ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (692-698), Month: April - June 2015, Available at: www.researchpublish.com

HOG and CS-LBP Methods for Human Detection with Occlusion

[1]S.K UMA, [2]SRUJANA B J, [3]Dr. B R RAMACHANDRA

[1]Associate professor, [1,2]Dept of Computer Sscience and Engineering, PES College of Engineering, Mandya, Affiliated to VTU, Belagavi

[3]Professor and Head, Dept, of Electrical and Electronics, PES College of Engineering, Mandya, Affiliated to VTU, Belagavi

Abstract: **Accurately detecting humans under occlusion in images plays a critically important role in many computer vision applications. Extraction of effective features is the key to this task. Promising features should be discriminative, robust to various variations and easy to compute. In this paper, we present center-symmetric local binary patterns (CS-LBP) and Histogram of oriented gradients (HOG), for human detection with occlusion. HOG feature calculates the gradient magnitude and the gradient direction of the local image. The main drawback of HOG feature extraction is that, it does not work well with different texture and pose of human and accuracy of detection is less. Rather than using LBP for feature extraction, we use CS-LBP because LBP produces long histogram and are not too robust on flat image areas. The CS-LBP feature captures both gradient information and texture information and works well on flat image areas. So we combine HOG and CS-LBP method for better human detection with occlusion. Experiments on the INRIA pedestrian dataset show that the combination of CS-LBP feature and HOG feature with linear support vector machines (SVMs) gives better result for human detection under occlusion.**

Keywords**: Center Symmetric - Local Binary Pattern (CS-LBP), Histogram of Oriented Gradients (HOG), INRIA pedestrian dataset, Feature Vector, Occlusion, Support Vector Machine (SVM).**

1. INTRODUCTION

The ability to detect humans under occlusion in image has a major impact on applications such as video surveillance [1], smart vehicles [2], [3], robotics [4]. Humans usually have many different appearances in pose and style, and the background of the images is often cluttered and has on general describable structure. So, human detection under occlusion in image is a challenging task for the variable appearance and various poses, which can influence the algorithm of choice. The articulated pose, style and color of clothes, illumination conditions in outdoor scene will affect the detection results.

Feature extraction is of the center importance here. In this work, features are extracted using the combination of HOG and CS-LBP.

The HOG descriptor was proposed in [5] for human detection. Since then, the descriptor has grown in popularity due to its success. These features are now widely used in object recognition and detection. They describe the body shape through a dense extraction of local gradients in the window. Usually, each region of the window is divided into overlapping blocks where each block is composed of cells. A histogram of oriented gradients is computed for each cell. The final descriptor is the concatenation of all the blocks features in the window. The main drawback of HOG is that, it doesn't work well with different texture and pose of human.

The LBP [6],[7],[8] has properties that favour its usage in interest region description such as tolerance against illumination changes and computational simplicity. Drawbacks are that the operator produces a rather long histogram and is not too robust on flat image areas and slightly bend areas. To address these problems, in this paper, we propose a new LBP-based texture feature, denoted as center-symmetric local binary pattern (CS-LBP) that is more suitable for the given problem.

The drawback of HOG and LBP is overcome with the use of combination of HOG and CS-LBP for feature extraction. In CS-LBP, instead of comparing each pixel with the center pixel, we compare center-symmetric pairs of pixels. CS-LBP method of feature extraction produces less number of binary patterns and generates small histogram. So that analysis of binary pattern and histogram is simple than compare to LBP method of feature extraction. CS-LBP is more efficient for different texture and pose of human since the diagonal elements are compared.

2. ARCHITECTURE OF OUR APPROACH

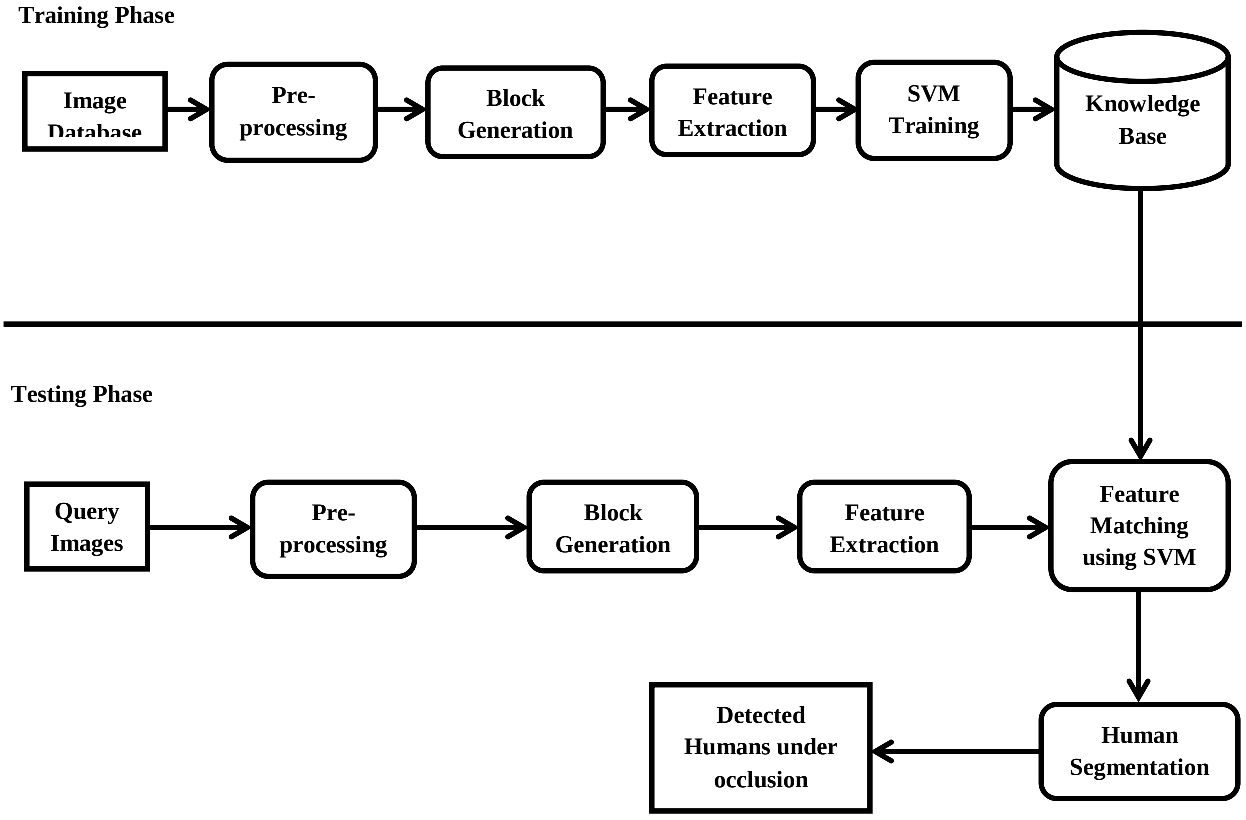

Fig. 1 The architecture of detection system

Fig. 1 illustrates our approach at two phases: **Training** and **Testing**. In Training phase, image database is pre-processed first. Pre-processing includes re-sizing the image to 256*256 pixel resolution and removal of noise area-wise, width-wise and height-wise mainly to eliminate illumination, shadow and other environmental effects. Image is divided into 64 blocks with each block having 32*32 dimensions. Features are extracted using HOG and CS-LBP method for each block and combined features of all block gives feature vector. SVM Training is done on the extracted features for classification.

In Testing level, query image is pre-processed same as we do in training phase. Query image is divided into 64 blocks with each block having 32*32 dimensions. Features are extracted using HOG and CS-LBP method for each block and combined features of all block gives feature vector. SVM classification is done where the classifier matches the training and testing features. If match found between training and testing features, then it detects human under occlusion else not.

3. HISTOGRAM OF ORIENTED GRADIENT (HOG)

HOG feature is an excellent descriptor, which calculates the gradient magnitude and the gradient direction of the local image. It has shown great success in object detection and recognition. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. In practice this is implemented by dividing the image window into small spatial regions for called cells. For each pixel, calculating a local 1-D histogram of gradient directions or edge orientations present in the cell. The combined histogram entries form the descriptor blocks which is referred to as *Histogram of Oriented Gradient (HOG)* descriptors. The following are the steps in HOG:

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (692-698), Month: April - June 2015, Available at: www.researchpublish.com

3.1 Gradient computation:

The gradient computation is done in two steps: the first step of gradient computation is the computation of centered mask. The most common method is to simply apply the 1-D centered point discrete derivative mask in both horizontal and vertical directions. This is done to smooth the image. Specifically, this method requires filtering the color or intensity data of the image.

The second step of gradient computation is to find the gradient angle and gradient magnitude for each pixel in a cell.

3.2 Orientation Binning:

The orientation binning involves creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. The cells are rectangular in shape, and the histogram channels are evenly spread over 0 to 180 degrees depending on whether the gradient is "unsigned" or "signed". The unsigned gradient used in conjunction with 9 histogram channels (0^0-20^0,20^0-40^0,40^0-60^0,60^0-80^0,80^0-100^0,100^0-120^0,120^0-140^0,140^0-160^0,160^0-180^0) performs best in the human detection.

3.3 Descriptor blocks:

We extract feature from each cells, and cells are concatenated to each other to construct a block descriptor. The final descriptor is obtained by the concatenation of all the blocks features in the window.

The above steps for extracting features using HOG is summarized in the Fig. 2:

Fig. 2 HOG detector.

4. CENTER-SYMMETRIC LOCAL BINARY PATTERN (CS-LBP)

After pre-processing, we extract a feature for each pixel of the region using the center-symmetric local binary pattern (CS-LBP) operator which was inspired by the local binary patterns (LBP). The LBP operator produces rather long histograms and is therefore difficult to use in the context of a region descriptor. To produce more compact binary patterns, we compare only center-symmetric pairs of pixels. As can be seen from Fig. 3, there are 8 neighbours for which LBP produces 256 different binary patterns, whereas CS-LBP produces only 16 different binary patterns. Furthermore, robustness on flat image regions is obtained by thresholding the grey level differences with a small value T :

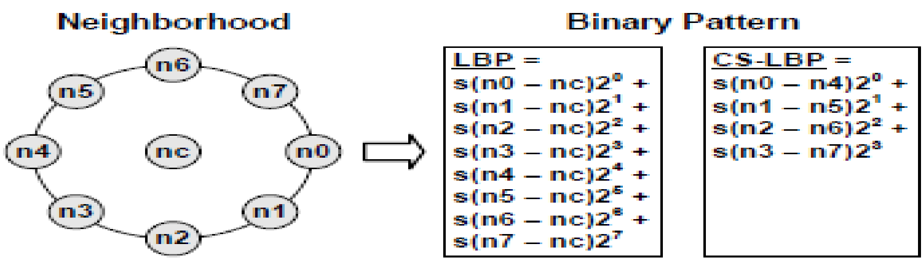

Fig. 3 LBP and CS-LBP features for a neighbourhood of 8 pixels

Center-Symmetric Local Binary Pattern is defined as:

$$\mathrm{CS-LBP}_{R,N,T}(x, y) = \sum_{i=0}^{\left(\frac{N}{2}\right)-1} s(n_i - n_{i+(N/2)})2^i \qquad \text{........Eqa 1}$$

Where,
n_i and $n_{i+(N/2)}$ correspond to the grey values of center-symmetric pairs of pixels.
N is the number of neighbourhood pixels.
R is the radius of equally spaced pixels on a circle.
T is the threshold value

T is set to 0.1. The radius is set to 2 and the size of the neighbourhood N is 8. All the experiments presented in this paper are carried out for these parameters ($CS - LBP$ 2, 8, 0.1) which gave the best result for the given test data. It should be noted that the gain of CS-LBP over LBP is not only due to the dimensionality reduction, but also to the fact that the CS-LBP captures better the gradient information than the basic LBP. Experiments with LBP and CS-LBP have shown the benefits of the CS-LBP over the LBP, in particular, significant reduction in dimensionality while preserving distinctiveness.

5. SUPPORT VECTOR MACHINE (SVM) FOR CLASSIFICTION

Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data Classification etc. Most of the existing supervised classification methods are based on traditional statistics, which can provide ideal results when sample size is tending to infinity. However, only finite samples can be acquired in practice. In this work, a learning method, Support Vector Machine (SVM), is applied on image data.

Support Vector Machines (SVM) has recently became one of the most popular classification methods. They have been used in a wide variety of applications such as text classification, facial expression recognition, gene analysis and many others. Support Vector Machines can be thought of as a method for constructing a special kind of rule called a linear classifier in a way that produces classifiers with theoretical guarantees of good predictive performance.

SVMs are set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classification. A special property of SVM is simultaneously minimizing the empirical classification error and maximizing the geometric margin. So SVM is also called Maximum Margin Classifiers.

SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is to maximize the distance between the two parallel hyper planes. An assumption is made that, the larger the margin or distance between the parallel hyper planes to achieve better generalization error of the classifier.

5.1 Support Vector Machine Architecture:

SVM can be represented as a network architecture Fig. 4 resembling artificial neural networks (ANNs) that have been pruned to obtain model to improve generalization. However, the process of determining the network architecture regarding both models is different. Determining an appropriate ANNs architecture generally involves manual trial-and-error procedures, and mostly depends on the past experience and the preference of users, and the derived weights are not interpretable. Conversely, the SVMs network architecture can be analytically determined by the SVM algorithm, and thus the optimal support vector network is automatically obtained.

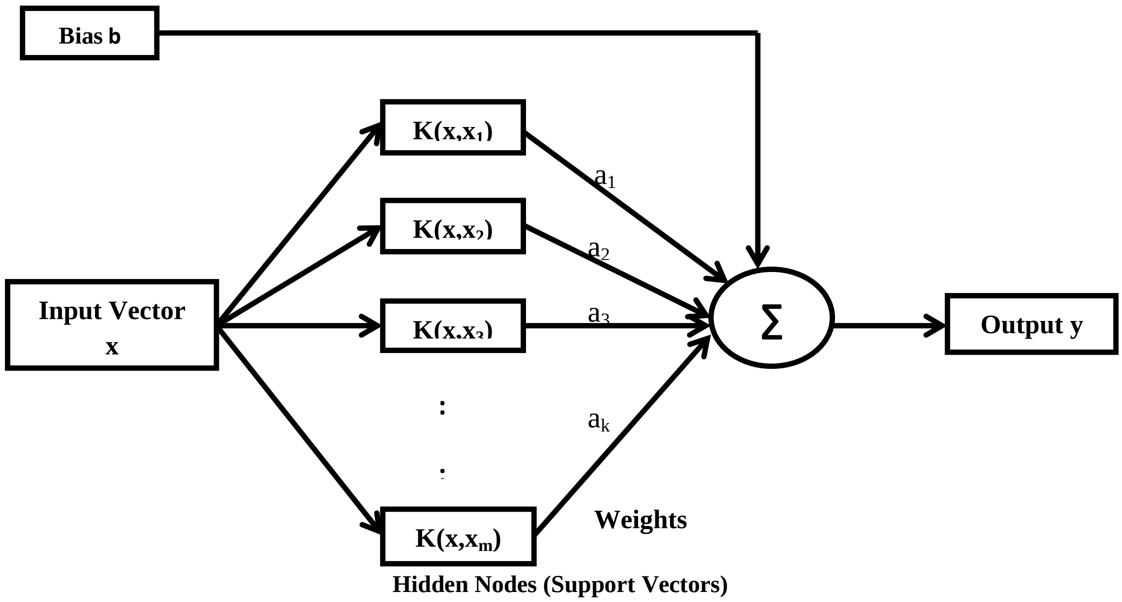

Fig. 4 Architecture of SVM

In Fig. 4, the input vector is the feature vector obtained from HOG and CS-LBP method. The centers of the hidden nodes are the support vectors (SVs) of the SVM. This support vectors compares feature vector of both training and testing images to produce classification output for each features.

$$K(x,x_k)$$

where: x is a feature vector of training image

x_k is a feature vector of testing image.

K is a kernel function of SVM.

After comparing each training and testing features, we get corresponding weights (a_k) for each features as a lagrange multiplier. The final output is obtained by adding all the weights along with the bias factor (b) as shown in equation 2.

$$y=f(x) = \sum_{k=1}^{m} (a_k * K(x, x_k) + b) \qquad \text{....... Eqa 2}$$

If both training and test features matches, then it detects human else not. If there is an occlusion in an image, then SVM detects occluded person and this can be highlighted by considering width and height parameters of an occluded person.

6. RESULTS AND DISCUSSION

6.1 Specification of Datasets:

The publicly available INRIA person dataset is used in this work which contains images in png format as shown in Fig. 5:

Fig. 5 INRIA person dataset images

6.2 Training and Evaluation:

We first choose training samples and re-size into 256*256 pixels. We divide each samples into 64 blocks for training. For each blocks, we do the feature extraction using HOG and CS-LBP. Later SVM is trained on the extracted features. The overall process is shown in the Fig. 6.

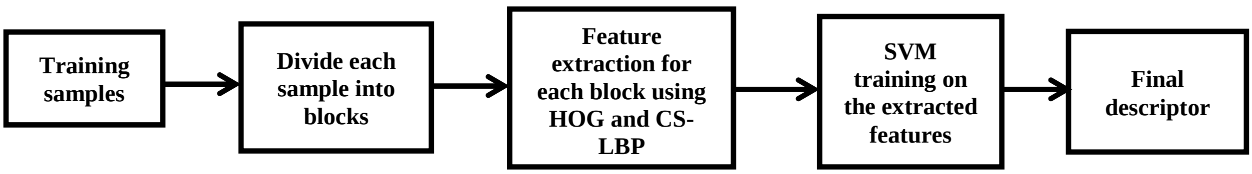

Fig. 6 The entire training process

6.3 Results after Classification using HOG features:

The classification of human using HOG feature is shown in the Fig. 7 below:

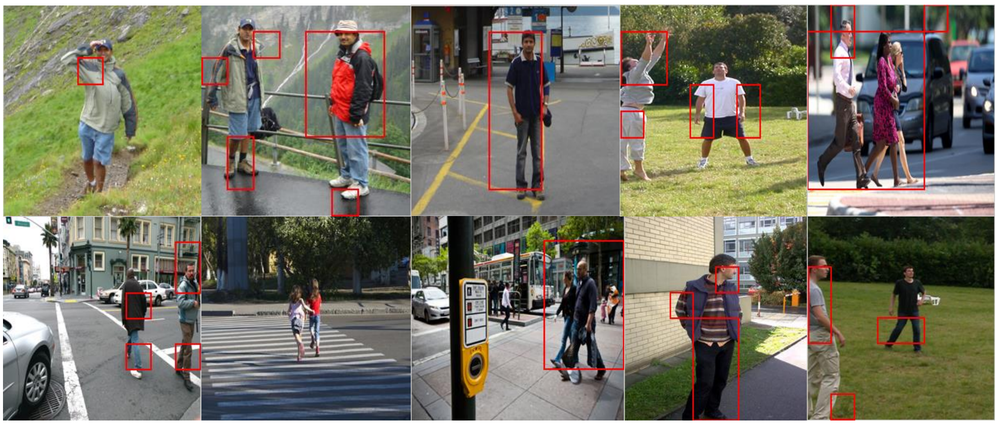

Fig. 7 Classification results using HOG

6.4 Results after classification using HOG and CS-LBP features:

The classification of human using HOG feature is shown in the Fig. 8 below:

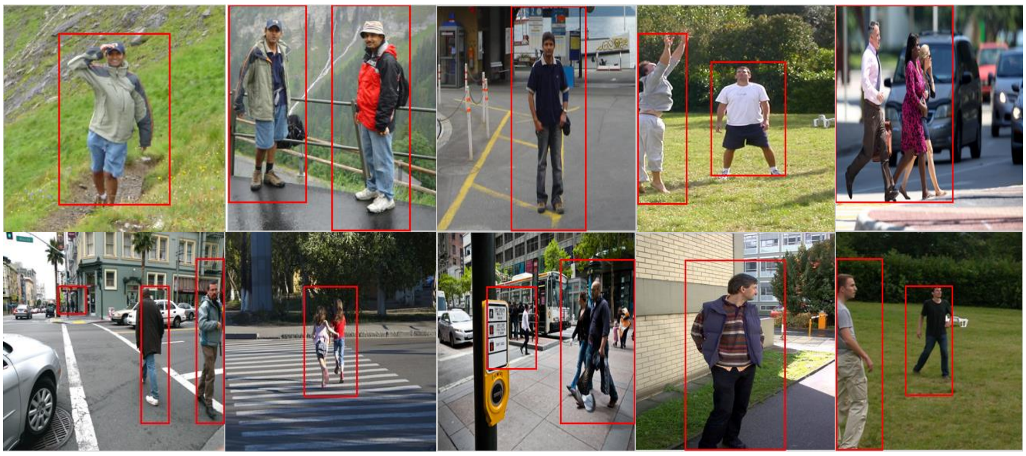

Fig. 8 Classification results using HOG and CS-LBP

6.5 Results of Occlusion using HOG and CS-LBP features:

Detection of occlusion using HOG and CS-LBP features are shown in the Fig. 9 below:

Fig. 9 Occlusion results using HOG and CS-LBP

6.6 Discussion:

The Fig. 5 shows the original images of INRIA person dataset. The Fig. 7 shows the classification of human using HOG features. HOG calculates gradient direction or edge direction for each pixel in an image. Detection using HOG is not efficient for different texture and pose of human. Also extracting features using HOG is time consuming.

The Fig. 8 shows the classification of human using combination of HOG and CS-LBP. Detection using the combination of HOG and CS-LBP gives better result compare to HOG for different texture and pose of human, because CS-LBP works well with different texture and pose of human since diagonal pixels are compared. Also CS-LBP produces less number of binary patterns and hence detection rate is more.

The Fig. 9 shows occlusion detection using the combination of HOG and CS-LBP features. HOG and CS-LBP efficiently detects partial occlusion if present in an image. From the experimental results it can be observed that the classification of human and occlusion using HOG and CS-LBP features gives more efficient result.

7. CONCLUSION

In this paper we made an attempt to compare HOG and HOG+CS-LBP method for feature extraction and the extracted features are classified by using SVM to classify human and occluded person. In HOG, gradient angle and magnitude is computed for each pixel in a block and assigning gradient magnitude to bin to get feature vector. In CS-LBP method, each pixel is compared against neighbourhood pixel in a diagonal manner and produces less number of binary patterns compared to LBP method.

The classification using HOG method is less accurate and do not work well with different texture and pose of human. Also extracting features using HOG is time consuming. CS-LBP method classifies human with different texture and pose of human and detection rate is more since it produces less number of binary patterns. Hence we classify using HOG and CS-LBP method and it confirms both HOG and CS-LBP gives more efficient result.

8. FUTURE WORK

In this paper, features are extracted using HOG and CS-LBP method, so we can also use some other feature extraction methods such as HAAR. Classification is done using SVM classifier. In future, we can also use cascaded classifiers for better human detection.

REFERENCES

[1] Haritaoglu, D. Harwood, and L. S. Davis, “W4: real-time surveillance of people and their activities,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 809–830, 2000.

[2] D. M. Gavrila and S. Munder, “Multi-cue pedestrian detection and tracking from a moving vehicle,” Int. J. Comp. Vis., vol. 73, no. 1, pp. 41–59, 2007.

[3] D. M. G. abd J. Giebel and S. Munder, “Vision-based pedestrian detection: the protector system,” in Proc. IEEE Int. Conf. Intell. Vehic. Symposium, Parma, Italy, 2004, pp. 13–18.

[4] T.Nakada, S.Kagami, and H.Mizoguchi, “Pedestrian detection using 3d optical flow sequences for a mobile robot,” in Proc. IEEE Conf. Sens., Leece, Italy, 2008, pp. 776–779.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. CVPR, San Diego, CA, USA, 2005, pp. 886–893.

[6] T. Ojala, M. Pietik¨ainen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition 29 (1) (1996) 51–59.

[7] T. Ojala, M. Pietik¨ainen, T. M¨aenp¨a¨a, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971–987.

[8] Http://www.ee.oulu.fi/mvg/page/lbp bibliography