

圖(一) 原架構



圖(二) 簡化架構

**8.3 The Image Classifier**

In this section, we evaluate the performance of the skin detection algorithm first and then the performance of the algorithm for recognizing pornographic images.

**8.3.1 Skin Detection**

Skin modeling. The choice of the number of Gaussians is important for mixture models. The number of Gaussians used in previous work on skin modeling varies greatly from

1 to 15 [38], [39], but, as yet, no comparisons have been made between mixture models with different numbers of Gaussians, partly due to the complexity of conventional mixture modeling.

In the experiments, we tested our histogram-based mixture models with different numbers of Gaussians using the most popular color spaces, namely, red, green, blue (RGB), Hue Saturation Value (HSV), YCbCr, and normalized r-g. The setup of the test is:

1. The test was carried out on the Compaq skin database [29], which contains about 5,000 images with skin color, including over 80,000,000 million labeled skin pixels.

2. The experiment included a fivefold cross validation. The database was divided into five disjoint subsets with equal size; in each round of the test, one subset was chosen as the test set and the remaining four subsets comprised the training set. The training and test process was repeated for five times and the results were averaged.

3. For each model, there are different false positive rates and different false negative rates associated with different thresholds used to determine whether pixels are classified as skin. For the best choice of the threshold, the false positive rate reaches equilibrium with the false negative rate. Therefore, the Equal Error Rate (EER) (where the false positive rate equals the false negative rate) is used as the performance index.

Table 7 compares the results with a varying number of Gaussians and for different choices of the color space. From Table 7, we have the following interesting discoveries:

1. For all the 3D color spaces, namely, RGB, HSV, and

YCbCr, the Gaussian mixtures generally outperform

the single Gaussian. The EERs decrease overall as

the number of Gaussians increases from 1 to 5. The EERs change very slightly except for minor fluctuations when the number of Gaussians increases from 6 to 10. When the number of Gaussians is more than 10, the performance of the models reaches saturation

and begins to downgrade due to overfitting.

2. For the 3D color spaces, the performance of YCbCr is comparable with that of RGB, as they are linearly related; the HSV space in which the chrominance and luminance information are separated provides the best accuracy in skin detection.

3. The single Gaussian model is not very weak. It is somewhat (instead of greatly) worse than the Gaussian mixtures except in the normalized r-g space.

4. The normalized r-g space, which was thought to be well suited for skin detection due to its robustness to changing illuminations, does not perform as well as expected. The reduction from a 3D color space to a 2D color space plays a role here, as useful information

for skin classification is lost. For the normalized r-g space, there is little benefit in using more than one Gaussian.

Fig. 4 shows the average time spent for training skin models. In Fig. 4, it can be seen that the proposed method is efficient in training skin models. Given the same number of bins, comparable efficiencies are obtained for all the three 3D color spaces, as shown in Fig. 4a. The time spent on training the model with the maximum number of 15 Gaussians is less than 250 seconds. For the normalized r-g space, even less time is needed, as shown in Fig. 4b. Test on synthetic data. Our method was further tested on synthetic data sets to show its efficiency. Random samples were generated from Gaussian mixtures in 3D space, with five Gaussians for each model by default. Fig. 5a shows the speedup ratios of our method compared with the conventional EM algorithm when the number of data points varies from 1,000 to 1,000,000. The dashed line shows the results of histogram bin modeling with a fixed bin size, whereas the solid line shows the results with adaptive bin size. Fig. 5b shows the speedup ratios with respect to different bin sizes (in terms of variance). In Fig. 5, we can see that the efficiency of our method improves considerably when the number of data points or the size of bins increases, and modeling with adaptive bin size brings a further speedup compared with fixed bin size. Table 8 shows the relative errors in log-likelihood values between the estimated model and the ground truth. In Table 8, we can see that both strategies with fixed and adaptive bin sizes approximate the ground truth model well, given that the bin size is not too large. A bin edge size about 1/10 of variance seems to be a good trade-off between

accuracy and efficiency.

TABLE 7

Comparison of Gaussian Mixture Models

8.3.2 Pornographic-Image Recognition

The performance of the contour-based pornographic-image recognition algorithm is evaluated in our previous work [1], when the image plane is partitioned into 4 4 rectangular

blocks (readers may read our previous paper [1] for more

details). In the following, the performance of the contour-based algorithm is compared with the performance of a

connected skin-region-based algorithm.

The connected skin-region-based pornographic-image

recognition algorithms are the most popular in current

research. We implemented a region-based algorithm based

on the work in [21], [22], [23], [24]. Referring to the result of

the comparison, in [21], [22], [23], [24], between contribu-

tions of the different features of the connected skin regions,

the following features are selected for classifying images:

. ratio of the area of skin regions to the area of the image,

. average skin probability of all pixels in the image,

. number of connected skin regions,

. area ratio of the largest skin region to all skin regions,

. image width, and

. image height.

In real situations, images of women wearing bikinis or swimsuits or close range images of faces are often mistakenly classified as pornographic. Three sets of images are constructed: a set of pornographic images, a set of images of women wearing bikinis, swimsuits, and so forth (legitimate sex-related images), and a set of face images, where each set contains 2,000 images. Half of the images in each set are used as training samples and the remaining images as test samples. Table 9 shows the comparison between the correct classification and false acceptance rates of the contour-based algorithm and the correct classification and false acceptance rates of the region-based algorithm, tested in the three sets of images. In Table 9, it can be seen that, compared with the contour-based algorithm, for the pornographic images and the legitimate sex-related images (the images of women wearing bikinis, swimsuits, and so forth), the correct classification rates of the region-based

algorithm are only slightly decreased, but the false