Robust Real-Time Face Detection Using Face Certainty Map

Bongjin Jun and Daijin Kim

Department of Computer Science and Engineering
Pohang University of Science and Technology,
{simple21,dkim}@postech.ac.kr

Abstract. In this paper, we present a robust real-time face detection algorithm. We improved the conventional face detection algorithms for three different steps. For preprocessing step, we revise the modified census transform to compensate the sensitivity to the change of pixel values. For face detection step, we propose difference of pyramid(DoP) images for fast face detection. Finally, for postprocessing step, we propose face certainty map(FCM) which contains facial information such as facial size, location, rotation, and confidence value to reduce FAR(False Acceptance Rate) with constant detection performance. The experimental results show that the reduction of FAR is ten times better than existing cascade adaboost detector while keeping detection rate and detection time almost the same.

1   Introduction

Face detection is an essential preprocessing step for face recognition[1], surveillance, robot vision interface, and facial expression recognition. It also has many application areas such as picture indexing, tracking, clustering and so on. However face detection has its intrinsic difficulties for the following reasons. First, face is not a rigid object, i.e. every person has different facial shape and different form/location of facial features such as eyes, nose, and mouth. Second, face of the same person looks differently as the facial expression, facial pose, and illumination condition changes. Finally, it is almost impossible to train infinite number of non-face patterns, consequently unexpected false acceptance or false rejection could be occurred.

The procedure of face detection could be divided as preprocessing step, face detection step, and postprocessing step. First, for preprocessing step, illumination compensation techniques like histogram equalization[2], normalization to zero mean and unit variance on the analysis window[3], and modified census transform[4] have been proposed. Second, for detecting faces, many classification algorithms have been proposed to classify the face and non-face patterns such as: skin color based approaches[5][6], SVM[7][8], gaussian mixture model[9], maximum likelihood[10], neural network[11] [8], and adaboost[3][4]. Finally, for postprocessing step, the algorithms usually group detected faces which is located in the similar position. Then, they select only one face from each face group and
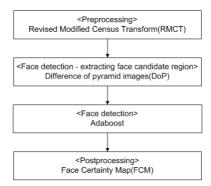
Fig. 1. The overall procedure of the proposed algorithm

determine the size, location, and rotation of the selected face. These methods usually show good performance, but have difficulties in learning every non-face patterns in natural scene. In addition, these methods are somewhat slow due to much computation steps.

In this paper, we present a novel face detection algorithm. For preprocessing step, we revise the modified census transform to compensate the sensitivity to the change of pixel values. For face detection step, we propose we propose difference of pyramid(DoP) images for fast face detection. Finally, for postprocessing step, we propose face certainty map(FCM) which contains facial information such as facial size, location, rotation, and confidence value to reduce FAR(False Acceptance Rate) with constant detection performance(Fig. 1).
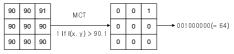
The outline of this paper is as follows: in section 2, we explain modified census transform and our proposed revised modified census transform for preprocessing step. In section 3, we explain face detection using adaboost[4], our proposed face detection method, and the combined system. In section 4, we propose face certainty map for postprocessing step. In section5, we show the experimental results and analysis. Finally, some conclusions of this work are given in section 6.
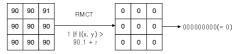
2   Preprocessing Step

2.1   Revised Modified Census Transform

Zabin and Woodfill proposed an illumination insensitive local transform method called census transform(CT) which is an ordered set of comparisons of pixel intensities in a local neighborhood representing which pixels have lesser intensity than the center[12]. Let $N(x)$ define a local spatial neighborhood of the pixel at x so that $x \notin N(x)$, a comparison function $C(I(x), I(x'))$ be 1 if $I(x) < I(x')$, and $\bigotimes$ denote the concatenation operation, then the census transform at x is defined as

$$T(x) = \bigotimes_{y \in N} C(I(x), I(y)). \tag{1}$$

(a) Modified Census Transform(MCT)

(b) Revised Modified Census Transform(RMCT)

Fig. 2. Comparison of MCT and RMCT

Since census transform transforms pixel values by comparison with center pixel value, it can not transforms the pixel values equal to center pixel value.

Fröba and Ernst proposed modified census transform(MCT) to solve this problem[4]. Let $N'(x)$ be a local spatial neighborhood of the pixel at x so that $N'(x) = N(x) \cup x$. The intensity mean on this neighborhood is denoted by $\bar{I}(x)$. With this, they reformulate equation 1 and write the modified census transform as

$$\Gamma(x) = \bigotimes_{y \in N'} C(\bar{I}(x), I(y)). \tag{2}$$

Using equation 2, they could determine all of the 511 structure kernels defined on a 3×3 neighborhood, while CT has only 256 structure kernels. However, as you can see in Fig 2-(a), MCT is sensitive to subtle changes of pixel values in local region. To solve this problem, we revised MCT by addition of a small value $r(=2 \text{ or } 3)$ as

$$\Upsilon(x) = \bigotimes_{y \in N'} C(\bar{I}(x) + r, I(y)). \tag{3}$$

We call equation 3 as revised modified census transform(RMCT). RMCT transforms the pixel values to one of 511 patterns in 3×3 neighborhood. Since the local pixel value changes in 3×3 neighborhood is insensitive to illumination change, this transform is robust to illumination change(Fig 2-(b)). Moreover, since RMCT has regular patterns which can represent facial features, it is good to classify face and non-face patterns.

3   Face Detection Step

3.1   Face Detection Using RMCT and Adaboost

In this section, we present our proposed face detection algorithm using RMCT and adaboost. RMCT transforms the pixel values to one of 511 patterns in 3×3 neighborhood. Then, using the training images transformed by RMCT, we construct the weak classifier which classifies the face and non-face patterns

and the strong classifier which is the linear combination of weak classifiers. The weak classifier consists of the set of feature locations and the confidence values for each RMCT pattern.

In test phase, we scan the image plane by shifting the scanning window and obtain the confidence value for strong classifier in each window location. Then, we determine the window location as face region, when the confidence value is above the threshold.

Moreover, we construct multi-stage classifier cascade for fast face detection. We expect that the image patches which contain background are rejected from early stage of cascade. Accordingly the detection speed may increase[3][4]. For training multi-stage classifier cascade, the non-face training images for first classifier are composed of arbitrary non-face images, while that of later classifiers are composed of the images which is falsely accepted by the former classifiers.

3.2   Speed Up the Algorithm

The face detector analyzes image patches of pre-defined size. Each window has to be classified either as a face or non-face. In order to detect faces in the input image, conventional face detection algorithms have been carried out by scanning all possible analysis windows. In addition, to find faces of various size, the image is repeatedly down-scaled with a pre-defined scaling factor. This is done until the scaled image is smaller than the sub window size. Although face detection by this full-search algorithm shows optimal performance, it is computationally expensive and slow.

To solve this problem, we propose difference of pyramid(DoP) images coupled with two-dimensional logarithmic search. First, we obtain face candidate region using DoP. Since the algorithm does not search the whole input image but the face candidate region, we expect it will reduce the computational time.

Difference of Pyramid Images. Extracting face candidate region using motion difference has been adopted for fast face detection in many previous works. However this method assumes that the location of camera is fixed and the background image is constant. Accordingly, it is difficult to adopt this method to still image or image sequence from moving camera. We propose difference of pyramid images(DoP) for compensating this problem. In order to obtain face candidate region using motion difference, at least two images are required such as background image and input image or image of previous frame and image of current frame. However, we can obtain face candidate region from single image using DoP, since it is computed not from image sequence but from single image. Thus, we are able to adopt this algorithm to still image and image sequence from moving camera as well.

Many face detection algorithms construct several numbers of down-scaled pyramid images and then scan each pyramid image using scanning window with pre-defined size. In this paper, we construct n down-scaled images which constitute an image pyramid. Then we obtain $n-1$ DoP image by subtract $i-th$ pyramid image from $(i-1)-th$ pyramid image. Since the size of $i-th$ pyramid
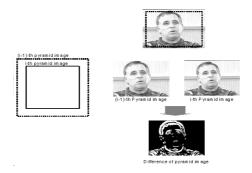
Fig. 3. Difference of Pyramid Images

image and $(i-1)-th$ pyramid image are different, we first align each image with the center point then obtain the DoP image by subtracting corresponding pixel points. Since plain background has little changes in DoP image, we can perform fast face detection. The pixel points in DoP image which have higher value than threshold are selected as face candidate region. Fig. 3 shows some example images of DoP.

4   Postprocessing Step

4.1   Face Certainty Map

For minimizing FAR(False Acceptance Rate) and FRR(False Rejection Rate), existing face detection algorithms concentrate on learning optimal model parameters and devising optimal detection algorithm. However, since the model parameters are determined by the face and non-face images in the training set, it is not guaranteed that the algorithms work well for novel images. In addition, there are infinite number of non-face patterns in real-world, accordingly it is almost impossible to train every non-face patterns in natural scene. As a result, the face detection algorithms, which showed good performance during the training phase, show high FRR and FAR in real environment.

The face detector we described in section 3.1 determines the image patch in the current scanning window as face when the confidence value is above the threshold and determines it as non-face when the confidence value is beyond the threshold. We call the scanning window of which the confidence value is above the threshold as detected face window. Fig. 4 shows some examples of the face detection algorithm proposed in section 3.1: the rectangles in each figure represent the detected face windows. We can see that while the face regions are detected perfectly, there are several falsely accepted regions(blue circle regions). In addition, when we investigate the figures precisely, there are numbers of detected face windows near real face region, while few detected face windows are near falsely accepted regions. With this observation, we propose face certainty
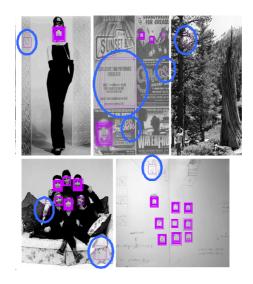
Fig. 4. Face Detection Results Without FCM

map(FCM) which can reduce FAR with constant detection performance and no additional training.

As you can see in Fig. 4, there are multiple detected face windows, even though there is only one face in the input image. For real face region, there are detected face windows with same location but different scale(Fig. 5-(a)) and detected face windows with same scale but different location(Fig. 5-(b)). However, falsely accepted regions do not show this property. Consequently, we can determine the regions where multiple detected face windows are overlapped as face region and the regions with no overlapped detected face windows as falsely accepted region. By adopting this methodology, we can reduce FAR greatly.

For now, we explain how to adopt FCM to the face detection algorithm we described in section 3.1. The detailed explanation of the procedure is given below.

1. For each scanning window centered at (x, y), we compute the confidence value.
$$H_i(\Upsilon) = \sum_{p \in S_i} h_p(\Upsilon(p)),$$
   where i represents the $i-th$ cascade, p represents the $p-th$ feature location, and S_i is the set of feature locations, respectively.
2. The confidence value cumulated for all n cascade is like following:

$$S(x, y) = \sum_{i=1}^{n} H_i(\Upsilon), \tag{4}$$

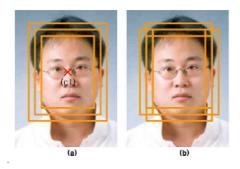   if $H_i(\Upsilon)$ for all i is above threshold, otherwise 0.

Fig. 5. Detected Face Windows;(a)Detected face windows which have the same center point(C1) but different scale (b) Detected face windows which have the same scale but different center point

    We compute the cumulated confidence value for every pixel position in the image.

3. We also compute equation 4 for all pyramid images, then we have $S_p(x, y), p = 1, \ldots, m$, where m is the total number of constructed pyramid images and the pixel locations (x, y) of each down-scaled pyramid image are translated to its corresponding original image locations.

4. The FCM for location (x, y) consists of four items such as $S_{max}(x, y)$, $W_{max}(x, y)$, $H_{max}(x, y)$, and $C(x, y)$. $S_{max}(x, y)$ is the maximum confidence value among $S_p(x, y), p = 1, \ldots, m$, $W_{max}(x, y)$ and $H_{max}(x, y)$ is the width and height of the detected face window which has the maximum confidence value, and $C(x, y)$ is the confidence value cumulated for all m pyramid images $C(x, y) = \sum_{p=1}^{m} S_p(x, y)$.

5. Since we constructed FCM, we can determine the face region using it. First, we look for the values above threshold in $S_{max}(x, y)$. Then we determine the location (x, y) as the center of face when $C(x, y)$ is above threshold. The non-face region where the maximum confidence value is above threshold is not classified as a face region, since $C(x, y)$ is lower than the threshold. Consequently, we can reduce the FAR using our proposed FCM.

5   Experimental Results and Discussion

For constructing training face data, we gathered 17,000 face images from internet and Postech DB[13]. Gathered face images contain multiple human species, variety of illumination conditions, and expression variation. Each image is aligned by eye location, and we resize images to 22×22 base resolution. In addition, for the robustness to image rotation, we generated another 25,060 face images by rotating gathered face images to -3, 0, 3 degrees.

    For non-face training data, we collected 5,000 images which include no face image from internet. Then, we extracted image patches from collected internet images by random size and position. After that, we generated 60,000 non-face

Fig. 6. Experimental Results

images by resizing extracted image patches to the same scale of training face images. We used these 60,000 non-face images as the training non-face data for the first stage of cascade. For the next stages of cascade, we used non-face data which are considered as face image by the previous cascade(i.e. we used false positives of previous cascade as training non-face data for training current cascade). A validation training data set is used for obtaining threshold value and stop condition of each stage of cascade. The validation set face and non-face images exclude the images used for training. We constructed validation set of 15,000 face images and 25,000 non-face images by the same way as we used for training data.

For each cascade, we preprocessed each face and non-face image using RMCT. Then we chose the feature positions for classification (S_i) and obtained classification value($H_i(\Upsilon)$) and threshold value(T_i) for each position. We constructed the face detector with 4 cascade and the maximal number of allowed position for each cascade is 40, 80, 160, and 400, respectively. In addition, the RMCT is defined 511 patterns in 3×3 neighborhood. Accordingly, we can not apply the 3×3 RMCT to the training images which have the size 22×22. Thus, we excluded the outer areas of each image and used the inner 20×20 areas of it.

We tested our algorithm on CMU+MIT frontal face test set. Fig. 6 and table 1 represents the results of face detection. When we used FCM, the reduction of FAR

Table 1. Results of Face Detection

Detector	Number of False Detection
RMCT, adaboost and FCM	3
RMCT and adaboost	93
Viola-Jones	78
Rowley-Baluja-Kanade	167
Bernhard Froba	27

is ten times better than the cascade adaboost detector with the same detection rate, while the detection time is almost the same. The cascade adaboot detector needs computations for grouping and eliminating overlapped face candidate region. In contrast, proposed detector does not need these computations but needs the computation for FCM. Operating on 320 by 240 pixel images, faces are detected at 23 frames per second on a conventional 3.2 GHz Intel Pentium IV system and 6 frames per second on OMAP5912(ARM9 system).

6   Conclusion

In this paper, we proposed a face and robust face detection algorithm using difference of pyramid(DoP) images and face certainty map(FCM). The experimental results showed that the reduction of FAR is ten times better than existing cascade adaboost detector while keeping detection rate and detection time almost the same. Existing adaboost face detection algorithms have to add more cascade stages for non-face images in order to reduce FAR. However, since it needs more weak classifier for constructing strong classifier, the processing time increases. Moreover, as the number of stages in cascade increase, FRR also increase. We were free from these drawbacks and increased detection performance by applying FCM to existing adaboost face detection algorithm. Since we can reduce FAR, the number of stages in cascade is also minimized, while preserving the same performance as existing algorithm which has more stages of cascade. Accordingly training time and processing time is faster than existing algorithm. Furthermore, FCM can be applied to any face detection algorithm beside adaboost which obtains confidence value or probability.

In this work, we applied the algorithm to only frontal faces. A future extension of this work could be pose and rotation invariant face detection algorithm

Acknowledgements

References

1. Lee, H.-S., Kim, D.: Facial expression transformations for expression-invariant face recognition. In: Proc. of International Symposium on Visual Computing, pp. 323–333 (2006)
2. Sung, K.K.: Learning and Example Selection for Object and Pattern Recognition. PhD thesis, MIT, AI Lab, Cambridge (1996)
3. Viola, P., Jones, M.: Fast and Robust Classification using Asymmetric Adaboost and a Detector Cascade. In: Advances in Neural Information Processing System, vol. 14, MIT Press, Cambridge (2002)
4. Froba, B., Ernst, A.: Face detection with the modified census transform. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition. IEEE Computer Society Press, Los Alamitos (2004)
5. Yang, J., Waibel, A.: A real-time face tracker. In: Proc. 3rd Workshop on Appl. of Computer Vision, pp. 142–147 (1996)
6. Dai, Y., Nakano, Y.: Face texture model based on sgld and its application in face detection in a color scene. Pattern Recognition 29, 1007–1017 (1996)
7. Osuna, E.: Support Vector Machines: Training and Applications. PhD thesis, MIT, EE/CS Dept., Cambridge (1998)
8. Mohan, A., Papageorgiou, C., Poggio, T.: Examplebased object detection in images by components. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 349–361 (2001)
9. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 39–51 (1998)
10. Schneiderman, H., Kanade, T.: A statistical method for 3d object detection applied to face and cars. In: Computer Vision and Pattern Recognition, pp. 746–751 (2000)
11. Rowley, H., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 23–38 (1998)
12. Zabih, R., Woodfill, J.: A non-parametric approach to visual correspondence. IEEE Transactions on Pattern Analysis and Machine Intelligence (1996)
13. Kim, H.C., Sung, J.W., Je, H.M., Kim, S.K., Jun, B.J., Kim, D., Bang, S.Y.: Asian Face Image Database PF01. Technical Report, Intelligent Multimedia Lab, Dept. of CSE, POSTECH (2001)