Face Detection using SMQT Techniques

Y. Sangeetha

Lecturer
IT Department
VRSEC
Vijayawada

P. Madhavi Latha

Lecturer
IT Department
VRSEC
Vijayawada

Prof Ch.Narasimhan

Professor
ITDepartment
VRSEC
Vijayawada

Prof R.Satya Prasad

Professor &Head
CSE Department
Acharya Nagarjuna University
Vijayawada

ABSTRACT-This paper focuses on an extension to the Snow classifier, the split up Snow, for the chosen classification task. The study comprises of a threefold plan: firstly, the local Successive Mean Quantization Transform (SMQT) features are proposed for illumination and sensor insensitive operation in object recognition. Secondly, a split up Sparse Network of Winnows (Snow) is presented to speed up the original classifier. Finally, the features and classifier are combined for the task of frontal face detection. Detection results are presented for the MIT+CMU and the BioID databases. With regard to the face detector, the Receiver Operation Characteristics curve for the BioID database yields the significant outcome.

INTRODUCTION

Face detection is defined as to locate the position of face in a given image and verify the correctness of face in an image. Face detection is a procedure in which the whole image is examined to find regions that are identified as "FACE". In simple words face detection can be determined as "a computer technology that determines the locations and the sizes of the human faces in arbitrary images. It detects the facial features and ignores everything such as buildings, trees and bodies.

Over the last decade, face detection has become an active area of research in computer vision, neuroscience and psychology. Face detection is a required first step in the face recognition system. It has several applications in the areas of video coding and video conference and human computer interfaces. Illumination and sensor variation are major concerns in visual object detection. It is desirable to transform the raw illumination and sensor-varying image so the information only contains the structures of the object. Some techniques previously proposed to reduce this variation are computationally expensive operation in comparison with SMQT & SNOW classifier. The Successive Mean Quantization Transform (SMQT) can be viewed as a tunable trade off between the number of quantization levels in the result and the computational load. The split up Snow will utilize the result from the original Snow classifier and create a cascade of classifiers to perform a more rapid detection. It will be shown that the number of splits and the number of weak classifiers can be arbitrary within the limits of the full classifier. Further, a stronger classifier will utilize all information gained from all weaker classifiers. Face detection is a required first step in face recognition systems.

A novel learning approach for human face detection using a network of linear units is presented. The SNOW learning architecture is a sparse network of linear functions over a predefined or incrementally learned feature space and specifically tailored for learning in the presence of very large no of features. A wide range of face images in different poses, with different expressions and under different lighting conditions are used as training set to capture the variations of human faces. Furthermore, learning and evaluation methods using the SNOW based method are significantly more efficient than with other methods. The purpose of this paper is threefold: firstly, the local Successive Mean Quantization Transform features are proposed for illumination and sensor insensitive operation in object recognition. Secondly, a Split up Sparse Network of Winnows is presented to speed up the original classifier. Finally, the features and classifier are combined for the task of frontal face detection. Detection results are presented for the Bio ID databases. With regard to this face detector, the Receiver Operation Characteristics curve for the Bio ID database yields the best published result. The result for the database is comparable to state-of-the-art face detectors.

STATE OF ART

There are some methods proposed for the face detection. Few of them are 1. ADABOOST ALGORITHM for face detection. 2. SMQT TECHNIQUE for the face detection.

ADABOOST ALGORITHM:

In this adaboost algorithm first, skin color model in the YCbCr chrominance space is built to segment the non-skin-color pixels from the image. Then, mathematical morphological operators are used to remove noisy regions and fill holes in the skin-color region, so we can extract candidate human face regions. Finally, these face candidates are scanned by cascade classifier based on AdaBoost for more accurate face detection. This system detects human face in different scales, various poses, different expressions, lighting conditions, and orientation. Experimental results show the proposed system obtains competitive results and improves detection performance substantially. The ultimate goal is to develop fast detection algorithms with high practicalities. To train a

classifier, Viola and Jones select from a large number of very efficiently computable features. Every weak classifier implements a simple threshold function on one of the features. In order to ensure fast classification, we build a simple and efficient classifier using the AdaBoost learning algorithm to select a small number of critical visual features from a very large set of potential features.AdaBoost is an algorithm for constructing a "strong" classifier as linear combination of "weak" classifiers ht(x). The ht(x) can be thought of as one feature.

## PROPOSED METHOD & IMPLEMENTATION

The SMQT (SUCCESSIVE MEAN QUANTISATION TRANSFORM TECHNIQUE)   uses an approach that performs an automatic structural breakdown of information. Our previous work with the SMQT can be found in. These properties will be employed on local areas in an image to extract illumination insensitive features. For example, a straightforward method is to divide the image into blocks of a predefined size. Another way could be to extract values by interpolate points on a circle with a radius from a fixed point. Nevertheless, once the local area is defined it will be a set of pixel values. Let x be one pixel and $D(x)$ be a set of $|D(x)| = D$ pixels from a local area in an image. Consider the SMQT transformation of the local area

$$SMQT_L: D(x) \rightarrow M(x)$$

This yields a new set of values. The resulting values are insensitive to gain and bias. These properties are desirable with regard to the formation of the whole intensity image $I(x)$ which is a product of the reflectance $R(x)$ and the illuminance $E(x)$. Additionally, the influence of the camera can be modeled as a gain factor g and a bias term b. Thus, a model of the image can be described by

$$I(x) = g\,E(x)\,R(x) + b.$$

In order to design a robust classifier for object detection the reflectance should be extracted since it contains the object structure. In general, the separation of the reflectance and the illuminance is an ill posed problem. A common approach to solving this problem involves assuming that $E(x)$ is spatially smooth. Further, if the illuminance can be considered to be constant in the chosen local area then $E(x)$ is given by $E(x) = E$. Given the validity of the SMQT on the local area will yield illumination and camera-insensitive features. This is implies that all local patterns, which contain the same structure, will yield the same SMQT features for a specified level L see Fig. 1. The number of possible patterns using local SMQT features will be $(2^L{}^D)$. For example the 4×4 pattern at $L = 1$ in Figure has $4*4 = 65536$ possible patterns.
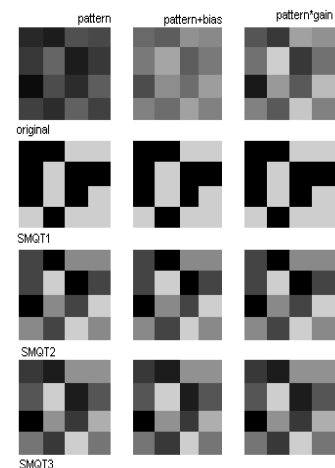
SPLIT UP SNOW CLASSIFIER:

Consider a patch W of the SMQT features $M(\mathbf{x})$, then a classifier can be achieved using the no face table H no face x, the face table H $face\ x$ and defining a threshold for θ.

$$\theta = \text{Sigma}\ (x{\sim}W)\ H\ \text{no face}\ (M(\mathbf{x})) - Sigma(X{\sim}W)H\ \text{face}(x)(M(\mathbf{x}))$$

Since both tables work on the same domain, this implies that one single lookup-table can be created for single lookup-table classification.

$$H\,x = H\,x\ \text{no face} - H\,x\ \text{face}.$$

Let the training database contain i $=1, 2 \ldots N$ feature patches with the SMQT features $M\,i(\mathbf{x})$ and the corresponding classes $c\,i$ (face or no face). The no face table and the face table can then be trained with the Winnow Update Rule. Initially both tables contain zeros. If an index in the table is addressed for the first time during training, the value (weight) on that index is set to one. There are three training parameters; the threshold γ, the promotion Parameter $\alpha > 1$ and the demotion parameter $0 < \beta < 1$. If $\mathbf{X}{\sim}W\ h$ face $x\ (M\,i(\mathbf{x})) \le \gamma$ and $c\,i$ is a face then promotion is conducted as follows h face $x\ (Mi(\mathbf{x})) = \alpha\ h$ face $x\ (Mi(\mathbf{x}))$. If $c\,i$ is a no face and $\mathbf{X}{\sim}W\ h$ face $x\ (Mi(\mathbf{x})) > \gamma$ then demotion takes place h face $x\ (Mi(\mathbf{x})) = \beta\ h$ face $x\ (Mi(\mathbf{x}))$ .This procedure is repeated until no changes occur. Training of the no face table is performed in the same manner, and finally the single table is created. One way to speed up the classification in object recognition is to create a cascade of classifiers. Here the full SNoW classifier will be split up in sub classifiers to achieve this goal. Note that there will be no additional training of sub classifiers; instead the full classifier will be divided. Consider all possible feature combinations for one feature, $Pi, i = 1, 2, \ldots, (2L)D$, then $v\,x =(2L)D\ X\ (i=1)|\ H\,x(Pi)|$ results in a relevance value with respective significance to all features in the feature patch. Sorting all the feature relevance values in the patch will result in an importance list. Rejecting no

faces within the training database, but at the cost of an increased number of false detections. The desired threshold used on θ is found from the face in the training database that results in the lowest classification value. Extending the number of sub classifiers can be achieved by selecting more subsets and performing the same operations as described for one sub classifier. Consider any division, according to the relevance values, of the full set W. Then W has fewer features and more false detections compared to W and so forth in the same manner until the full classifier is reached. One of the advantages of this division is that W will use the sum result from $W\_$. Hence, the maximum of summations and lookups in the table will be the number of features in the patch W. In order to scan an image for faces, a patch of 32×32 pixels is applied. This patch is extracted and classified by jumping $\Delta x = 1$ and $\Delta y = 1$ pixels through the whole image. In order to find faces of various sizes, the image is repeatedly downscaled and resized with a scale factor $Sc = 1.2$. To overcome the illumination and sensor problem, the proposed local SMQT features are extracted. Each pixel will get one feature vector by analyzing its vicinity. This feature vector can further be recalculated to an index.

$m = \text{Sigma } (I = 1 \sim D) \, \mathbf{V}(x \, I)(2^{\wedge}L^{\wedge}(I-1)).$

Where $\mathbf{V}(x_i)$ is a value from the feature vector at position i. This feature index can be calculated for all pixels, which results in the feature indices image. Face features with indices, with and with out masking.
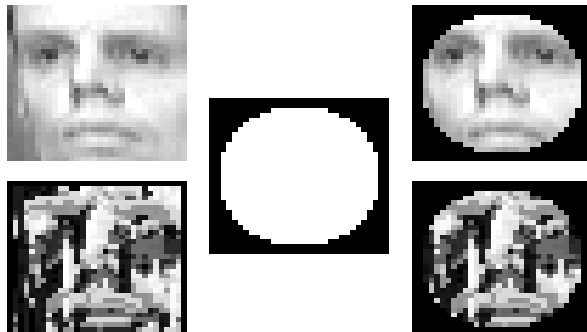

Masking of pixel image and feature indices image.
The features are here found by using a
3 *3 local area and L=1.

A circular mask containing $P = 648$ pixels is applied to each patch to remove background pixels, avoid edge effects from possible filtering and to avoid undefined pixels at rotation operation. With the Snow and the split up Snow classifier, the lookup table is the major memory-intense issue. Consider the use of $N \, bit = 32$ bit floating numbers in the table, then the classifier size (in bits) will be

$$S h \, x = N \, bit \, .P. \, (2(^{\wedge}L) \, ^{\wedge}D)$$

In this paper, the 3 *3 local area and level $L = 1$ are used and found to be a proper balance for the classifier. Some tests with 3 *3 and $L = 2$ were also conducted. Although these tests showed promising results, the amount of memory required made them impractical, see Tab. 1. The face and no face tables are trained with the parameters $\alpha = 1.005$, $\beta = 0.995$ and $\gamma = 200$. The two trained tables are then combined into one table according to Eq. 5. Given the Snow classifier table, the proposed split up Snow classifier is created. The splits are here performed on 20, 50, 100, 200 and 648 summations. This setting will remove over 90% of the background patches in the initial stages from video frames recorded in an office environment. Overlapped detections are pruned using geometrical location and classification scores. Each detection is tested against all other detections. If one of the area overlap ratios is over a fixed threshold, then the different detections are considered to belong to the same face. Given that two detections overlap each other, the detection with the highest classification score is kept and the other one is removed. This procedure is repeated until no more overlapping detect. Images are collected using a web camera containing a face, and are hand-labeled with three points; the right eye, the left eye and the center point on outer edge of upper lip (mouth indication). Using these three points the face will be warped to the 32X32 patch using different destination points for variation, see Fig. 3. Currently, a grand total of approximately one million face patches are used for training. Ions are found.

Initially the no face database contains randomly generated patches. A classifier is then trained using this no face database and the face database.
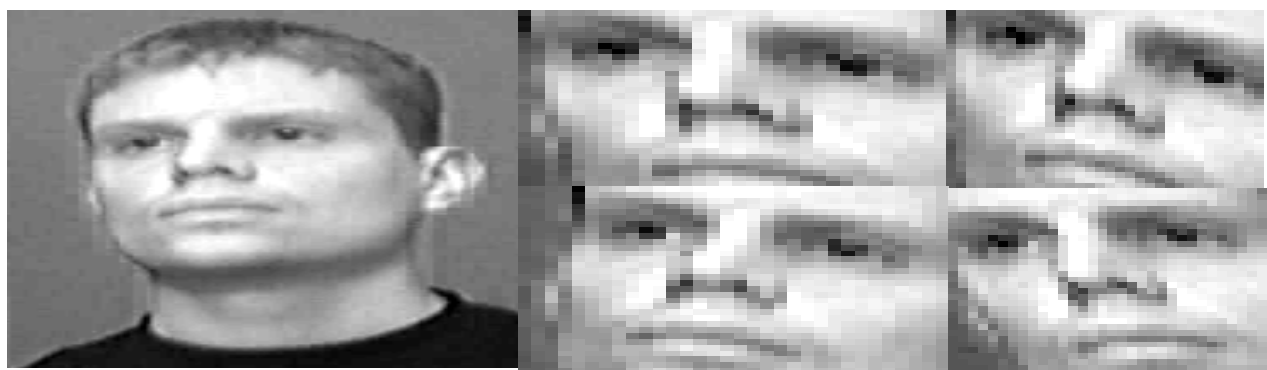
A collection of videos Prepared from clips of movies containing no faces and is used to bootstrap the database by analyzing all frames in the videos. Every positive detection in any Frame will be added to the no face database. The no face database is expanded using this bootstrap methodology. In final training, a total of approximately one million no face patches are used after bootstrapping.

Observations:
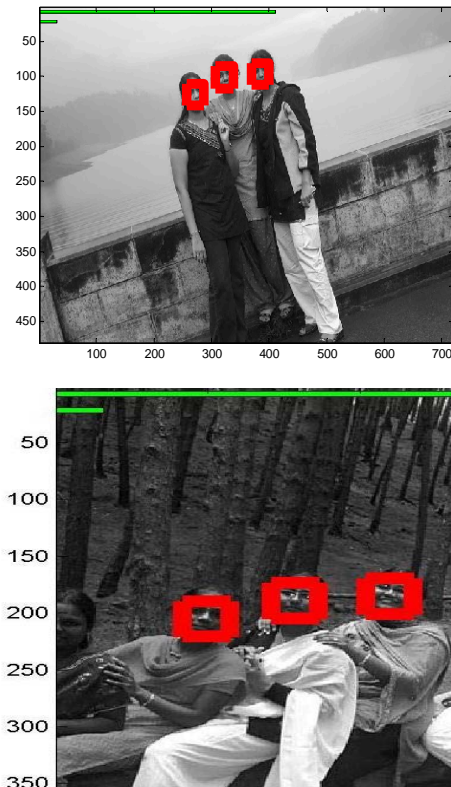
CONCLUSION

The Local SMQT features which can be used as feature extraction for object detection. Properties for these features were presented. The features were found to be able to cope with illumination and sensor variation in object detection. Further, the split up Snow was introduced to speed up the standard Snow classifier. The split up Snow classifier requires only training of one classifier network, which can be arbitrarily divided into several weaker classifiers in cascade. Each weak classifier uses the result from previous weaker classifiers which makes it computationally efficient. A face detection system using the local SMQT features and the split up Snow classifier was proposed. The face detector achieves the best published ROC curve for the Bio ID database, and a ROC curve comparable with state-of-the-art published face detectors for the CMU+MIT database.

REFERENCES

[1] O. Lahdenoja, M. Laiho, and A. Paasio, "Reducing the feature vector length in local binary pattern based face recognition," in *IEEE International Conference on Image Processing (ICIP)*, September 2005, vol. 2, pp. 914–917.

[2] B. Froba and A. Ernst, "Face detection with the modified census transform," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, May 2004, pp. 91–96.

[3] M. Nilsson, M. Dahl, and I. Claesson, "The successive mean quantization transform," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2005, vol. 4, pp. 429–432.

[4] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 1, pp. 34–58, 2002.

[5] E. Hjelmas and B. K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 3, no. 3, pp. 236–274, 2001.