Portrait Image Clipping with Rough Face Designation

ユーザの簡易指定を伴うポートレート画像切り抜き

船山竜士*　竹澤　創*　紺矢峰弘*　斗谷充宏*
Ryuji Funayama　Hajime Takezawa　Minehiro Konya　Mitsuhiro Hakaridani

Abstract

This paper describes a method of portrait image clipping for mobile information tools with poor computational power and easy handling interactive devices like Zaurus. User's designation of rough position and size of a face enables robust and fast clipping. Adaptive face region segmentation and the subsequent process of detecting face parameters make determination of clipping area. We have experimentally proven that our method has a good computational performance and is robust to variations in backgrounds, lighting conditions, individuals and so on.

本論文は，ザウルス等の携帯型情報機器のように，高い処理能力を持たないがペンのような簡易な入力装置を持つツールにおいて，ポートレート画像切り抜きを行う手法について述べている。ユーザによって顔のおおよその位置と大きさが与えられることで，ロバストで高速な切り抜き処理を実現する。適応的顔領域分離により顔肌領域を抽出し，その内部のみを処理することで背景の影響を受けずに，切り抜き矩形を決定するための顔パラメータを求めることができる。実験により，本手法は高速かつ，背景・照明条件・個人差等に対してロバストであることが確認された。

Introduction

Going with expanding market for personal computers and digital still cameras, there exists little barrier to get and edit digital images. In addition, mobile personal information tools (mobile tools) which have easy handling interactive device like stylus pen are ready to equip a CCD camera.

In general, most of the images which are captured by such equipments usually contain persons, and many applications based on human face image processing are considered promising. A main function of mobile information tools is a personal information management mechanism which usually appears in forms of address book or something. Until today, most of the information is stored in text. If those tools have an image capturing device, the technology which enables users to handle face images easily would be required because face is one of the most important personal information. To make a good use of a face image, a mechanism is required which can extract a face region from an image that contains people in it.

There are not many pieces of related works whose purpose is to clip an image which contains a portrait well-balanced in it. It is usual to use an image editing tool or a photo retouch tool to clip portrait images manually. The well-balanced portrait means that (1) horizontal position of a face, (2) vertical position of a face and (3) size of a face, have a proper value in the clipped image. Adjusting these parameters manually using the above image/photo tools is very difficult and requires a skill, especially if a face is near the edge or corner of an image or covers the most area of an image.

There are a number of works for face detection[1]. Typical works are using template matching[2], eigenfaces[3], color information[4], deformable models[5] and so on[6)7]. Most of the works are for preprocessing for personal identification or face tracking. The purpose of our work is to develop a method to clip a well-balanced portrait image from an input image on mobile tool which usually have poor computational power. So existing methods mentioned above do not satisfy both computational power requirement and detection accuracy of face position.

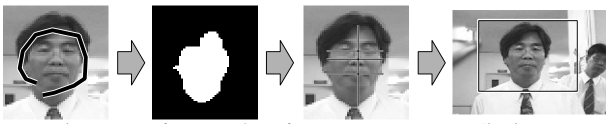
In this paper, we propose a portrait image clipping method which can be applied on mobile tools whose computational performance is poor but pointing device can be handled easily. After a face in an image is designated roughly using such device as stylus pen, a rectangular region which contains portrait well-balanced in it is clipped from the original image. We explain the detail of the proposed algorithm and experimental results in the

* 情報システム事業本部　情報商品開発研究所

following sections.

1. About Portrait Image Clipping with Rough Face Designation

We work out the following strategy to achieve portrait image clipping. First, we utilize an easy handling pointing device like stylus pen which is a distinctive feature of mobile tool. Position and size of a face are obtained through the interactive device. We then segment a face region robustly using the above information. Face parameters are derived by processing an input image inside a face mask which is made from the segmented image. Finally, a clipping area is found out based on the derived face parameters (see **Fig. 1**).



user's input　face mask　face parameters　clipping area

Fig. 1 General flow of portrait image clipping.

2. Rough Face Designation by Simple User Interface

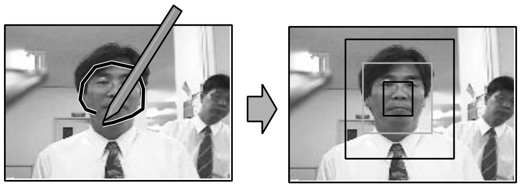
In this paper, we have the assumption that target machine is a mobile tool. We utilize an easy handling pointing device like stylus pen equipped by mobile tool because of restrictions on computational power and requirements for detection stability.

Pen type input device is suitable for pointing any spot within an image and drawing any curve. So user can input position and size of a face using this device. However, it is difficult to designate precise position and size of a face, and is not user-friendly to require to do that. Therefore, as shown in **Fig. 2(a)**, by roughly surrounding a face with pen drawing, approximate position and size are obtained, and these are used for the subsequent process. User's designation has the following merits:

(1) Make it possible to segment face region adaptively,

(2) Improve performance by limiting a processing area.

Fig. 2(b) shows a sampling area (small rectangle) and a target area (large rectangle). The former is used to define a face skin region extraction function, and the latter is a target rectangle area in which all the subsequent process will be done. Each region is come out to scale-down or scale up a circumscribed rectangle of user's designation pattern respectively. Adaptive face skin region segmentation method utilizes information about color distribution analysis of the sampling area. This is described in detail in section 3. The resolution of the target area is reduced to improve computational performance and eliminate an influence of noise.



(a) user's designation by pen　(b) sampling and target area

Fig. 2 Sampling and target area from user's designation.

3. Adaptive Face Region Segmentation and Face Mask Image

3•1 Face Region Segmentation Using Color Information

There exists a face region segmentation method which applies a face region extraction function to an input image[5]. This function is defined by analyzing color distribution of face skin in the training set of images in a specific color space. The method has a number of varieties of using different color spaces or different fitting functions[4)8)9)].

The segmentation method based on color information is usually easy to implement and can be efficiently performed. But it is difficult to apply it to practical applications directly because it is not robust to differences of face color in appearance caused by lighting condition, individuals and races or complex backgrounds. Then, the face skin region segmentation function which is suitable for the input image is interactively defined by representing color distribution of the sampling area shown in **Fig. 2(b)** as the distribution of the face focused there. This will enable robust segmentation of a face with less load to users[10].

3•2 Adaptive Face Region Segmentation with User's Designation

In our former research[5], it is shown that hue and saturation of face skin color go with normal distributions (see **Fig. 3**). A face skin segmentation function can be derived from the both distributions.

However, the probability density function is a representative from face skin color of more than one person. **Fig. 4** shows the results of applying the function learned from training set of images to unknown images. You can see face regions are not segmented correctly or

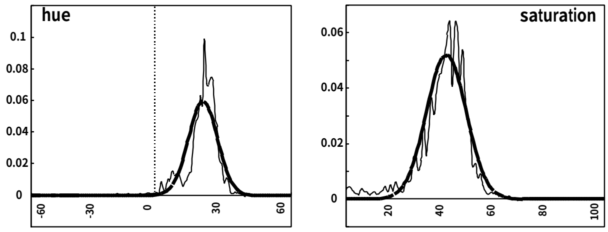


Fig. 3 Distribution of hue and saturation in face.

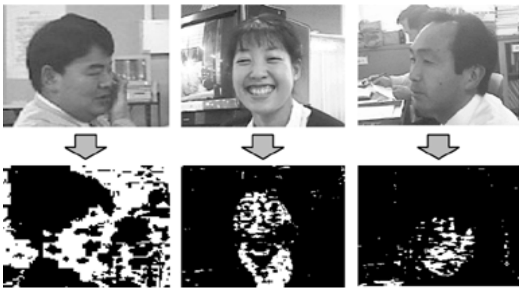


Fig. 4 Unsuccessfully segmented face region.

backgrounds are needlessly segmented.

Then color distribution within the sampling area is investigated as mentioned in section 2. Same as in the case of training set of images, new probability density functions are defined as follows by calculating mean and variance of both hue and saturation ($\mu_{hue}*$, $\sigma^2_{hue}*$, $\mu_{sat}*$, $\sigma^2_{sat}*$).

$$P_{hue}(h) \sim N(\mu_{hue}*, \sigma^2_{hue}*)$$
$$P_{sat}(s) \sim N(\mu_{sat}*, \sigma^2_{sat}*)$$

To apply the above terms $P_{hue}(h)$ and $P_{sat}(s)$, to the following formula, a face region segmentation function which is suitable for the input image can be obtained as:

$$I(p)=\begin{cases} 1 (\text{if } P_{hue}(hue(p)) > PH_{th} \text{ and } P_{sat}(sat(p)) > PS_{th} \text{ and } \\ \quad val(p) > V_{th}) \\ 0 \text{ (otherwise)} \end{cases}$$

where $hue(p)$, $sat(p)$ and $val(p)$ are hue, saturation and intensity value at pixel p, and PH_{th}, PS_{th} and V_{th} are proper thresholds. $I(p)$ is a binary image which discriminates face skin from other regions. **Fig. 5** shows correctly segmented face images from the input images in **Fig. 4** The proposed method can segment face region robustly, and it does not depend on face color in appearance based on lighting conditions, camera systems, individuals and races[10].

Fig. 5 Correctly segmented face region.

3•3 Face Mask Image

A face mask image is generated through the following four steps (see **Fig. 6**).

(1) Generate a face segmented image from an input image,
(2) Eliminate all regions except the region of maximum area,
(3) Plug up holes in the above region,
(4) Expand and contract the region.

After a contour of the region is smoothed by expansion-and-contraction operations (expanding→contracting→expanding), finally a face mask image is obtained.

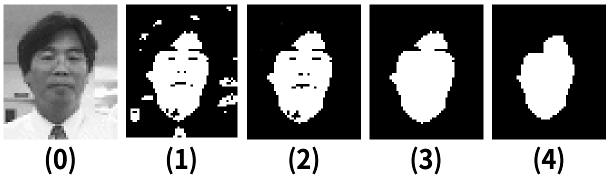


Fig. 6 Generation of face mask image.

4. Detection of Face Parameters

The following parameters are required for image clipping:

(1) an axis of symmetry in face,
(2) vertical position of eyes, nose and mouth,
(3) width of face.

In the clipped image, horizontal and vertical position of a face is placed in the center of the image and size of a face is regularized in proper balance. It is difficult to find out the above parameters correctly so that these values are estimated from above mentioned parameters (1)~(3).

4•1 Detection of Axis of Symmetry in Face

Here we consider an axis of symmetry in a face as horizontal center of a face. A rectangle image is clipped as the axis overlaps with the center of the clipped image. Owing to that, important parts in a face such as eyes, nose and mouth are placed in the center of the image. An axis of symmetry in a face is detected in the following way (see **Fig. 7**):

Step-1: Convert an input color image to a gray scale image,
Step-2: Differentiate the gray scale image vertically,
Step-3: Project the differential image vertically within a region which correspond to the face mask,
Step-4: Search an axis which minimizes SSD of frequency value across the axis in the projection histogram.

In general, spatial pattern of a face is symmetrical with

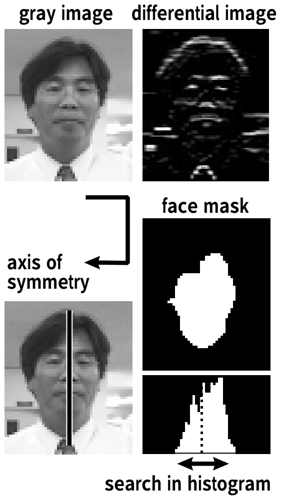


Fig. 7 Axis of symmetry in face.

an axis which penetrates the center of both eyes and mouth, that is caused by shape and placement of face components. In step-2, differentiating a gray scale image vertically makes a shape of the histogram also symmetrical by emphasizing the face components. If a gray scale image is differentiated directly, a shape of the histogram will not be symmetrical because of light and shade on the face.

In step-3, limiting projection target as inside face mask prevents complex background from deforming the shape of the histogram. Finally obtained axis is considered as an axis of symmetry in face.

4•2 Detection of Vertical Position of Face Components

Detecting a vertical position of eyes, nose and mouth makes it possible to estimate a vertical center of a face and detect the width of a face which is used to approximate the size of a face. As mentioned before, it is important to implement the algorithm as a fast system. We thus adopt a projection based approach. As images are projected horizontally and face components are detected from the shape of the projection histogram, some rotation of a head does not have influence on the detection efficiency because the shape of the histogram does not deform widely with such rotations.

4•2•1 Detection of Vertical Position of Nose

Vertical position of nose is detected in the following way (see **Fig. 8**):

Step-1: Convert an input color image to a gray scale image,

Step-2: Project the gray scale image horizontally inside face mask,

Step-3: Reduce the resolution of the projection histogram and find out a central peak.

The histogram obtained from projecting a gray scale image horizontally has large three peaks corresponding to forehead, nose and chin (and neck). So a central peak in low-resolutionized histogram is considered as the neighborhood of nose. Upper side from nose is for detecting position of eyes and lower side is for detecting position of mouth. The way to detect vertical position of eyes and mouth described below is the same as detecting nose in this section up to step-2.

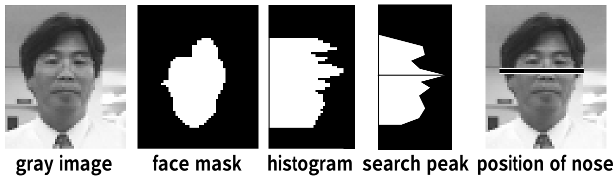


Fig. 8 Vertical position of nose.

4•2•2 Detection of Vertical Position of Eyes and Mouth

The vertical position of eyes is searched in the histogram upper side from the nose detected above. Step-3 is replaced by the following:

Step-3: Probe projection histogram upper side from the nose and select the lower valley position from first and second deepest valley as a vertical position of eyes.

In general, the histogram has deep valleys at eyes and eyebrows, and the lower valley is considered as eyes (see **Fig. 9(a)**).

The vertical position of a mouth is searched in the histogram lower side from the nose. Step-3 is replaced by the following:

Step-3: Probe projection histogram lower side from the nose and select the middle valley position from first to third deepest one as a vertical position of mouth.

In general, the histogram has deep valleys at the bottom of nose, mouth and chin, and the middle valley is considered as mouth (see **Fig. 9(b)**).

The vertical position of each face component is detected

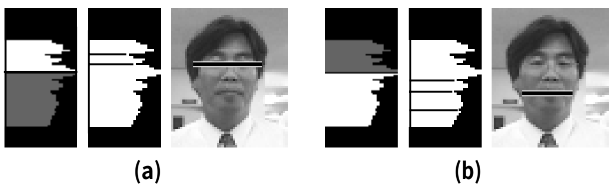


Fig.9 Vertical position of eyes and mouth.

through the above mentioned process. In practice, it is required to take steps to cope with the case that the number of peaks and valleys in the histogram is not the assumed one .

4•3 Detection of Width of Face

The width of a face is detected from a face mask image. If we consider the maximum value of width of a face mask as the width of a face, it is estimated wider because of ears and so on. We then apply the following approach:

Step: Measure the width of a face mask at height between eyes and mouth and consider the median value as the width of the face.

There is less fluctuation on width of face between eyes and mouth and the influence of ears is eliminated by adopting median value.

5. Portrait Image Clipping

Now face parameters, (1)axis of symmetry in a face, (2)vertical position of eyes, nose and mouth and (3)width of a face, are detected through the above process. Then face clipping rectangle is defined by those parameters.

First, a face is placed in the clipped image as the detected axis of symmetry in face overlaps with the horizontal center of the image. That makes face components such as eyes, nose and mouth put near the center of the clipped image. Detected vertical position of nose is overlapped with the vertical center of the image.

After deciding the center position of the clipping rectangle, size must be solved. It depends on the value of width of face and height between eyes and mouth. As is in our first experiment, it only depended on the height between eyes and mouth. But the ratio of the height between eyes and mouth to height of the face (height between the top of the head and chin) has individual variation, and it makes clipped image varied in size. We then made it by width of a face. This resulted in good clipping but a face image which dose not looking front was clipped in a bit small rectangle because width in appearance tended to detect wider. On the other hand, height between eyes and mouth does not depend on face direction.

Finally, we decided that clipping size should be computed from weighted geometrical average. Here in our method, clipping size mainly depends on width of a face which is stable for front looking face, and it makes possible that faces without looking front are clipped successfully by giving a little weight on value of height between eyes and mouth which does not depend on the direction of a face. We experimentally set the ratio of width to height as 3:1.

6. Experiments and Results

To demonstrate the performance of the proposed algorithm, we have experimented on a number of images. Aspect ratio of clipped image is set to 4:3 in crosswise to lengthwise. **Fig. 10** shows the results which contain images with user designation patterns, detected face parameters and clipped portrait image (whose size is regularized). If a face is placed near the edge of an input image, a part of the clipped image corresponding to the exterior of the input image is filled with gray pixels and it also results in aspect ratio of 4:3. Portrait image is clipped correctly regardless of backgrounds, lighting condition, direction of face, face size in appearance and so on.

Fig. 11 shows incorrectly clipped images. It is caused by the following reasons:

- Incorrect face mask is obtained because of extremely bad color reproduction,
- Incorrect face mask is obtained because of strong light and shade on a face,
- Face mask is distorted because of occlusion,
- Indistinct projection histogram is obtained because of drooped face,
- Indistinct projection histogram is obtained because of small face in appearance.

The number of images experimented here is 105, which can be divided to two groups:

(A) uniform background + fixed lighting + frontal face ... 33 images,

(B) complex background + irregular lighting + frontal and non-frontal face ... 72 images.

Tables 1 and **2** show the performance of portrait image clipping on the image sets (A) and (B), respectively. Here we consider the correct clipping as what the reference face rectangle region selected manually is placed in the center of the clipped image and have proper size. The tables show the horizontal and vertical position of clipping area detected correctly in high reliability but size correctness results in 60 to 70 percent. The image set (A) which consists of regular conditioned images also results in the similar ratio at “size” because the weight for geometrical average mentioned before is adjusted to make every images include non-frontal face clipped correctly. So in the case of the image set what all is frontal face, results are influenced

Fig. 10　Correctly clipped portrait image.

Fig. 11　Incorrectly clipped portrait image.

Table 1 Results on image set (A).

	correct	incorrect	correct ratio
horizontal	33	0	100%
vertical	31	2	93.9%
size	23	10	69.7%
total	21	12	63.6%

Table 2 Results on image set (B).

	correct	incorrect	correct ratio
horizontal	72	0	100%
vertical	66	6	91.7%
size	45	27	62.5%
total	44	28	61.1%

by the height between eyes and mouth as a whole.

Fig. 12 shows some screen shots of PowerZaurus(MI-500/600 series), a CCD camera attachable mobile information tool, which employs the proposed portrait image clipping method. This application is an address book on which user can paste a portrait image as an individual information. Processing time on 32bit RISC CPU (30MHz) per one clipping is less than 0.8 sec. at any size up to VGA-size (640×480 pixels) of an input image.

Fig. 12 Application on PowerZaurus.

Conclusion

We have proposed a portrait image clipping method using easy handling interactive device like stylus pen equipped by mobile information tool. The portrait image clipping is an application that clips a rectangle region in which a face is placed in proper balance from an input image. This method is proven to be robust through the experiments with variations in backgrounds, lighting conditions, direction of face and so on.

The proposed method produces a fast and robust portrait image clipping with help of user's designation of rough position and size of a face. We will improve this method that image can be clipped with easier designation.

Acknowledgements

The authors would like to thank Professor Yokoya at NAIST(Nara Institute of Science and Technology) for many helpful suggestions and comments on making this manuscript. The authors also thank Mr.Sakata, General Manger of our laboratories.

References

1) R. Chellappa, C. L. Wilson and S. Sirohey,“ Human and Machine Recognition of Faces: A Survey ”, Proc. IEEE, Vol.83, No.5, pp.705-740 (1995).

2) 小杉,“ 個人識別のための多重ピラミッドを用いたシーン中の顔の探索・位置決め ”, 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.4, pp.672-681 (1994).

3) M. A. Turk and A. P. Pentland,“ Face Recognition Using Eigenfaces ”, Proc. CVPR '91, pp.586-591 (1991).

4) Q. Chen, H. Wu and M. Yachida,“ Face Detection by Fuzzy Pattern Matching ”, Proc. 5th ICCV, pp.591-596 (1995).

5) R. Funayama, N. Yokoya, H. Iwasa and H. Takemura,“ Facial Component Extraction by Cooperative Active Nets with Global Constraints ”, Proc. ICPR '96, Vol.2, pp.300-305 (1996).

6) R. J. Qian and T. S. Hung,“ Object Detection Using Hierarchical MRF and MAP Estimation ”, Proc. CVPR '97, pp.186-192 (1997).

7) M. S. Lew and N. Huijsmans,“ Information Theory and Face Detection ”, Proc. ICPR '96, Vol.3, pp.601-605 (1996).

8) E. Saber, A. M. Tekalp, R. Eschbach and K. Knox,“ Automatic Image Annotation Using Adaptive Color Classification ”, GMIP, Vol.58, No.2, pp.115-126 (1996).

9) C. H. Lee, J. S. Kim and K. H. Park,“ Automatic Human Face Location in a Complex Background Using Motion and Color Information ”, PR, Vol.29, No.11, pp.1877-1889 (1996).

10) 船山, 竹澤, 紺矢, 斗谷,“ ユーザーの簡易指定を伴う顔領域の適応的セグメンテーション ”, 情報処理学会第55回全国大会, 6AB-3 (1997).

11) R. Brunelli and T. Poggio,“ Face Recognition: Features versus Templates ”, IEEE Trans. PAMI, Vol.15, No.10, pp.1042-1052 (1993).

(received June 2, 1998)