Efficient Face Detection from News Images
by Adaptive Estimation of Prior Probabilities and Ising Search

Takio Kurita†, Masaru Tanaka‡, Kazuhiro Hotta‡, Hiroyuki Shimai‡, and Taketoshi Mishima‡
† Electrotechnical Laboratory,
1-1-4, Umezono, Tsukuba, 305-8568 JAPAN
kurita@etl.go.jp,
‡ Saitama University
255, Shimo-Okubo, Urawa, 338 JAPAN
{mtanaka,hotta,shimai,mishima}@me.ics.saitama-u.ac.jp

Abstract
This paper presents an efficient method to detect faces from image sequences such as news or visual surveillance. To speed up the search, the prior probabilities of the face locations in the image are adaptively estimated and are used in the Ising search algorithm. The search points in the Ising search are selected depending on the estimated prior probabilities. The information obtained by the previous search points in the given image is effectively utilized through spin flip dynamics of the Ising search. If a face is found, the prior probabilities are updated with forgetting. This makes adaptation to the changes of the environment possible.

1 Introduction
Recently considerable attention have been paid to face detection and several methods have been proposed [1, 2, 3, 4, 5] because automatic face recognition is regarded as one of the important functions for man-machine interface and face detection is one of the essential steps for face recognition. However, in general, it takes considerable computation time to find a face location in a give image. It is difficult to know how many faces there are and where the faces are in advance. Therefore it is necessary to check all points on the given image. If we select the search points randomly, it is expected to decrease the average numbers of trials of face matching which are necessary to find a face. We can speed up the search further by using the information obtained at the previous search points. The authors have proposed an efficient search method which adopts Ising model to make good use of the likelihood of face obtained at the previous search points [6].

In the applications such as face detection from news images or surveillance system, it is expected that prior probabilities of face locations in the image can be estimated to a certain extent. For example, in the case of news images, the center region gives higher probability than peripheral region because reporters are usually captured at the camera center. To speed up the search, regions with higher probabilities should be checked first.

In this paper, the idea from reinforcement learning [7] is adopted. The prior probabilities of the face locations in the image are adaptively estimated and are used in the Ising search algorithm. The search points in the Ising search are selected depending on the estimated prior probabilities. The information obtained by the previous search points in the given image is effectively utilized through spin flip dynamics of the Ising search. If a face is found, the prior probabilities are updated with forgetting. This makes adaptation to the changes of the environment possible.

2 Face Detection Method
We designed the matching method by taking attention mechanism into consideration [5]. Attention points are selected as the points where the outputs of Gabor filters applied to the contrast filtered image (Gabor features) have rich information. We use the contrast filter [8, 9, 10] which resembles the profiles of retinal ganglion cells. It is well known that Gabor filters resemble the receptive field profiles of the simple cells in V1 area and [11].

The matching is performed by using saliency map which indicates attention candidates. The information of Gabor features extracted from contrast filtered im-
age is used as saliency value. The output probabilities of Gabor filters can be estimated from a set of general images. In this paper 1500 different TV images are obtained by using the cut detection method based on the robust correlation [12]. Then a set of local regions with 9 × 9 pixels are randomly selected from the TV images and the probability distributions of the outputs of Gabor filters are estimated.

Since the outputs of each Gabor filters can be regarded as independent [13], the output probability density \( p(x) \) to the Gabor feature vector \( x = (x_1, \ldots, x_N) \) can be calculated by

\[
p(x) = \prod_{j=1}^{N} p(x_j),
\]

(1)

where \( N \) denotes the number of Gabor filters which have different orientations and \( p(x_j) \) is the output probability density of \( j \)th Gabor filter. Then the information at the point with Gabor feature vector \( x \) can be defined by

\[
I(x) = -\log p(x) = -\sum_{j=1}^{N} \log p(x_j).
\]

(2)

In this paper, the probability density of an output of a Gabor filter is approximated by using normal distribution with mean zero. Thus the probability densities \( p(x) \) and the information (saliency value) \( I(x) \) to Gabor output \( x \) are computed as

\[
p(x) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{x_j^2}{2\sigma_j^2}\right),
\]

\[
I(x) = \sum_{j=1}^{N} \left\{ \log\left(\frac{1}{\sqrt{2\pi}\sigma_j}\right) + \frac{x_j^2}{2\sigma_j^2} \right\},
\]

(3)

where \( \sigma_j^2 \) denotes the variance of output of the \( j \)th Gabor filter and is estimated from the training samples.

The saliency values of all points on the model face image are calculated and those values are used as the weight for matching. The matching similarity is defined as the weighted sum of the correlations of Gabor features extracted from the contrast image. This means that the attention points which have high saliency (information) values have much influence. The similarity measure is defined by

\[
WC = \sum_{i=0}^{M} \frac{I_i}{\sum_{k=0}^{M} f_k} \text{Cor}(i),
\]

(4)

where \( I_i \) is the information of \( i \)th pixel on the model image and \( \text{Cor}(i) \) is the correlation between Gabor features of \( i \)th pixel on the model image and the input image.

To cope with the scale changes of a face on the image, several images with different scales are generated by interpolation from the input image and these images are used as the matching candidates. The best match is searched from all candidates.

This matching method is applied to face detection and face classification. The examples of the face detection results are shown in Figure 1 (d) and (e). It is noticed that faces are detected at correct location and size. The model face image and its saliency map are shown in Figure 1 (a) and (b). The model image is the mean image of over 10 person’s faces. It is noticed that eyes, nose, and mouth have high information values. Since the attention points with low saliency values may have bad influence, the upper 80% are used. The attention points used for matching are shown as white pixels in Figure 1 (c).
cation $i$ in the image, then the probability $h(j)$ at the location $j$ ($\sum_j^M h(j) = 1.0$) is updated by
\[
h(j) = ah(j) + (1 - a) Gw(i, j), \quad (j \in Nn(i))
\]
\[
h(j) = ah(j), \quad \text{(otherwise)} \quad (5)
\]
where $Nn(i)$ is the neighboring regions of detected location (point) $i$, $0 < a < 1$ is the forgetting factor, and $Gw(i, j)$ is the Gaussian weights at the center $i$ ($\sum_{j \in Nn(i)} Gw(i, j) = 1.0$). In following experiment, $\alpha$ is set to 0.99 and $Nn(i)$ to $15 \times 15$ pixels region. By this adaptive estimation, the probability is increased at the points in which faces are frequently detected. On the other hand, the probability is gradually decreased by forgetting factor at the point in which faces are seldom detected.

3.2 Ising Search

Ising model is one of the simplest models of magnetization [14]. It consists of two states of Ising spins: “up” and “down”. The state of a spin depends on both the states of neighboring spins and an external magnetic field. The spin flip dynamics works to minimize a given energy function which is defined by the states of neighboring spins and the external magnetic field. In face detection, there are also two states: “face” and “not face”. Here we set “face” to $-1$ (“down”) and “not face” to 1 (“up”). The measured likelihood of face is integrated into the energy function of Ising model as the external magnetic field. The direction and magnitude of the external magnetic field are changed depending on the measured likelihood of face. If the measured likelihood of face is high, then that direction is toward “down” (“face”) to increase the possibilities to flip the spins of the neighboring regions to “down”. The states of neighboring regions of the selected spin can be estimated through the spin flip dynamics and the search space can be reduced.

The energy function integrated into the measured likelihood of face ($m_d(a)$) of a selected location $a$ are given by
\[
E_i = -J \sum_{j \in nn(i)} s_is_j - H_d(\theta_d - m_d(a))s_i, \quad (6)
\]
where $E_i$ is the energy of the spin $s_i$ which is in the neighboring $5 \times 5$ lattice centered at the selected location $a$, $nn(i)$ represents the nearest neighboring spins of the spin $s_i$, $J$ is the strength of the interaction between spins, $H_d$ is the coefficient of the external magnetic field, $m_d(a)$ is the measured likelihood of face at the location $a$, and $\theta_d$ is the threshold to classify “face” and “not face”, respectively. The direction of the external magnetic field is changed by the threshold.

Then the state of each spin is updated according to the probability which is proportional to $\exp(-\beta\Delta E_i)$. $\Delta E_i$ is the energy change caused by flipping the spin $s_i$, that is,
\[
\Delta E_i = 2J \sum_{j \in nn(i)} s_is_j + 2H_d(\theta_d - m_d(a))s_i, \quad (7)
\]
and $\beta$ is a reciprocal of the temperature. The face candidates (search space) are narrowed down through this spin flip dynamics.

The search point is selected by using the prior probabilities obtained through the exploration to previous images. Among the face candidates, that are the points with “down” spin, one search point are selected with the probability proportional to the prior probability at that point. Then the likelihood of face is measured at the selected point. This search process is repeated until a face is detected.

The algorithm for the search method is given as
1. Set all the spins to $-1$ (face candidate).
2. Select one spin from face candidate depending on the estimated prior probability.
3. Measure the likelihood of face at the selected search point (spin).
4. The spin flip dynamics are applied.
5. Repeat from 2 to 4 until a face is detected (the likelihood of face becomes larger than the threshold).
6. If face is found, prior probabilities are updated.

4 Experiments

The proposed method was applied to news images and the number of search points to find a face was evaluated. Two kinds of TV news programs were captured during about a week. Then news images from different scenes were obtained by using cut detection method based on robust correlation [12]. The size of these images is $240 \times 180$ pixels. To obtain the news images which include faces, we applied the face detection method described in section 2 to all points of the gathered news images (exhaustive search). Then 1849 face images were obtained from the news programs. The number of errors (which classify “not face” regions as “face”) is 15, namely the correct rate of face detection is 99.19% in this exhaustive search.
Figure 2 shows the estimated prior probabilities after 10, 100, 500, 1000, 1500 images are processed. It is noticed that the center regions give high probabilities from this Figure. This is reasonable because only one person is often taken at the center in these news images. The probabilities of the both sides are also relatively high probabilities. This is because two announcers often appear in the same frame. During the interval from 900th image to 999th image, many faces are located at the center of the news images. The center regions are emphasized in the estimated prior probabilities (Figure 2 (d)). After this interval, many faces are also located at right and left side on the news images. During the first 100 trials, it was needed to check about 170 search points until a face was found. But 60 search points were enough to detect a face after 100 trials. This shows the effectiveness of the proposed method.

For the comparison, Ising search without prior probabilities estimation was applied to the same news images. The median of the necessary search points to find a face was 663. The search performance is about 10 times slower than the proposed method which utilizes the estimated prior probabilities. This shows the importance of the adaptive prior probabilities estimation.

References