in component-wise. Certainly, supervised learning can be utilised here to optimize the weight of Distance 1 to 7 to boost the performance, but for computation efficiency and precise evaluation of our features, we

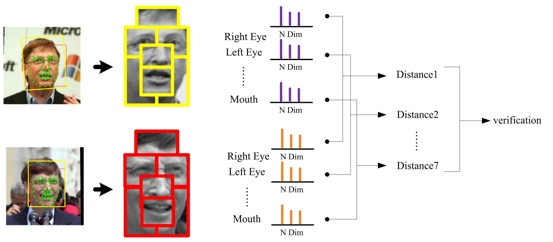


Figure 5: Process of face matching.

Table 1: Performance comparison in different parameters of initial vector ν and distance metric ξ.

ν \ ξ	Euclidean	L1	cosine	correlation
sample	0.647	0.653	0.682	0.642
uniform	0.635	0.678	0.662	0.674
cluster	**0.658**	**0.669**	**0.713**	**0.685**

just sum them and use a optimal threshold in training data to predict the testing data. Actually, all the experiments in this paper are conducted in this process.

Moreover, we find out that optimizing parameters in clustering can also help improve the performance. As we all know, the initial vectors in K-means impact much on the total accuracy, so we compare the three common initialization mechanisms, namely, 'sample', 'uniform' and 'cluster'. While 'sample' and 'uniform' simply select initial vectors at random and uniformly respectively, 'cluster' perform a preliminary clustering phase on a random 10% subset of the whole data set to derive the initial vectors. In addition, as Figure 4 shows that the distance metric also influence the result, and in K-means algorithm, the distance metric is extremely important. So we also compare the four common distance metrics—Euclidean distance, Cosine distance, cityblock(i.e. L1) distance and correlation distance. All the results are shown in Table 1, from which, we can find that 'cluster' method in initialising vectors and cosine similarity metric in K-means outperform the other parameters. And the highest recognition rate is 71.3%. Since the chosen data set is really challenging, the result is quite satisfactory.

6 EXPERIMENT AND RESULTS

In this section, we illustrate our experiment and report the final face recognition results on LFW benchmark. We use View 2 in LFW dataset for evaluation. This data is provided for researchers to evaluate and compare the performance in face recognition. Although the dataset is fixed, its reasonable construction by randomly sampling data from a much larger dataset makes it convinced enough for correct evaluation. To avoid over-fitting by using machine learning mechanisms with parameters selection and evaluate the strength and rationality of our proposed features, we just calculate the dissimilarity of the distance vectors of each pair images as what described in subsection 5.3. In addition, as subsection 5.3 reveals, we choose cosine distance measures and 'cluster' as default parameter of distance measures and initial vector respectively in unsupervised learning to derive the final LBP-Faces.

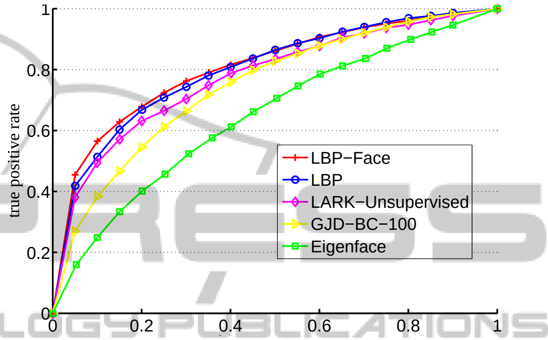


Figure 6: Face recognition comparison on the restricted LFW benchmark.

Figure 6 shows the ROC curve and the point on the curve of our result-the red one, represents the average over the 10 folds of false positive rate and true positive rate for a fixed threshold. The other 4 curves are chosen for comparison. We choose Eigenface because it has a similar idea with us. We both try to project a face image into another feature space according to some projection basis, but use totally different method. And LBP are classical descriptors and our method are based on it, thus included. Additionally, since our model use unsupervised method, we should choose some unsupervised results to compare. GJD-BC-100 and LARK unsupervised (Verschae et al., 2008) are two unsupervised descriptors enumerated in LFW, which can be benchmarks to evaluate our work. From the figure, we can see our method have much better performance over the two unsupervised method and Eigenface while outperform the LBP a little. Specifically, we achieve 72.3% accuracy, a little higher than 71.4% of LBP but much higher than 60.0% of Eigenface. Although our result has not extremely exciting, it has proven the value and rationality of our novel idea.

7 CONCLUSIONS AND FUTURE WORK

This work constructs the computation efficient face verification mechanism. Without any complex machine learning approach, the designed mechanism verifies people using only the LBP-Faces and its similarity distances. Although the stage of unsupervised clustering need to take a period of time to process, the new feature can immediately generated by the LBP-Faces in the stage of testing. The new feature is calculated only by the similarity distance, which means the computation complexity is very low and can speed up the verification procedure. Our proposed method can automatically derive the LBP-Faces and verifying people in nearly real-time, which is applicable in intelligent mobile phone and embedded system design. Experimental results show that our method can achieve higher recognition accuracy than that of the LBP and Eigenface in the Labeled Faces in the Wild (LFW). Even though the recognition of our method might be less promising when the face is partially occluded or the head pose is severely varied, we believe that the improvement can be achieved by utilizing the 3D information to enhance the LBP-Faces and clustering the LBP-Faces into different poses. In conclusion, this work is a good initial start, which prove the reasonableness of our novel idea in face verification and still have a long way to go for future stronger work.

In this work, we just choose K-means as default unsupervised learning algorithm, so in the future, we will firstly attempt on more clustering mechanisms. In addition, we will take 3D scenario into consideration in unsupervised learning stage, that is, we will derive LBP-Faces according to different poses to improve the accuracy. Moreover, the data set—for example, the size and the sampling images, used in learning stage can be also further researched on.

ACKNOWLEDGEMENTS

The work was supported in part by the Image and Vision Lab at National Taiwan University. The author was supported by MediaTek Fellowship.

REFERENCES

Ahonen, T., Hadid, A., and Pietikinen, M. (2004). Face recognition with local binary patterns. In *Proc.7th European Conference on Computer Vision(ECCV)*, pages 469–481.

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

Cao, Z., Yin, Q., Tang, X., and Sun, J. (2010). Face recognition with learning-based descriptor. In *Proc. 23th IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, pages 2707–2714.

Cootes, T., Edwards, G., and Taylor, C. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 23(6):681–685.

Huang, G. B., Mattar, M., Berg, T., and Learned-miller, E. (2007). E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report.

Liu, C. and Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4):467–476.

Nguyen, H. V. and Bai, L. (2011). Cosine similarity metric learning for face verification. In *Proc. 10th Asian conference on Computer vision (ACCV)*, ACCV'10, pages 709–720, Berlin, Heidelberg. Springer-Verlag.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.

Rod, Z. P., Adams, R., and Bolouri, H. (2000). Dimensionality reduction of face images using discrete cosine transforms for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Sharma, G., Hussain, S., and Jurie, F. (2012). Local higher-order statistics (lhs) for texture categorization and facial analysis. In *Proc.15th European Conference on Computer Vision (ECCV)*, pages 1–12. Springer Berlin Heidelberg.

Tan, X. and Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6):1635–1650.

Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586–591.

Verschae, R., Ruiz-Del-Solar, J., and Correa, M. (2008). Face Recognition in Unconstrained Environments: A Comparative Study. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France. Erik Learned-Miller and Andras Ferencz and Frédéric Jurie.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc. 14th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–511–I–518 vol.1.