Object Detection by Selective Integration of Higher Order Local Autocorrelation Mask Feature

Akinori Hidaka
University of Tsukuba
hidaka.akinori($\alpha\tau$)aist.go.jp

Takio Kurita
National Institute of Advanced
Industrial Science and Technology

Nobuyuki Otsu
National Institute of Advanced
Industrial Science and Technology

Abstract

Higher order Local Autocorrelation (HLAC) proposed by Otsu [5] is often used in the recent computer vision application such as gate recognition, object tracking, or video surveillance. The feature value of HLAC is the integral of the product between local pixels' intensity, and usually the integral is calculated in entire images. However, in the image recognition, feature selection is often effective for the both of classification accuracy and processing speed. In this paper, we propose HLAC Mask Feature extracted from arbitrary local region, and its feature selection algorithm base on Adaboost technique. We show Adaboost can select HLAC Mask having higher classification power and lower computational cost than usual HLAC for face detection task.

1. Introduction

Recently, the importance of the image recognition technology has been increasing with the advance of the computers and network technology. Especially, the interest in advanced video surveillance systems is rapidly increasing; although surveillance cameras are installed at many places for security purposes, huge costs are required to analyze those data by manually in real time.

For the purpose of the video surveillance, human face detection is one of the most basic issue. Viola [6, 7] proposed Rectangular Feature (RF) based boosted detector. RFs are local image features and there are much variety depending on the focused local region. Viola used a variant of Adaboost [2] for RF selection to extract a small subset of superior RFs from many (not always superior) candidates. As shown by Viola, feature selection is an important issue for object detection. Both of accuracy and speed will be improved by eliminating unnecessary local features for the detection task.

Higher order Local Autocorrelation (HLAC), one of a primitive image feature, is proposed by Otsu [5]. HLAC is known as effective features for many applications in the image recognition. Especially for the face detection task, Kurita [3] used HLAC with Linear Descriminant Analysis [1].

In this paper, we propose HLAC Mask features by treating HLAC as local features. Usually, the feature value of HLAC is calculated by integral of the mask value that is calculated from all 3×3 pixel regions on an image. On the other hand, HLAC Mask has the arbitrary rctangular integral range. Furthermore, HLAC Mask is considering multiresolutional feature value for robust face detection. Depending on the focused resolutions and integral ranges, there are much variety of HLAC Mask as well as Viola's RF. Therefore, a variant of Adaboost could be used to select a small subset of superior HLAC Mask from many candidates. In our experiment, we show that Adaboost based feature selection algorithm can select the powerful HLAC Masks, and as a result we obtain more accurate face detector with less HLAC Mask than usual HLAC based detector.

2. Previous Method

In this section, we briefly introduce previous works: Higher order Local Autocorrelation[5], and variant of Adaboost for local feature selection [6, 7].

2.1. Higher order Local Autocorrelation

Otsu [5] proposed Higher order Local Autocorrelation (HLAC) as the primitive image feature. It is known that HLAC features are very effective for many applications

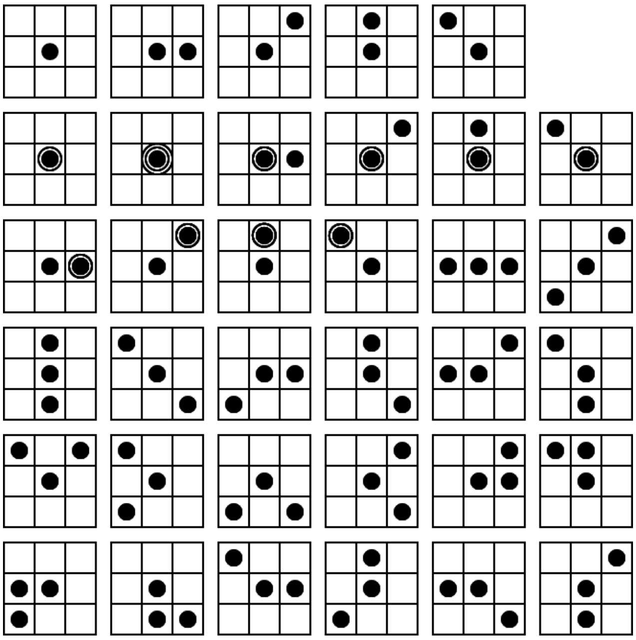

Figure 1. 35 mask patterns of HLAC.

in the image recognition. HLAC features are subset of the higher order autocorrelations [5]. The feature value of HLAC is the auto-correlation functions calculated from the object image by Eq. (1).

$$x(a_1, \cdots, a_N) = \int I(r)I(r + a_1)\cdots I(r + a_N)dr \quad (1)$$

where r is a referred pixel, $I(r)$ is the intensity value of r, and $a_1, \cdots, a_N$ are the displacements from r. The number of these autocorrelation functions obtained by the combination of the displacements over the image is enormous. For practical application, they are reduced to define HLAC features. At first, we restrict the order N up to the second. Then, we also restrict the range of displacements to within a local 3×3 window, because the correlation within local region is much higher than the correlation between far points. By eliminating the displacements which are equivalent by the shift, the numbers of patterns of the displacements are reduced to 35. Fig. 1 shows the 35 mask patterns of HLAC feature.

2.2. Adaboost for Local Feature Selection

Adaboost [2] is the ensemble learning method that trains multiple base-classifiers and assembles these to create a more powerful classifier. In the iterations of Adaboost, a classifier that assists the weakness of assembled ensemble is chosen and added into the ensemble. Therefore, the assembled ensemble will effectively obtain a perfect classification power for given training samples.

The algorithms of original Adaboost (Adaboost.M1) is shown in Figs. 2 and 3. In the training of Adaboost, each training sample I_i is assigned weight w_i that implies the "difficulty" of sample I_i. The cost function of Adaboost is designed as weighted classification error rate for training samples. Weight w_i is made big if the base-classifier selected newly misclassified sample I_i. Therefore, at the next iteration, the base-classifier that can correctly classify the samples which assembled ensemble fails will be chosen.

Recently, a variant of Adaboost was proposed to create local-feature based face detector, by Viola and Jones [6, 7]. For convenience, we call the algorithm Adaboost.LFS. The optimization function for Adaboost.LFS is shown in Fig. 4. Adaboost.LFS only added the phase of feature selection to Adaboost.M1. Thus, there are two optimization phases in Adaboost.LFS;

1) parameters fitting for candidate classifiers, and

2) selection of the best classifiers.

If ones consider that variety of local-features are kinds of base-classifier's parameters, Adaboost.LFS is equivalent to Adaboost.M1.

3. HLAC Mask and Feature Selection

We propose a new feature extraction method that calculates mask values of HLAC in variety resolutions and integral range, and its feature selection algorithm by using a variant of Adaboost.

3.1 HLAC Mask Feature

HLAC calculates the product of local intensity values (mask values) in all image regions, and integrates them by uniform weights. However, in some cases, the recognition accuracy will be improved when the locations where the feature value is calculated are appropriately selected; for example, in the face recognition task, the significant parts such as eyes or mouths will be considered more infomative than the flat parts such as cheek or foreheads. Furthermore, the image resolution should be appropriately set for each focused region.

In this paper, we propose HLAC Mask Feature that calculates mask values of HLAC in variety resolutions and integral range. HLAC Mask has the following additional parameters; the mask's resolution k ($k = 1, 2, \cdots, 8$) and the integral range R (Fig 5). In the k-th resolutional mask, each mask pattern is expanded to $3k \times 3k$ pixel region composed from 9 elemental regions of $k \times k$ pixel (Fig. 6). The feature value of HLAC Mask is not based on pixel intensity, but based on the mean intensity of expanded elemental regions. Actually, the feature value of k-th resolutional HLAC Mask

- Input labeled samples $\{I_i, y_i\}_{i=1}^N$. ($I_i \in R^d$: sample, $y_i \in \{0, 1\}$: class label.)
- Initialize samples weights: if $y_i = 1$ then $w_i = \frac{1}{2p}$, otherwise $w_i = \frac{1}{2q}$. (p: # of face, q: # of non-face)
- for $t = 1, \cdots, T$
  - Normalize samples weights: $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{i=1}^N w_{t,i}}$.
  - Optimize base-classifiers $\{b_c\}_{c=1}^C$: $\{b_t, err_t\} = OptFunc(\{b_c\}_{c=1}^C, \{I_i, w_i, y_i\}_{i=1}^N)$
  - Compute $\alpha_t = \log((1 - err_t)/err_t)$
  - Update samples weights: $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot \delta(y_i - b_t(I_i))]$. ($\delta(x) = 1$ (if $x = 0$), 0 (otherwise).)
- Final classification function is:
  $H(I) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t b_t(I) \geq \Theta \sum_{t=1}^T \alpha_t. \\ 0 & \text{otherwise.} \end{cases}$ (Θ: Threshold)

Figure 2. Common Adaboost algorithm. $OptFunc$ **for the described methods are shown in Figs. 3 to 4.**

- Input arguments $\{b, \{I_i, w_i, y_i\}_{i=1}^N\}$.
- Fit a classifier b to the training samples using weights w_i.
- Compute $err = \frac{\sum_{i=1}^N w_i \cdot \delta(y_i - b(I_i))}{\sum_{i=1}^N w_i}$.
- Return $\{b, err\}$

Figure 3. Optimization function for Adaboost.M1.

- Input arguments $\{\{b_c\}_{c=1}^C, \{I_i, w_i, y_i\}_{i=1}^N\}$.
- For $c = 1, \cdots, C$,
  - Fit a classifier b_c to the training samples using weights w_i.
  - Compute $err_c = \frac{\sum_{i=1}^N w_i \cdot \delta(y_i - b_c(I_i))}{\sum_{i=1}^N w_i}$.
- Choose the classifier b^* with the lowest error err^*.
- Return $\{b^*, err^*\}$

Figure 4. Optimization function for Adaboost.LFS.

is calculated as follows:

$$f_k(a_1, \cdots, a_N) = \int_R \prod_{n=1}^N M(e_{kn}) dr$$

where e_{kn} is the n-th expanded elemental region used for product calculation, and $M(x)$ is the mean intensity of elemental regions x.

In our method, the multiresolutional mask values are obtained by expanding the calculation region of the mask instead of changing the image resolution. The low-resolutional mask value is calculated base on the averaged intensity in extended regions. As a result, the feature values of HLAC Mask in arbitrary resolutions can be calculated rapidly by using Integral Image method [6, 7].

HLAC Mask is equivalent to usual HLAC if only the original resolution is used and integral range is fixed at entire image.

3.2. Feature selection and integration for HLAC Mask

To treat HLAC Mask as base-classifiers for Adaboost.LFS, we define the weak-classification function as follows;

$$b(I) = \begin{cases} 1 & \text{if } pf(I) > p\theta \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where I shows an input image, $f(I)$ implies a feature value of HLAC Mask at I, and $p \in \{1, -1\}$ and $\theta \in R$ are the parameters determined by training (see [6, 7]).

Resolution	Origin	1/2	1/3	1/4	1/5	1/6	1/7	1/8	All
Classification rate (%)	95.9	96.1	95.5	94.4	93.2	91.4	90.0	84.0	97.5

Table 1. The comparison of single-resolutional mask and multiresolutional mask. Each classification rate is the best one from up to $1,000$ selected masks. The columns from 'Origin' to '1/8' show the result by the single-resolutional mask using its resolution, and the column 'All' shows the result by multiresolutilnal masks using all resolutions.

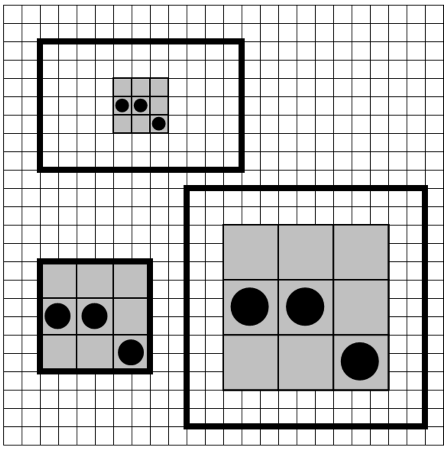

Figure 5. The examples of HLAC Mask in 24×24 pixel images. Each bold frame shows one HLAC Mask Feature. Bold frame region implies the integral range R of the mask, and the size of the mask pattern in the frame implies the mask resolution (Fig.6).

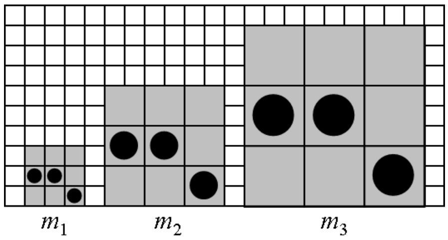

Figure 6. Multiresolutional Mask. m_1 is the mask of the original (highest) resolution. m_2, m_3 is the mask of $1/2, 1/3$ resolution, respectively. The mask value of m_1 is the product of intensity of the referred points (black circle), and the value of m_k $(k = 2, 3, \cdots)$ is the product of the averaged intensity in referred regions instead.

4. Experiments

In this paper, we tested the performance of HLAC Mask Feature for the face classification task. We used 725 face and $2,200$ background images of 24×24 pixel for the experiments. We divided all images into training and test set; 300 face and $1,000$ background images are used for the training, and remained images are used for the test. Three pairs of training and test sets were generated randomly. All of the accuracies presented in this section are the averaged values calculated from the three pairs. We describe classification experiments as follows.

First, we performed a preliminary experiment to fix the integral range because exhaustive search for the region by Adaboost is too expensive. As a result, it has been understood that the minimum size of integral range, i.e. the single HLAC Mask without any integral, have the highest classification rate. Therefore, in the actual experiments, we performed feature selection with integral range that is fixed at minimum.

Next, we tested the effect of multiresolutional masks. Table 1 shows the result of feature selection using the mask of single-resolution only or all resolution as candidate features. As a result, it was understood that 2nd or 3rd resolutional mask can obtain the comparable classification performance with 1st (original) resolutional mask that is equivalent to usual HLAC. When all resolutional masks were used as candidates for feature selection, we obtained the best result, 97.5%.

Fig. 7 shows the comparison result of proposed HLAC Mask trained by Adaboost and usual HLAC trained by Linear Support Vector Machine (SVM) [4]. In the usual HLAC, $(24-2)^2 * 35 = 16940$ masks were used to calculate the feature vector in 24×24 pixel images, and the classification rate is 96.6%. On the other hand, proposed method achieved comparable accuracy with 328 masks, and marked the highest result, 97.5%, with 601 masks.

5. Conclusion

In this paper, we proposed HLAC Mask Feature that has multiresolutional feature value and arbitrary integral

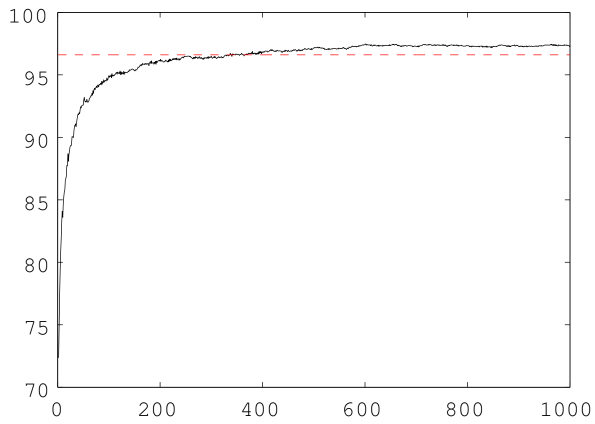

Figure 7. The classification rate of proposed method. The horizontal axis shows the steps of Adaboost, i.e. the number of selected HLAC Mask, and the vertical axis shows the average of classification rate for the test samples (%). The solid line is the rate of the proposed method, and the broken line is the rate of linear SVM with usual 35 dimenasional HLAC. The cost parameter C of SVM is selected from $C \in \{2^{-5}, 2^{-4}, \cdots, 2^{14}, 2^{15}\}$ as the best one for the test samples.

range, and its feature selection algorithm based on Adaboost technique. We showed the combination of multiresolutional HLAC Mask yields more better result than single-resolutional mask. Furthermore, feature selection of HLAC Mask was able to improve both classification power and processing speed of face detector compared with the SVM constructed from usual HLAC.

References

[1] R. A. Fisher, "The Use of Multiple Measurements in Taxonmic Problems," Ann. Eugenics, Vol. 7, No. Part II, pp. 179-188, 1936.

[2] Y. Freund and R. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting." Journal of Computer and System Sciences, 1997.

[3] T. Kurita, N. Otsu, and T. Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," Proc, of 11th International Conference on Pattern Recognition, Aug.30-Sep.3, The Hague, Vol.I, pp.530-533, 1992.

[4] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," In Proc. of Conf. Computer Vision and Pattern Recognition, pp.130-136, 1997.

[5] N. Otsu and T. Kurita, "A new scheme for practical, flexible and intelligent vision systems," In Proc. of IAPR Workshop on computer Vision, pp.431-435, 1988.

[6] P. Viola and M. Jones, "Robust real time object detection," In IEEE ICCV Workshop on Statistical and Computational Theories of Vision, July 2001.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," In Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, Dec. 2001.