ENHANCEMENT OF FACE DETECTION USING SPATIAL CONTEXT INFORMATION

*Hyun-Seok Min, Young Bok Lee, Sihyoung Lee and *Yong Man Ro*

Image and Video Systems Lab.
Information and Communications University
Daejeon, Korea
E-mail: yro@icu.ac.kr

ABSTRACT

Significant attention has recently been drawn to digital home photo albums that use face detection technology. The tendency can be found in home photo albums that people prefer to allocate concerned objects in the center of the image rather than the boundary when they take a picture. To improve detection performance and speed that are important factors of face detection task, this paper proposes a face detection method that takes spatial context information into consideration. Experiments were performed to verify the usefulness of the proposed method and results indicate that the proposed face detection method can efficiently reduce the false positive rate as well as the runtime of face detection.

Keywords: face detection, spatial context information, region of interest

1. INTRODUCTION

Face detection is a technology which finds the positions of faces which are present and determines the dimensions of faces on digital image. It is used for home photo management and search in home photo album. And recent digital cameras for home photos often use face detection technology to detect and focus faces automatically. However, there are still many problems existing in the face detection area [1].

The most conventional face detection methods have applied classifier such as neural network and Support Vector Machine (SVM) [2][3]. These approaches show acceptable performance and they are used in many face detection systems. However, those classifiers are computationally expensive, and need to be trained with many non-face training data. Some conventional face detection approaches use template matching [4][5] and facial features such as eyes, nose, mouth, and skin color [6][7]. Although these approaches with template and facial features do not require training steps with many non-face training data, it is difficult to find facial features in a diverse environment.

The number of faces in an image is usually small. Therefore, in the face detection procedure, the runtime for checking the non face region is much larger than the runtime for the face region, which resulted many conventional approaches to attempt to reduce the runtime for the non-face region in order to improve the speed of face detection. Some conventional face detection systems have applied facial characteristics such as human skin color and motion information to reject the background region and other objects containing non face [8]-[10]. While these methods can reduce the runtime of face detection by rejecting the non-face region with a simple feature, these methods are limited and are conditional approaches. Color information such as human skin color is not invariant to the illumination conditions. Also, approaches involving skin color can reject positive face regions due to hair and accessories on face such as eyewear and caps. Background rejection with motion information cannot be used for a still image. Camera motion and motion of other objects can affect the quality of performance.

Home photos are taken by ordinary people, as opposed to those taken by professional photographers for commercial objects. Ordinary people usually arrange concerned object in the center of the scene to distinguish the object from other objects and to accentuate objects in the image. Therefore the spatial context information about the spatial distribution of concerned objects is more likely to exist in home photos than other images. This tendency also exists when people take a picture containing faces. Thus, faces in the image tend to be located on the center rather than the boundary of the image. Taking above factors into consideration, we propose a face detection method using the spatial context information. Specifically, we verify the inclination of arranging face in the center of the scene and extracted spatial context information about the region where faces are mainly allocated on the image. Also, we apply the spatial context information on face detection to improve the performance of the face detection. Experimental results show that the face detection method using spatial context information does improve the performance and reduce the runtime.

The rest of the paper is organized as follows. Firstly, we present the extraction of spatial context information in section II. In section III, we describe the proposed method that uses the spatial context information. In section IV, we show the performance evaluation. Finally, we conclude the paper in section V.

2. SPATIAL CONTEXT INFORMATION

Correspondence: Prof. Yong Man Ro; Email: yro@icu.ac.kr;
Image and Video Systems Laboratory, Information and Communications University, Munji 103-6, Yuseong, Daejeon, 305-732,
South Korea, Phone: +82-42-866-6129, Fax: +82-42-866-6245

This section covers spatial context information that can exist in an image and extraction process of spatial context information.

2.1 Spatial Context Information

Image includes not only the concerned objects but also background and other objects. However, people are inclined to allocate concerned objects in the center of the scene rather than the outline of the scene when they take a picture. As shown in Fig.1, objects in the center of the scene stand out.

Fig. 1: Images with concern objects located in the center

Home photos are typical photos taken by normal people to record their lives, as opposed to those taken by professional photographers for commercial purposes. Therefore, especially, the spatial context information is more likely to exist in home photos than other images. The spatial context information also exists when people take a picture containing faces as shown in Fig. 2. Unlike pixel based information such as skin color and motion, the spatial information is independent on other information in the image such as illumination and motion of other objects. Therefore, the spatial context information can improve face detection performance without affecting pixel based information.

Fig. 2: Facial image with faces located in the center

2.2 Spatial Context Information Extraction

To use this information for face detection, we should verify the spatial context information. In order to verify the spatial context information, we analyze spatial distribution of faces on home photo images in MPEG-7 database which contains abundant actual images. We use 669 images in MPEG-7 database of which contained 1191 faces in total.

Due to the fact that the size of the images is not uniform, we use histograms to observe the distribution of faces in a database. When a new image for an observation is presented, we divide the images into 400 regions with the same size. As shown Eq. 1, we determine the region value for observed face region distribution. In Eq. 2, let *face region* be the region containing a face. If each divided region includes some face region, the value of corresponding *Bin(i,j)* is 1; otherwise, *Bin(i,j)* is 0.

$$Bin(i,j) = \begin{cases} 1, & if \quad region(i,j) \cap face\ region \neq \phi \\ 0, & otherwise \end{cases}$$
$$where \quad 1 \le i \le 20, \ 1 \le j \le 20 \qquad (1)$$

To recognize the spatial distribution of faces included in all images of database, we evaluate the sum of Bin(i,j) as the following equation.

$$SumofBin(i,j) = \sum_{All\ image} Bin(i,j) \qquad (2)$$

Fig. 3 shows the distribution of values of SumofBin(i,j), which implies the spatial distribution of faces in images. As shown in Fig. 3, regions containing face are largely located in the center rather than the boundary of the images on home photos. This confirms that ordinary people are inclined to allocate concerned faces in the center of the scene when they take a picture.

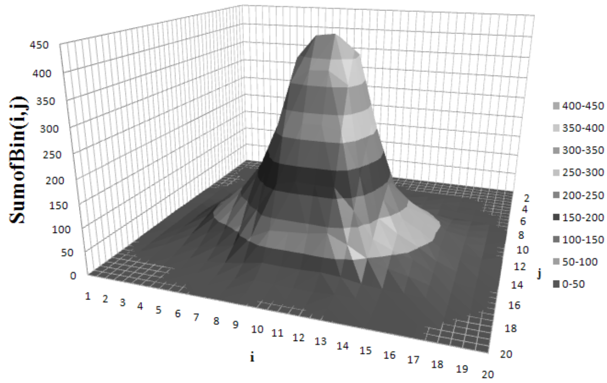


Fig. 3: Distribution of faces in images

To apply the verified spatial context information to face detection, we should know the areas of faces that are largely allocated in the image. In this paper, we define the 'region of interest (ROI)' to use the spatial context information. ROI is the region where people mainly allocate faces when they take a picture.

As shown in above graph, the distribution of faces around horizontal-axis and vertical-axis follow Gaussian distribution. Mean and standard deviation of the accumulated value of horizontal axis are 11.63 and 2.44. And mean and standard deviation of the accumulated value of vertical axis are 10.53 and 2.85. In this paper, we determine the ROI with 99% confidence interval. Defined ROI is shown in Fig. 4. ROI in the image is an area where faces are mostly concentrated and distributed, and is 10.4%, 26.7%, 15.9% and 10.6% off from the top, bottom, left and right of the image respectively. The size of the ROI is 46% of the original image size.

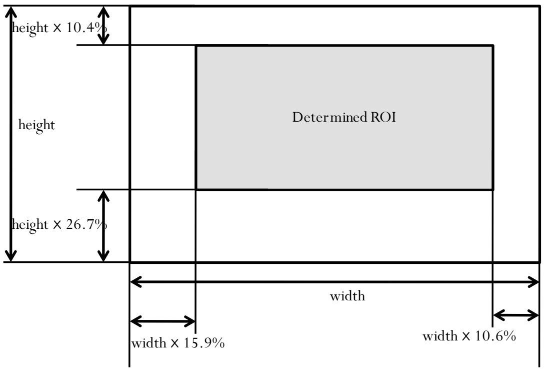


Fig. 4: Determined region of interest model

3. PROPOSED FACE DETECTION METHOD

We determined the region of interest with the spatial context information. Most faces are allocated on the ROI rather than on the Non-ROI. On the other hand, the Non-ROI has many background and non-face regions which can cause false alarms. This means that the ROI and Non-ROI have different characteristics from each other. However, the conventional face detections perform through entire image at once. Therefore, it is important to apply different face detection methods on the ROI and Non-ROI according to the characteristics of regions.

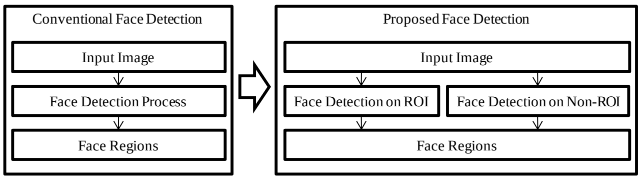


Fig. 5: The proposed Face Detection Method

3.1 Face Detection using Candidate Regions

In this paper, the face detection using candidate regions is used to apply spatial context information for face detection [6], [7]. The candidate region is a rectangular sub-region on the image. Face detection using candidate region determines that each candidate region includes face region or not. The process of the proposed face detection using candidate regions is described in Fig. 6.

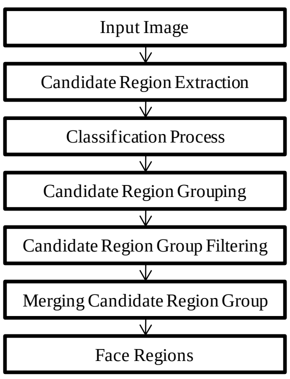
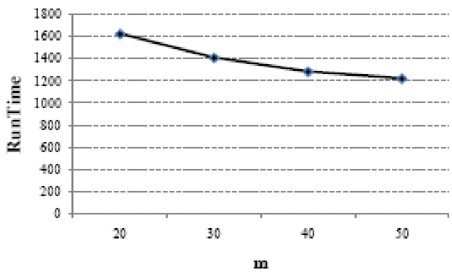


Fig. 6: Process of face detection using candidate regions

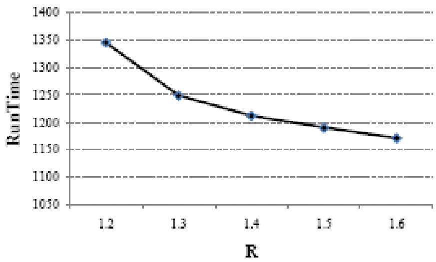
As shown above, when a new image is presented, the candidate regions of diverse sizes and positions are extracted. The classifier determines whether each candidate region includes a face region or not. Candidate regions can contain large amount of regions which overlap each other. After the classification, the candidate regions that include large amount overlapping regions are clustered into candidate region groups. Because each positive face region actually generates many hits from classifier with rectangle feature, candidate region group are filtered by the number of candidate region that belong to a candidate region group. After the filtering step, each candidate region group is merged into a single face region.

3.2 Criteria determining the characteristics of face detection

Face detection method using candidate regions have some condition values which determine characteristic of face detection [7]. We have two condition values to decide the number of candidate regions: minimum size of candidate region (*m*) and scale increase rate of candidate region (*R*). Fig. 7 shows the runtime of face detection for 669 images with respect to m and R.



(a)



(b)

Fig. 7: Runtime of face detection for 669 images

(a)Runtime with respect to m (b)Runtime with respect to R

As shown in Fig. 7, the runtime of face detection process depends on m and R. If either m or R is decreased, the runtime of face detection using candidate region is also decreased. Since each candidate region undergoes classification process, the runtime of face detection process depends on the number of candidate regions. Given m and R, we have the number of candidate regions (*NumOfCR(m,R)*) as shown in Eq. 3 where, *W* and *H* are the width and height of image respectively.

$$NumOfCR(m,R) = \sum_{k=0}^{N} (W - m \times \lfloor R^k \rfloor + 1)(H - m \times \lfloor R^k \rfloor + 1)$$
$$where \quad N = \left\lfloor \min\left(\log_R \frac{W}{m}, \log_R \frac{H}{m}\right) \right\rfloor \tag{3}$$

As shown Eq.3, *m* and *R* determine the number of candidate regions. Therefore, the runtime of face detection using candidate regions is dependent on *m* and *R*. If either *m* or *R* is decreased, the number of candidate regions is decreased as well as false positive rate. However, true positive rate is also decreased. On the other hand, if m and R are increased and the number of candidate regions is increased, then not only true positive rate but also runtime and false positive rate are increased. Therefore *m* and *R* are the criteria that determine the characteristics of face detection.

Generally, each positive face region generates many candidate regions that contain the face region. Thus, many candidate regions which include positive face region are classified into face region by classifier with rectangle features. The positive face region itself generates large cluster of candidate regions. Therefore, we discard candidate region groups with smaller amounts of candidate regions than the *Min Neighbors* threshold. *Min Neighbors* is the threshold value for filtering candidate groups. If *Min Neighbor* is increased, false positive rate is decreased. However, true positive rate is also decreased. *Min Neighbors* operates on true positive rate and false positive rate.

3.3 Proposed Face Detection

There are some criteria such as *m*, *R* and *Min Neighbors* for face detection which determine the characteristic of face detection. These criteria adjust the true positive rate, false positive rate, and running time. Fig. 8 illustrates the process of the proposed face detection method which uses ROI information.

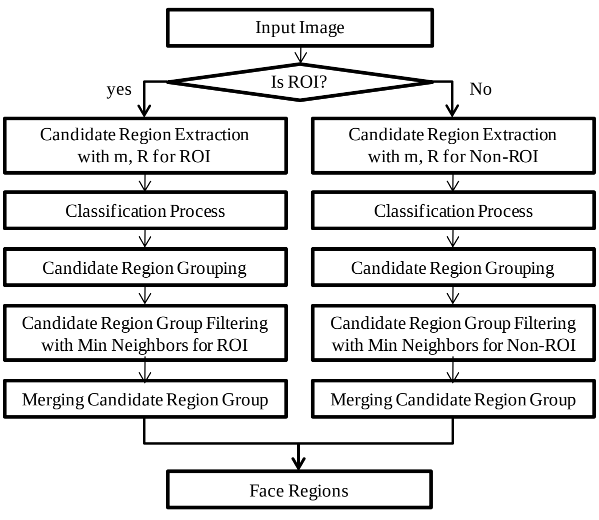


Fig. 8: Process of proposed face detection method

In the proposed system, we applied ROI information on the face detection using candidate region. Most faces are allocated on ROI rather than on Non-ROI. On the other hand, Non-ROI has many background and non-face regions which can cause false alarms. *m* and *R* control true positive rate, false positive rate and runtime of system by determining the number of candidate regions. And *Min Neighbors* operates on true positive rate and false positive rate. However, if we set the values of criteria to improve true positive rate, false positive rate and runtime are also increased.

To solve this problem, the proposed method sets different values of criteria for the ROI and Non-ROI. Most faces are located on the ROI rather than Non-ROI. Therefore, if the criteria are determined to lower false positive rate, true positive rate should be decreased more in the ROI than in the Non-ROI. On the other hand, Non-ROI has many background regions and regions with non face which can cause false alarms. Likewise, if the criteria are determined to make higher true positive rate, false positive rate should be increased more in the Non-ROI than ROI. Thus, we set the criteria for ROI to improve true positive rate, and set criteria for Non-ROI to reduce false positive rate. To apply different characteristics on the ROI and Non-ROI, we use different criteria value such as *m*, *R* and *Min Neighbors* on the ROI and Non-ROI.

As shown in Fig. 8, when a new image is presented, the proposed face detection system extracts the candidate regions on ROI and Non-ROI with different criteria. And the proposed method performs face detection procedure with different criteria on ROI and Non-ROI.

4. EXPERIMENTAL RESULTS

To verify the performance of the proposed method, we use 669 images in MPEG-7 database that is different to images for determining ROI. There are 1161 faces in 669 images for experiments in total. And we use the open source for comparable experiments with conventional face detection [11]. The experimental environment is CPU 2.40GHz. And we use two terms to evaluate face detection performance: TP rate (True Positive Rate) and FP rate (False Positive Rate). TP rate and FP Rate are described as following equations. In equation 4, 5, N_{True} is the number of faces in images, N_{TP} is the number of positive detections, and N_{FP} is the number of negative detections. Table 2 shows criteria values for face detection.

$$True\ Positive\ Rate = \frac{N_{TP}}{N_{True}} \tag{4}$$

$$False\ Positive\ Rate = \frac{N_{FP}}{N_{TP} + N_{FP}} \tag{5}$$

Criteria values for experiments are described in Table 1. In Table 1, *m* is the minimum size of candidate region; *R* is the scale increase rate of candidate region. And *Min Neighbors* means threshold value for candidate region group filtering.

Table 1: Criteria values for ROI and Non-ROI.

Criteria	m	R	Min Neighbors
Criteria for ROI	20	1.1	3
Criteria for Non-ROI	50	1.2	6

To verify the usefulness and performance of the proposed method, we perform face detection separately with criteria for ROI and Non-ROI. And we compare the results of them with the results of the proposed method. Table 2 shows the result of face detection with criteria for ROI.

Table 2: Face detection with criteria for ROI

N_{TRUE}	N_{TP}	N_{FP}	TP rate	FP rate	Runtime
1161	1123	162	0.967	0.126	1650

As shown in above table, face detection with criteria for ROI shows high TP rate, but FP rate is also high. And it takes 1650 sec to evaluate the face detection on 669 images. Table 4 shows the result of face detection with criteria for Non-ROI.

Table 3: Face detection with criteria for Non-ROI.

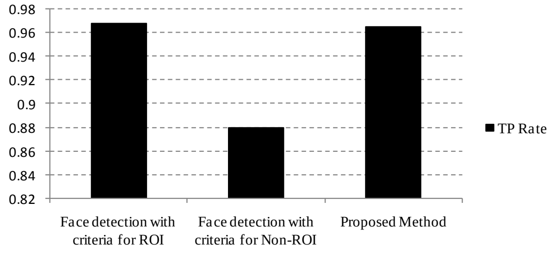
N_{TRUE}	N_{TP}	N_{FP}	TP rate	FP rate	Runtime
1161	1021	3	0.879	0.00293	870.1

Result of face detection with criteria for Non-ROI shows low FP rate while runtime is decreased. However, TP rate is also decreased almost 10%. If we perform uniform criteria for whole image, TP rate has trade-off relationship with FP rate and runtime. Table 4 shows result of the proposed method which uses both criteria for ROI and Non-ROI.

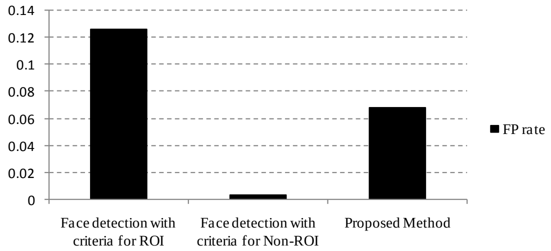
Table 4: Proposed method

N_{TRUE}	N_{TP}	N_{FP}	TP rate	FP rate	Runtime
1161	1120	82	0.964	0.0682	961.2

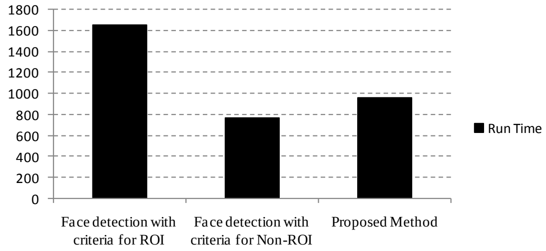
In our proposed method, TP rate and FP rate are 96.5% and 6.82% respectively. And runtime is 961.2 sec. TP rate of the proposed method is almost equal to TP rate of face detection with criteria for ROI. FP rate and runtime are decreased almost 50% when compared to results of face detection with criteria for Non-ROI. Fig. 9 shows the results of face detections with the criteria for ROI and Non-ROI and the proposed method.



(a)



(b)



(c)

Fig. 9: Result

(a)TP Rate (b)FP Rate (c)Runtime

As shown above Fig. 9, unlike face detection using criteria for ROI or using criteria for Non-ROI, the proposed method can reduce FP rate and runtime while preserving TP rate. The proposed method reduces the time to examine the region containing non face on Non-ROI and does not omit faces.

5. CONCLUSION

The conventional face detection system used only pixel based information. However, not only pixel based information but also, context information can be used to improve face detection performance. In this paper, we verified the spatial context information which people allocate the concern object in the center of the scene when they take a picture, and determined the region of interest (ROI) based on the spatial context information. Most faces are allocated on the ROI rather than Non-ROI. On the other hand, the Non-ROI has many background regions and regions containing non-face which can cause false alarms. We applied the ROI information on the face detection method. In our system, face detection with different characteristics of ROI and Non-ROI are used. Experimental results show that the proposed face detection method using spatial context information does improve the performance and reduce the runtime.

We can affirm through this work that not only pixel based information but also context information and tendency or preferences of people can also enhance the performance of face detection.

REFERENCES

[1] R. R. Duda, P. E. Hart, and D. G. Strok, "Pattern

Classification," Second Edition by John Wiley & Sons, Inc, 2001.
[2] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, pp. 23-38, 1998.
[3] E. Osuna, R. Freund, and Girosi, "Training Support Vector Machines: An Application to Face Detection,"Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 130-136, 1997.
[4] R. Brunelli and T. Poggio, "Face Recognition: feature versus templates," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, pp. 1042-1052, 1993.
[5] K. C. Tow, R. Clipolla, "Feature-Based Human Face Detection," Image and Vision Computing, Vol.15, No.9, pp.713-735, 1997.
[6] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 511-518, 2001.
[7] R. Lienhart, J. Maydt, " An Extended Set of Haar-like features for rapid object detection," IEEE Int. Conf. Inmage Processing, Vol. 1, pp. 900-903, 2002.
[8] C. Garcia, G. Tziritas, "Face Detection using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," IEEE Multimedia, Vol.1, No.3, pp.264-277, 1999.
[9] D. Chai and K.N. Ngan, "Face segmentation using skin colour map in videophone applications," IEEE Trans. on Circuits and Systems for Video Tech., Vol. 9, No. 4, pp. 551-564, 1999.
[10] M. C. Hwang, L.T Ha, N. H. Kim, C.S. Park, and S.J. Ko, "Person identification system for future digital tv with intelligence," IEEE Trans. Consumer Electronics, Vol.53, pp. 218-226, 2007.
[11] Open Source Computer Vision Library: http://www.intel.com/research/mrl/research/opencv