

Eye-Strip based Person Identification based on Non-Subsampled Contourlet Transform

Hemprasad Y. Patil

Dept. of ECE

Visvesvaraya National Institute
of Technology, Nagpur, India

Ashwin G. Kothari

Dept. of ECE

Visvesvaraya National Institute
of Technology, Nagpur, India

Kishor M. Bhurchandi

Dept. of ECE

Visvesvaraya National Institute
of Technology, Nagpur, India

ABSTRACT

Many state-of-the-art face recognition systems fail to identify a person when most portions of the face are occluded. This paper addresses an intriguing problem of face recognition only with eye-strip samples as testing images and full images or again eye-strips as database images. Non-sub-sampled Contourlet transform is a distinguished algorithm for extracting soft and smooth contour-like edges without any loss of information. It also produces eminent features due to its localization and directionality preserving abilities and has strong resemblance with abilities of human visual cortex to extract features. This does not require any boosting of sub-band coefficients. We have proposed a novel approach that adds all the sub-bands at each pyramidal level of non-subsampled contourlet transform to achieve a hybrid high frequency composite sub-band and minimize the dimensionality followed by feature extraction using Weber Local Descriptor (WLD) on the hybrid high frequency sub-band. Linear Discriminant Analysis is further used for insignificant feature reduction followed by nearest neighbor matching using Euclidean distance measure. JAFFE, Yale and Faces94 Essex university databases are used for experimentation and benchmarking. The analysis indicates that the proposed approach yields a robust feature vector and very good recognition rates using only eye-strip features.

General Terms

Biometrics, Person Identification, Pattern Recognition

Keywords

Eye-Strip, Biometrics, Face Recognition, NSCT, WLD, Wavelets.

1. INTRODUCTION

Automatic face recognition is a popular research area and has attracted many researchers since last three decades [1], [2], [3]. Face recognition has been explored by pattern recognition scientists, neuroscientists, anthropologists, physiologists and psychologists. Face recognition is an instinctive, non-intrusive and contact free biometric method. Face recognition is also a special case of pattern recognition where the classifier has to distinguish between the multiple identities when drastic inter and intra class feature variations are subjected for classification. The robust feature recognition and extraction plays very important role in the overall recognition process. Facial cues are the most prominent features used by human brain to distinguish between persons [4]. Many face recognition algorithms are deployed as commercial biometric authentication systems under constrained conditions across the world. Unfortunately, the recognition accuracy of most of the state-of-the-art face recognition systems affects due to illumination, pose, expression, rotation, ageing, makeup and

occlusion related variations. To assess the accuracy of commercial face recognition systems various tests have been conducted by biometric agencies. Face recognition vendor test (FRVT)-2000 indicated the need to improve algorithms on the count of illumination and scale variations. FRVT-2002 report suggested that outdoor face recognition systems needed to be improved in terms of recognition performance. FRVT-2006 experimentation concluded that machine based face recognition is more robust than humans under limited illumination variations [5]. An ideal face recognition system should have high false rejection rate (FRR) and low false acceptance rate (FAR). Equal error rate (EER) is a point at which both FAR and FRR values are same is a single valued measure of face recognition accuracy [4]. The True Acceptance rate is expected to be 100%.

Although most of the research directions were focused on person identification with whole face, a few approaches have addressed the problem of person identification when most parts of the face are occluded. Face recognition from occluded images is a challenging task and requires highly discriminative feature extraction methods to achieve successful biometric identification. Occlusions due to scarf, sunglasses and hair falling on the face are commonly addressed problems in the existing literature [6], [7], [8], and [9]. These approaches consider the commonly occurring occlusions. However, when only the eyes portion of the face is available, the research problem becomes more complex.

In this research, person identification when most of the facial area is masked and only eye portion is available is addressed under constrained conditions. For example, professionals like doctors, biotechnologists, chemists and nurses etc. have to hide their mouth area and hair with a mask during day-to-day practices. Proposed approach processes an eye strip image using non-subsampled contourlet transform (NSCT) based features and subsequently extracts the features from frequency domain sub-bands using Weber local descriptor (WLD). The rest of the paper is organized as follows. Section 2 discusses the prominent related works. Section 3 describes non-subsampled contourlet transform and Weber local descriptor. Section 4 presents the proposed approach. Section 5 includes the results and discussions followed by conclusion.

2. RELATED WORK

Since the proposed method deals with occlusion based facial comparison using automatic face recognition algorithms, we primarily focus on previous methods that address occlusion invariant face recognition.

Huang and Yang [7] proposed an approach to address partial occlusions based on grammatical facial models. A grammatical face model represents the whole facial structure by chin, nose, eyes, forehead and mouth. They constructed

separate models for sub-faces and occlusion variations which is a computationally very heavy approach. Li et al. [8] explored the sparse representation based approach for occlusion invariant face recognition. They have proposed graph based morphological model to illustrate the morphological arrangement of error. Their approach is based upon a non-linear similarity measure called correntropy based metric that follows exponential distribution for occlusion error measurement. This technique depends upon shape of the occlusion which is an uncertain parameter. Gao et al. [10] proposed an algorithm which subtracts the mean face from test face to obtain occluded regions. They have utilized sparse coding based classification technique for data segregation. However, mean face subtraction is a rudimentary technique and fails especially when the database is large in size. Amutha [11] presented color co-occurrence matrix (CCM) based algorithm. The principal component analysis is employed for dimensionality reduction on CCM classified faces. But, the availability of color images is not possible under every capturing environment. De Smet et al. [12] explored the use of 3-D morphable face model to estimate the facial parameters. They have utilized visibility map which is embedded into statistical facial image formation model as a collection of random variables that are correlated in spatial domain. Their approach depends upon computationally heavy 3-D morphable model based generalized expectation maximization algorithm for occlusion invariant face recognition. Guanglu et al. [13] demonstrated use of modular representation based technique to address face recognition with occlusion. Block distance based modular linear regression classification method has been improved to efficiently fuse the features provided by various blocks in the spatial domain image. The above methods address occlusion variations by considering whole facial region. The person identification issues while eyes' region act as a query and also provided as a set of training images have been rarely addressed in the literature so far. Chuan Chin et al. [14] presented an analysis of eye based person identification using non-negative matrix factorization (NMF), spatially confined NMF and local NMF. They have concluded that human eyes have ability to differentiate between distinct persons. The major lacunae of NMF is that it does not consider the geometric structure of input data [15]. However, geometric structure has a prime importance in case of facial components which are curved in nature.

3. MATERIALS AND METHODS

This section briefs upon the non-subsampled contourlet transform and Weber local descriptor which are further incorporated in the proposed algorithm to obtain person identification using only eye strips or cropped eyes' region of human faces.

3.1 Non-Subsampled Contourlet Transform (NSCT)

The non-subsampled contourlet transform is a combination of non-subsampled pyramid transform and oriented directional filter banks as proposed by Cunha et al. [16]. The NSCT sub-bands are fully shift-invariant, direction oriented and pyramidal multi-scale image decompositions. The redundancy in frequency domain transforms facilitates to obtain enhanced basis functions without much loss of information. These basis functions are able to capture the signal/image behavior more efficiently [16]. Wavelet like redundant transforms are more extensible, smooth to design and implement. NSCT is one of those redundant transforms which is effectively used for

applications like image denoising [17], illumination invariant face recognition [18], medical image fusion [19] etc. Fig.1. indicates an outline of non-subsampled contourlet transform. Procedure to obtain forward decomposition of an input image using NSCT consists of two stages namely, non-subsampled pyramid transform and multidirectional non-subsampled filter banks [16].

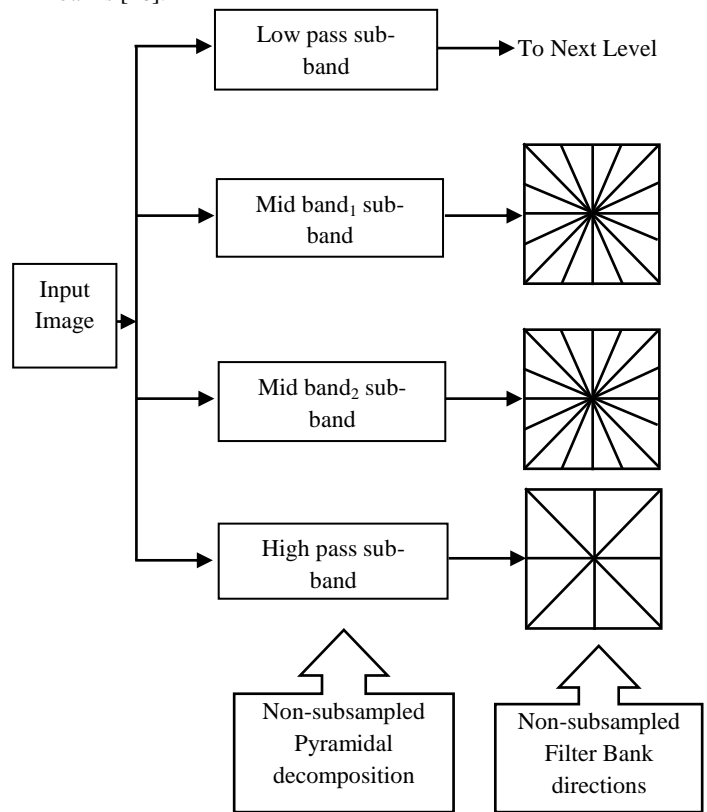


Fig 1: Outline of non-subsampled contourlet transform [16]

3.1.1 Non-Subsampled Pyramid Transform

Non-subsampled 2-channel filter banks are utilized to obtain sub-band decomposition of an input image. The non-subsampled filter bank is constructed in such a way that $P_L(z) = 1 - P_H(z)$ where, $P_L(z)$ is a low pass filter and $P_H(z)$ is a high pass filter. This construction of non-subsampled pyramids is a generalized case of Laplacian pyramid transform [20], [21].

3.1.2 Multidirectional Non-Subsampled Filter Banks

Bamberger and Smith [22] proposed the directional filter bank (DFB) to obtain wedge like tree structured frequency partitioning with the help of resampling matrix structures and oriented fan filters. Cunha et al. [16] modified the original structure of the DFB by removing up-samplers and down-samplers to obtain the sub-bands of the size similar to an input image. Two channel non-subsampled filter banks employed in the construction of non-subsampled contourlet transforms follow the Bezout relation as in (1) [16].

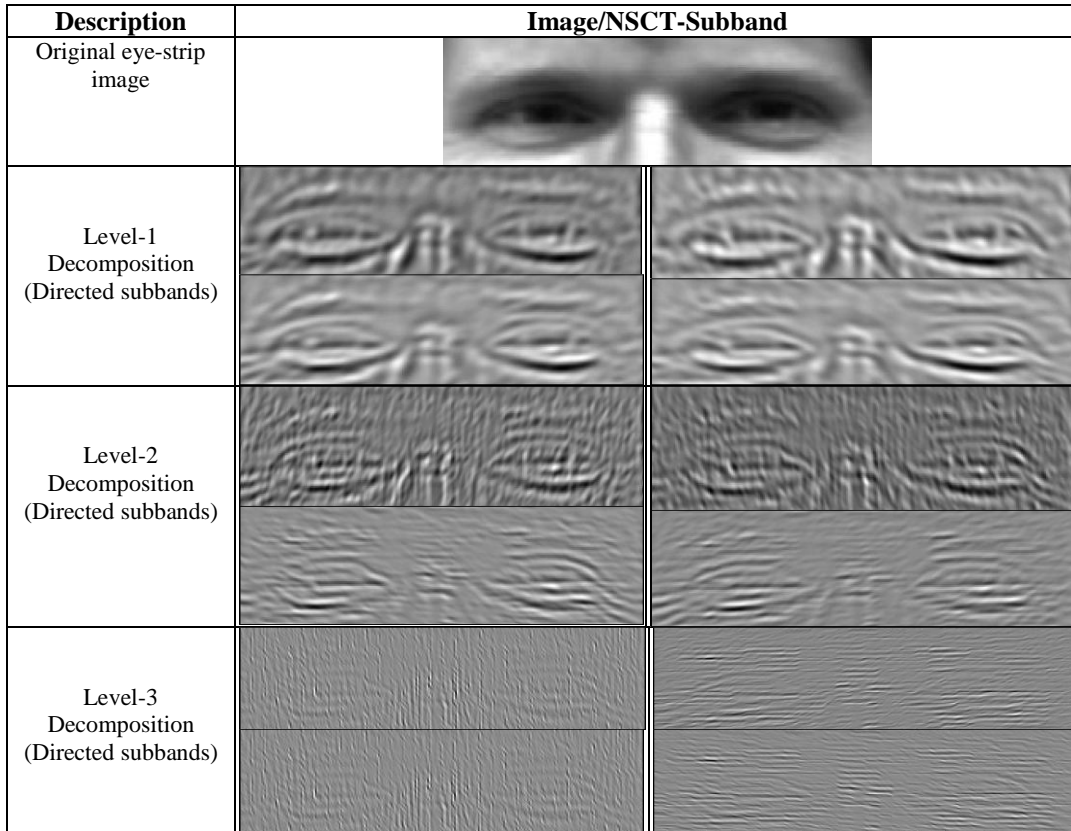


Fig 2: Forward decomposition using NSCT at scale (2,2,2)

$$P_L(z)Q_L(z) + P_H(z)Q_H(z) = 1 \quad (1)$$

Frame structure of NSCT satisfies the tight bounds criteria which depends upon constants X and Y as in (2) [16].

$$X\|i\|^2 \leq \sum_{b \in \Gamma} |\langle i, \mathcal{G}_b \rangle|^2 \leq Y\|i\|^2 \quad (2)$$

3.2 Weber Local Descriptor

Human eye's ability to sense a visual pattern depends upon original illumination intensity on the pattern as well as just noticeable changes in the reflected light source [23]. This inference is based on Weber's law and depends upon two components namely, differential excitation and orientation. The methods of facial texture classification are broadly classified into sparse descriptors like SIFT [24] and dense descriptors like LBP [25] and WLD [23]. Sparse descriptor initially locates the interest pixels in the original image and then forms a patch of appropriate size to extract meaningful features. Whereas, dense descriptor operates on each pixel and forms a discriminant feature vector. Weber local descriptor is applied on whole input image or input patch treated as an image to form a feature vector. Weber's law states that the ratio of variation in just noticeable intensity to the original intensity is constant. The gradient and its corresponding orientation assimilates the advantages of SIFT transform and local pixel level feature differential extraction gives eminent features as in LBP. Finally, a WLD histogram is computed using differential excitation and orientation components. To compute differential excitation, directional filters are employed which are shown in Fig.3. Differential excitation is computed in order to resemble the pattern perception by

human eyes. Input image I is convolved with $FILT_{00}$ and $FILT_{01}$ to obtain Λ^{00} and Λ^{01} respectively. The differential excitation (D) is given by (3) and (4).

$FILT_{00}$			$FILT_{01}$		
+1	+1	+1	0	0	0
+1	-8	+1	0	+1	0
+1	+1	+1	0	0	0

Fig 3: Directional filter masks $FILT_{00}$ and $FILT_{01}$ [23].

$$D(X_m) = \tan^{-1} \left[\frac{\Lambda^{00}}{\Lambda^{01}} \right] \quad (3)$$

$$D(X_m) = \tan^{-1} \left[\frac{\sum_{p=0}^{P-1} X_p - X_m}{X_m} \right] \quad (4)$$

Where X_m stands for central pixel in the 3x3 mask and X_p represents neighboring pixels. The orientation parameter (θ) is obtained upon filtering the input image with filters $FILT_{10}$ and $FILT_{11}$ as indicated by Fig.4. Input image I is convolved with $FILT_{10}$ and $FILT_{11}$ to obtain Λ^{10} and Λ^{11} respectively. The orientation (ϖ) is given by (5) [23].

0	-1	0
0	0	0
0	+1	0

0	0	0
+1	0	-1
0	0	0

Fig 4: Orientation filter masks $FILT_{10}$ and $FILT_{11}$

$$\varpi = \tan^{-1}\left(\frac{\Lambda^{11}}{\Lambda^{10}}\right) \quad (5)$$

Overall Weber local histogram descriptor is obtained by 2-D to 1-D concatenation of directional and orientation filtering result histograms.

4. PROPOSED APPROACH

We propose a system to acquire discriminant features for eye strip based person identification by using NSCT and WLD. Fig.5. demonstrates the overview of the proposed feature generation scheme for eye-strip based person identification.

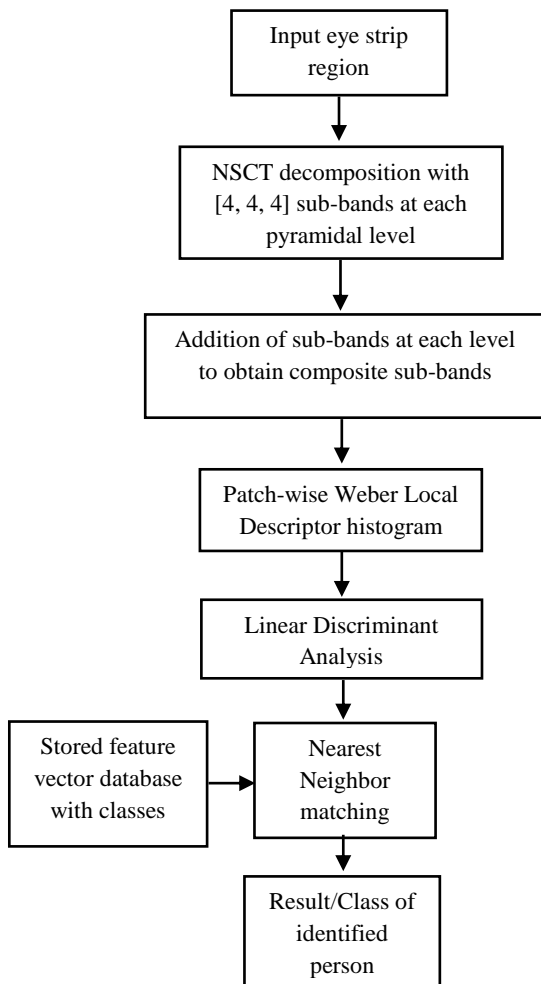


Fig 5: Proposed system for eye-strip based person identification.

For all non-subsampled contourlet sub-band images, patch-wise WLD histogram is computed which forms a transform domain feature descriptor. At first, the original eye-strip image is resized to 76 x 256 pixels. Subsequently, non-subsampled contourlet transform decomposition is computed at three pyramidal decomposition levels. Resultant forward decomposition after NS-DFB filtering at levels $S_d = \{2,2,2\}$.

This signifies $2^{\wedge}S_d = [4,4,4]$ directional sub-bands at each corresponding pyramidal level. Output of every NS-DFB filter is a non-subsampled contourlet coefficient matrix. At each pyramidal level, similar frequency band components are obtained in four directions. They are added to form one composite sub-band at each level. Thus, we obtained one low-pass, two composite band pass and one composite high pass sub-bands for the three levels. This greatly helps to reduce the computational complexity of the pattern recognition problem while preserving the features at multiple resolution. Further, patch-wise Weber local descriptor (WLD) histogram is obtained for low pass, composite band pass and high pass sub-bands.

WLD histograms are concatenated for entire eye-strip image which results in a composite feature vector. Since the size of composite feature vector is large, linear discriminant analysis (LDA) is utilized for dimensionality reduction. All the training composite vectors are stored in feature database. Nearest neighbour classifier with Euclidean distance is used for matching between different composite feature vectors. The nearest neighbour distance (d) between two feature vectors \vec{f}_{train} and \vec{f}_{test} is computed by (6).

$$d = \|\vec{f}_{train} - \vec{f}_{test}\| \quad (6)$$

5. EXPERIMENTS AND RESULTS

The experimentation is performed on JAFFE, Yale and Faces-94 databases. Fig. 6. illustrates extracted sample eye strip images from these databases. For color images we concatenate features detected in all the three bands of the color image at each level. The query image is subjected to non-subsampled contourlet transform to obtain forward decomposition coefficients. The subband coefficients have positive as well as negative values. Thus, the Weber Local Descriptor gives robust feature extraction using the directional subbands. The addition of pyramidal level subbands gives prominent edges in multiple directions congregated in a single composite subband which is computationally easier than obtaining features from each subband. The composite faces for a sample image are shown in Fig. 7. Further, the composite subbands are splitted patchwise to obtain the WLD histograms as a feature vectors. The concatenated feature vectors are subjected to linear discriminant analysis to perform dimensionality reduction. The compact feature vectors are used to represent the faces in the database as well as for the purpose of classification.

5.1 JAFFE facial database

Japanese Female Facial Expression (JAFFE) database [26] is first used for eye-strip based person identification. This database contains facial images of ten Japanese female models. The images are cropped in order to obtain a person specific database of eye-strip images. We have considered 20 eye-strip images/person for the experimentation purpose.



Fig 6: Sample faces from (a) JAFFE database (b) Yale database (c) Faces94 database.

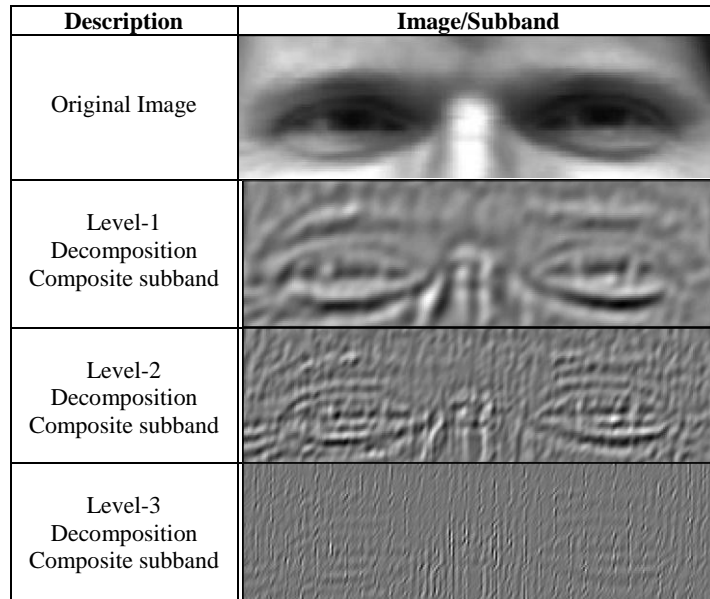


Fig 7: Composite faces at each pyramidal level after addition of NSCT subbands

Eleven eye-strips per person were selected as a gallery (database) whereas, the remaining nine eye-strips were used as query images. Fig.8 indicates cumulative match scores (CMC) on JAFFE database for three cases namely, NSCT+WLD with sub-band addition at each pyramidal level i.e. Proposed approach, NSCT+WLD without sub-band addition and NSCT+LBP with sub-band addition.

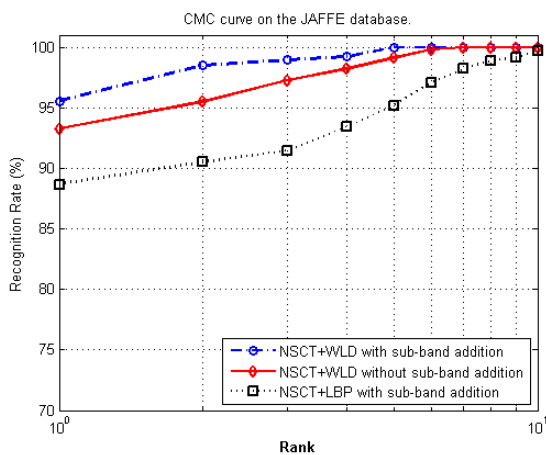


Fig 8: CMC curves on JAFFE database

After computing non-subsampled contourlet transform subbands at $[2, 2, 2]$ scales, and their subsequent addition, WLD features are computed for each composite band pass, low pass and high pass sub-band with 32×32 pixels as a patch size, which results in a composite feature vector. With the proposed approach, When 11 images per subject are denoted as a gallery set, rank-one recognition rate of 95.56% is achieved as an average of all the training sequence results of the combinations. The gallery set is selected sequentially from the set of person specific eye-strip images.

5.2 Yale facial database

Yale face database [27] consists of 165 gray-scale intensity images of fifteen subjects. Fig. 9 represents Cumulative Match Scores (CMC) on JAFFE database for three cases namely, NSCT+WLD with sub-band addition at each pyramidal level (Proposed approach), NSCT+WLD without sub-band addition and NSCT+LBP with sub-band addition. The images are cropped to obtain a person specific database of eye-strip images. Subsequently, WLD histograms are obtained on the composite band-pass, low-pass and high-pass sub-bands of non-subsampled contourlet transform (NSCT) coefficients at $[2, 2, 2]$ scales. Five images per subject are denoted as training images (gallery) and remaining images are treated as probe, we have achieved rank-one recognition rate of 89.33%. The gallery images are selected sequentially from the set of person specific eye-strip images

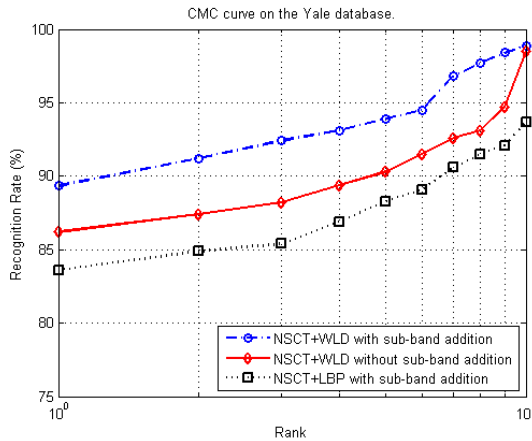


Fig 9: CMC curves on Yale database

5.3 Faces-94 Essex university face database

Faces-94 Essex university face database [14] consists of facial images of 153 subjects with 20 images/subject obtained at a fixed distance with considerable expression variations. In order to perform benchmarking, we have followed the same protocol with cropped image of the size 30x61 pixels as given in [14]. The benchmarking results are presented in Table 1. The gallery set consists of ten images/person whereas the remaining ten images were utilized as a query dataset. The training set is selected in a sequential order and it is mutually exclusive from testing data. In this experimentation, FRR and FAR tests are performed. Total Success Rate (TSR) is computed with (7) as in [14].

$$TSR = \left(1 - \frac{FAR + FRR}{Total\ accesses} \right) \times 100\% \quad (7)$$

Table 1. Benchmarking results with Faces-94 Essex database based approach [14]

Approach	Total Success Rate (TSR)
NMF [14]	93.23%
LNMF [14]	95.12%
SFNMf [14]	94.48%
Proposed Approach	97.32%

6. CONCLUSION

In this paper, an efficient and robust eye-strip based person identification scheme is presented which utilizes non-subsampled contourlet transform (NSCT) domain features extracted using Weber Local Descriptors. Addition of similar pyramidal sub-bands is advantageous for computational simplicity. Patch-wise Weber local descriptor gives discriminant eye-strip features. Experimentation results indicate significant increase in total success rate due to combination sub-bands at each pyramidal level. This technique is competent and yields satisfactory results with fewer eye-strip images in the gallery. The use of WLD as a feature descriptor results in more efficiency than LBP operator for eye-strip based person identification.

7. REFERENCES

- [1] A. Samal, and P. A. Iyengar 1992. Automatic recognition and analysis of human faces and facial expressions: a survey. *Pattern Recognition*. vol. 25, no. 1, 65-77.
- [2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld 2003. Face recognition: A literature survey. *ACM Comput. Surv.*, vol. 35, no. 4, 399-458.
- [3] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2007. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*. vol. 28, no. 14, 1885-1906.
- [4] A. A. Ross, K. Nandakumar, and A. K. Jain. 2006. *Handbook of Multibiometrics (International Series on Biometrics)*: Springer-Verlag New York.
- [5] D. Mou. 2010. *Automatic Face Recognition, " Machine-based Intelligent Face Recognition*. Springer Berlin Heidelberg.
- [6] Z. Wenchao, S. Shiguang, C. Xilin, and G. Wen. 2007. Local Gabor Binary Patterns Based on Kullback Leibler Divergence for Partially Occluded Face Recognition. *IEEE Signal Processing Letters*. vol. 14, no. 11, 875-878.
- [7] S. M. Huang, and J. F. Yang. 2012. Subface hidden Markov models coupled with a universal occlusion model for partially occluded face recognition. *Biometrics, IET*, vol. 1, no. 3, 149-159.
- [8] X.-X. Li, D.-Q. Dai, X.-F. Zhang, and C.-X. Ren. 2013. Structured sparse error coding for face recognition with occlusion. *IEEE transactions on image processing*. vol. 22, no. 5, 1889-1900.
- [9] W. Chia-Po, C. Chih-Fan, and Y. C. F. Wang. 2014. Robust Face Recognition with Structurally Incoherent Low-Rank Matrix Decomposition. *IEEE Transactions on Image Processing*. vol. 23, no. 8, 3294-3307.
- [10] G. Zhirong, L. Dongmei, X. Chengyi, H. Jianhua, and B. Huang. 2014. Face recognition with contiguous occlusion based on image segmentation. In *proceedings of International Conference on Audio, Language and Image Processing (ICALIP)*.
- [11] R. Amutha. 2012. A novel approach to face recognition under various facial expressions, occlusion and tilt angles. In *proceedings of International Conference on Emerging Trends in Science, Engineering and Technology (INCOSSET)*.
- [12] M. De Smet, R. Fransens, and L. Van Gool. 2006. A Generalized EM Approach for 3D Model Based Face Recognition under Occlusions. In *proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- [13] L. Guanglu, Y. Yan, and W. Hanzi. 2013. Robust Modular Linear Regression Based Classification for Face Recognition with Occlusion. In *proceedings of Seventh International Conference on Image and Graphics (ICIG)*.
- [14] T. Chuan Chin, N. Han Foon, and A. B. J. Teoh. 2007. A study on partial face recognition of eye region. In *proceedings of International Conference on Machine Vision*.
- [15] C. Deng, H. Xiaofei, W. Xiaoyun, and H. Jiawei. 2008. Non-negative Matrix Factorization on Manifold. In

- proceedings of Eighth IEEE International Conference on Data Mining.
- [16] A. L. da Cunha, Z. Jianping, and M. N. Do. 2006. The Nonsampled Contourlet Transform: Theory, Design, and Applications. *IEEE Transactions on Image Processing*. vol. 15, no. 10, 3089-3101.
- [17] Y. Zhou, and J. Wang. 2012. Image denoising based on the symmetric normal inverse Gaussian model and non-sampled contourlet transform,” *IET Image Processing*. vol. 6, no. 8, 1136-1147.
- [18] Y. Cheng, C. L. Wang, Z. Y. Li, Y. K. Hou, and C. X. Zhao. 2010. Multiscale principal contour direction for varying lighting face recognition,” *Electronics Letters*, vol. 46, no. 10, 680-682.
- [19] G. Bhatnagar, Q. M. J. Wu, and L. Zheng. 2013. Directive Contrast Based Multimodal Medical Image Fusion in NSCT Domain,” *IEEE Transactions on Multimedia*. vol. 15, no. 5, 1014-1024.
- [20] M. N. Do, and M. Vetterli. 2005. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*. vol. 14, no. 12, 2091-2106.
- [21] H. Y. Patil, A. G. Kothari, and K. M. Bhurchandi. 2014. Expression invariant face recognition using contourlet transform. In *proceedings of Image Processing Theory, Tools and Applications*.
- [22] R. H. Bamberger, and M. J. T. Smith. 1992. A filter bank for the directional decomposition of images: theory and design. *IEEE Transactions on Signal Processing*. vol. 40, no. 4, 882-893.
- [23] C. Jie, S. Shiguang, H. Chu, Z. Guoying, M. Pietikainen, C. Xilin, and G. Wen. 2010. WLD: A Robust Local Image Descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 32, no. 9, 1705-1720.
- [24] D. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. vol. 60, no. 2, 91-110.
- [25] T. Ahonen, A. Hadid, and M. Pietikainen. 2006. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 28, no. 12, 2037-2041.
- [26] M. Lyons, J. Budynek and S. Akamatsu. 1999. Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol 21, no. 12, 1357-1362.
- [27] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman. 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, 711-720.