Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 3, July 2021:3342- 3349

Research Article

Robust Extraction of Human Facial Components Using a Landmark-based Model

Seok-Woo Jang
Professor, Department of Software, Anyang University, 22, 37-Beongil, Samdeok-Ro, Manan-Gu,
Anyang 14028, Republic of Korea
swjang7285@gmail.com
Corresponding author*: mobile Phone: +82-10-9482-6547

Abstract

Background/Objectives: Unlike general cameras, high-speed cameras capable of capturing a very large number of frames per second can enable the advancement of image processing technologies that have been limited.

Methods/Statistical analysis: In this paper, we propose a method of removing noises from high-speed color images and then detecting human facial areas from the noise-removed image. In this paper, first, noise pixels included in the ultrafast image are effectively removed by applying a bidirectional filter. Then, using a retina face model, a face region representing a person's personal information is robustly detected from an image from which noise has been removed.

Findings: Experimental results show that the proposed algorithm removes noises from images and then robustly detects the face region using the generated model. In this study, the performance of the proposed model-based face detection approach was quantitatively compared and evaluated in terms of accuracy. In this study, the accuracy scale expressed as the ratio of the number of face regions accurately extracted through the introduced method and the number of face regions originally existing in the entire image data was used. For performance evaluation, we also implemented the method using the existing fixed model. The existing face detection has not been able to compensate for noise included in images. In addition, since an inflexible model was used, many errors occurred in face detection. In contrast, the proposed method removes undesirable noise contained in an image by applying a bidirectional filter. Then, a flexible model composed of five landmarks is created to detect a face from an image from which noise has been removed, so that accurate results can be obtained.

Improvements/Applications: The proposed face detection method is expected to be used for many application fields related to pattern recognition such as building monitoring, door management, and mobile biometric authentication.

Keywords: Color image, Landmark model, Preprocessing, Facial component, Performance evaluation.

1. Introduction

Recently, relatively inexpensive high-speed cameras, such as small cameras mounted on Samsung's Galaxy

S series, which can shoot more than hundreds of frames per second, have become more common, creating an environment where ordinary people can easily acquire video content shot with ultra-fast cameras [1]. Such ultra-high-speed video content can be usefully used in various applications such as fine motion measurement of objects, three-dimensional modeling and analysis of objects [2-4].

However, in addition to the ultra-high-speed image data that provides useful information to many users, the ultra-high-speed image data including exposed personal information such as a person's face or a specific part of the body is freely distributed without any restrictions, which can be a social problem. In particular, those who realize that their personal information is being exposed and shared with many people suffer a great deal of mental damage.

Therefore, there is a need for a study that robustly detects the area representing the personal information [5] of a person exposed, such as a face, from an input high-speed color image through an image processing technique. The personal information areas of the person detected in this way can be effectively protected through a blocking process such as a mosaic, which is carried out in the next step.

Related studies that have been conducted in the past to detect areas representing personal information of exposed persons such as faces from input color image data can be found in the literature. In the study [6], a feature aggregation network (FANet) was proposed to construct a novel single-phase facial extractor that not only produces excellent result but also runs efficiently. In the study [7], a partial face recognizer in the mobile domain was proposed. In this method, they describe two different approaches to segment-based face detection. In the study [8], a strategy for assisting proposal creation was proposed to acquire faces quickly in mobile devices. In the study [9], a method of normalizing the profile image to the front face was proposed for face recognition irrelevant to the pose. In addition to the existing methods described above, researches on detecting regions of interest including personal information such as faces through image analysis are ongoing [10].

The existing methods mentioned above detect a person's face area with some accuracy in a general environment. However, most of these methods are not for high-speed cameras, but for stills and videos shot with normal flat-speed cameras. Therefore, there are some limitations and constraints in simply applying the face detection algorithms used in the existing methods to ultra-high-speed image processing.

Therefore, in this work, we introduce a method to effectively remove unnecessary noise pixels contained within the input image using bidirectional filters from ultra-fast color images entered into the system, and then robustly detect human facial regions representing privacy from the denoised image using landmark-based facial models. Fig. 1 briefly illustrates the overall flow diagram of the model-based face area detection algorithm using five landmarks introduced in this work.

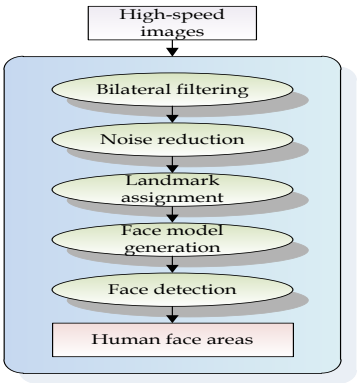

Figure 1. Overall flow chart of the introduced method

As can be seen in Fig. 1, the approach introduced in this study effectively reduces unnecessary noises present in the input image received first by applying a bidirectional filter. Then, from the color image from which noise has been removed, the face area of a person representing personal information is robustly extracted using a model generated based on the landmark.

In Chapter 1, we described the overall outline and background that led to this study. Chapter 2 describes the preprocessing to remove noise included in the input image. In Chapter 3, a landmark-based model is created and a method of detecting a face area from an input image using the generated model is described. In Chapter 4, we describe the experimental results conducted to compare and evaluate the performance of the model-based face region detection method presented in this study. Chapter 5 describes the conclusion of this study and future research plans.

2. Image Preprocessing

In this study, a bilateral filter [11] is applied to effectively remove unwanted noises in color images received at high speed. In general, noise often occurs when the intensity of light received by a high-speed camera is relatively weaker than a signal generated by an electrical signal. In addition, while the size of the image sensor mounted on the ultra-high-speed camera is small, it tends to generate a lot of noise even when the resolution is high. Therefore, in this study, after removing noise as much as possible through pre-processing of the image, we try to detect the human face region more stably.

Usually, in order to reduce various types of noises included in a color image, a two-dimensional Gaussian filter [12] is often used. However, while the Gaussian filter has the advantage of being easy to use, it has the disadvantage of blurring the contents of the image. In other words, when a Gaussian filter is applied to a color image, a phenomenon in which a boundary portion of the target existing in the color image is blurred occurs.

Therefore, we apply a bidirectional filter that operates to preserve the boundary of the object in the image as much as possible. The bidirectional filter used in this study is defined as Equation (1).

$$I^{filter}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) \times f_r(\| I(x_i) - I(x) \|) \times g_s(\| x_i - x \|) \quad (1)$$

In Equation (1), I(x) denotes an initial input image to be filtered, and $I^{filter}(x)$ denotes a result image subjected to bidirectional filtering. In addition, x represents the current position of a pixel to which filtering is applied, and Ω represents a search window located around x. Therefore, x_i represents another adjacent pixel. The function f_r denotes a range kernel for smoothing the difference between neighboring pixel values, and the function g_s denotes a spatial kernel for smoothing the difference between neighboring coordinates. In this study, kernels f_r and g_s use Gaussian functions.

$$W_p = \sum_{x_i \in \Omega} f_r(\| I(x_i) - I(x) \|) g_s(\| x_i - x \|) \quad (2)$$

The weight factor W_p used in Equation (1) is defined as Equation (2) using the spatial kernel g_s and the range kernel f_r. Let us consider the pixel at (i, j) that contains noise in the input image. Also, let us consider that one of the pixels adjacent to (i, j) is positioned at (k, l). Subsequently, the range kernel and the spatial kernel are set as a two-dimensional Gaussian function, and the weight factor set for the pixel located at (k, l) to remove noise from the pixel located at (i, j) is as shown in Equation (3).

$$w(i, j, k, l) = \exp\left(-\frac{(i-k)^2 + (j-l)^2}{2\sigma_d^2} - \frac{\|I(i, j) - I(k, l)\|^2}{2\sigma_r^2} \right) \quad (3)$$

In Equation (3), I(i, j) and I(k, l) represent the brightness values at positions (i, j) and (k, l), and σ_d and σ_r are range and space smoothing parameters. In this study, if the weighting factor is extracted and then normalized, the intensity value $I_D(x, y)$ of the pixel with reduced noise at the position (i, j) can be obtained as shown in Equation (4).

$$I_D(i, j) = \frac{\sum_{k,l} I(k, l) w(i, j, k, l)}{\sum_{k,l} w(i, j, k, l)} \quad (4)$$

Usually, as the value of the range parameter σ_r increases, the Gaussian convolution becomes wider and smoother. Furthermore, as the value of the spatial parameter σ_d increases, the larger feature is smoothed.

Bidirectional filtering considers correlations between current pixels and surrounding pixels, such as Gaussian filtering, and reduces the noise contained in the image by considering differences from the values of pixels in the region of interest. In two-way filtering, the corresponding pixel value is set to the weighted average value of the surrounding pixels. Furthermore, the weighting factors used may be established according to the general Gaussian distribution.

3. Model-based Face Detection

In this paper, from the ultra-high-speed color image with reduced noise through the two-way filtering performed in the previous step, the face area of a person including exposed personal information is to be robustly detected based on the face model. To this end, in this study, we intend to use the retina face model based on the landmark [13].

The retina face model used in this paper uses a single-layered face detection method, which performs pixel-wise face localization for different sizes of faces through combined self-supervised and extra-supervised multi-task

learning. A description of the main process of the corresponding learning method is as follows.

- Multitask loss

For all training anchors i, the retina model minimizes the loss as shown in Equation (5).

$$L = L_{cls}(p_i, p_i^*) + \lambda_1 p_i^* L_{box}(t_i, t_i^*) + \lambda_2 p_i^* L_{pts}(l_i, l_i^*) + \lambda_3 p_i^* L_{pixel} \quad (5)$$

In Equation (5), $L_{cls}(p_i, p_i^*)$ denotes the face classification loss, p_i denotes the forecasted probability of whether anchor i is a face, and p_i^* denotes 1 for the positive anchor and 0 for the opposite case. The categorization loss uses the softmax [14] for the binary class. In $L_{box}(t_i, t_i^*)$, the face region regression loss, we use $t_i = \{t_x, t_y, t_w, t_h\}_i$ and $t_i^* = \{t_x^*, t_y^*, t_w^*, t_h^*\}_i$ to denote predicted regions and ground truth regions for positive anchors, and for the face region, we use the face's center position, and the face's horizontal and vertical width to normalize the goal. $L_{pts}(l_i, l_i^*)$, a face landmark regression loss, uses five predicted face landmarks for positive anchors and ground truth, $l_i = \{l_{x_1}, l_{x_2}, ..., l_{x_5}\}_i$ and $l_i^* = \{l_{x_1}^*, l_{x_2}^*, ..., l_{x_5}^*\}_i$. Like the face area, landmark regression is also normalized based on the center of the anchor. The final loss L_{pixel}, the dense regression loss, uses the loss balance parameters $\lambda_1 - \lambda_3$ as 0.25, 0.1, and 0.01, respectively.

- Dense regression branch

In the retina face model, a mesh decoder is implemented via graph convolution based on a fast localized spectral filtering method. Furthermore, we use joint shapes and texture decoder similarities for faster decoding. Unlike general convolutions, graph convolutions use Euclidean response fields by computing their distance from their neighbors. Graph vertices in the retina face model are defined using colored face meshes with joint shape and texture information, and connections between vertices are defined via sparse adjacency matrices [15].

- Image quality classification

In the retina face model used in this paper, various facial images are classified into five categories based on the image quality. Here, the method of classifying the image quality is based on how difficult it is to display a landmark on the face area. As landmarks used here, both eyes of a person, the tip of the nose, and the area of the mouth on both sides are used.

In this study, the model learned through the above process and the ultra-high-speed camera are used in combination. In an ultra-high-speed camera environment, the image quality captured by the image sensor is not clear due to flickering and unremoved noise. However, the retina face model can exhibit relatively high performance even with poor image quality by using a method to which the concept of face classification, landmark, density check, and graphs between landmarks.

4. Results and Discussion

The computer used for the development and testing of the algorithm presented in this study was an Intel Core i7-6700 3.4 GHz CPU, 16 GB random access memory (RAM), Galaxy GeForce GTX 1080 Ti graphics card,

256 GB solid state drive (SSD). The computer used for the experiment has Windows 10 installed as the operating system. As a software tool for system development, Microsoft's Visual Studio 2017 was installed. In addition, the OpenCV image processing library was used to more efficiently develop the algorithm proposed in this study.

In this study, the performance of the proposed landmark model-based face area detection approach was quantitatively compared and evaluated in terms of accuracy. In this study, we use an accuracy scale such as Equation (6), expressed as a ratio of the number of precisely extracted face regions to the number of face regions originally present in the total image data used in the experiment. In Equation (6), $FACE_{extracted}$ represents the number of face areas accurately detected using the introduced method. In addition, $FACE_{total}$ represents the total number of face regions included in the entire color image data used in the experiment. As can be seen from Equation (6), the face detection accuracy scale defined in this study is expressed as a percentage.

$$M_{accuracy} = \frac{FACE_{extracted}}{FACE_{total}} \times 100 \ (\%) \tag{6}$$

Fig. 2 shows the comparison of the performance measurement results of face detection from ultra-high speed images using the conventional skin color model-based method and the proposed method. As can be seen in Fig. 2, the approach proposed in this study removes noises included in the image, and then more robustly detects the exposed human face area using the landmark-based model.

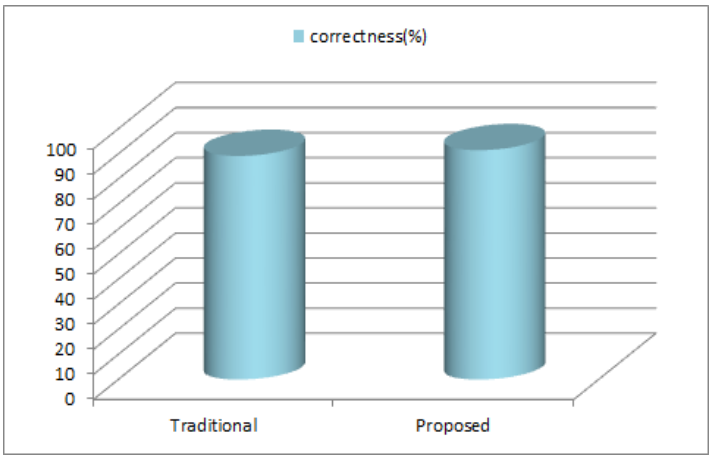

Figure 2. Performance graph

The existing face detection method has not been able to compensate for noises included in ultra-high-speed images correctly. In addition, since a fixed and inflexible model was used, many errors occurred in face detection.

In contrast, the method proposed in this study removes undesirable noises contained in an image by applying the bidirectional filter. Then, a flexible model consisting of five landmarks is created to try to detect a face from a color image from which noise has been removed, so that more accurate results can be obtained.

5. Conclusion

In recent years, as high-speed cameras, which are relatively inexpensive, are becoming more and more common, the general public can more and more easily acquire video content photographed with a high-speed camera. However, high-speed images with personal information, such as certain parts of the face and body, are

also freely distributed, which is a problem. Therefore, there is a need for a study to accurately detect an exposed object area containing personal information of a person from an input high-speed color image.

In this study, a technique was introduced to remove unwanted noise contained in ultra-high-speed color images filmed in general environments without special restrictions, and to accurately extract facial regions representing personal information from images from which noise was removed. In the introduced method, noise pixels generated in the image are removed as much as possible from the input color image by using the bidirectional filter. Then, by applying a face model composed of five landmark points representing the main features of the face, the face region, which is an area representing personal information, was robustly detected from the image data from which noise was removed. Experimental results represents that the approach described in this study removes noise from the ultra-high-speed color image data, and then accurately detects the human face region from the noise-removed image.

In the future, we plan to further improve the overall performance by robustly optimizing the landmark-based model of the face detection algorithm proposed in this study. In other words, in the input image, even if the face direction is not only the front side, but also the upper and lower sides, the direction of the face will be accurately measured and the landmark will be positioned according to the measured direction so that the overall face detection operation can proceed correctly.

6. Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2019R1F1A1056475)

7. References

1. Chen R, Li Z, Zhong K, Liu X, Chao YJ, Shi Y. Low-speed-camera-array-based high-speed three-dimensional deformation measurement method: principle, validation, and application. Optics and Lasers in Engineering. 2018 Mar;107:21-27. DOI: https://doi.org/10.1016/j.optlaseng.2018.03.009
2. Yu L, Pan B. Full-frame, high-speed 3D shape and deformation measurements using stereo-digital image correlation and a single color high-speed camera. Optics and Lasers in Engineering. 2017 Aug;95:17-25. DOI: https://doi.org/10.1016/j.optlaseng.2017.03.009
3. Jung W, Hurth C, Zenhausern F. Real-time monitoring of viscosity changes triggered by chemical reactions using a high-speed imaging method. Sensing and Bio-Sensing Research. 2015 Sep;5:8-12. DOI: https://doi.org/10.1016/j.sbsr.2015.05.003
4. Ma Z, Han M, Li Y, Gao H, Lu E, Chandio FA, Ma K. Motion of cereal particles on variable-amplitude sieve as determined by high-speed image analysis. Computers and Electronics in Agriculture. 2020 Jul;174:1-9. DOI: https://doi.org/10.1016/j.compag.2020.105465
5. Zafeiriou S, Zhang C, Zhang Z. A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding. 2015 Sep;138:1-24. DOI: https://doi.org/10.1016/j.cviu.2015.03.015
6. Zhang J, Wu X, Hoi SCH, Zhua J. Feature agglomeration networks for single stage face detection. Neurocomputing. 2020 Mar;380:180-189. DOI: https://doi.org/10.1016/j.neucom.2019.10.087
7. Mahbub U, Sarkar S, Chellappa R. Partial face detection in the mobile domain. Image and Vision Computing. 2019 Jan;82:1-17. DOI: https://doi.org/10.1016/j.imavis.2018.12.003

8. Zhang H, Wang X, Zhu J, Kuo CCJ. Fast face detection on mobile devices by leveraging global and local facial characteristics. Signal Processing: Image Communication. 2019 May;78:1-8. DOI: https://doi.org/10.1016/j.image.2019.05.016
9. Liu Y, Chen J. Unsupervised face frontalization for pose-invariant face recognition. Image and Vision Computing. 2020 Dec;106:1-10. DOI: https://doi.org/10.1016/j.imavis.2020.104093
10. Zhou Z, He Z, Jia Y, Du J, Wang L, Chen Z. Context prior-based with residual learning for face detection: a deep convolutional encoder–decoder network. Signal Processing: Image Communication. 2020 Jul;88:1-13. DOI: https://doi.org/10.1016/j.image.2020.115948
11. Geng J, Jiang W, Deng X, Multi-scale deep feature learning network with bilateral filtering for SAR image classification. ISPRS Journal of Photogrammetry and Remote Sensing. 2020 Jul;167;201-213. DOI: https://doi.org/10.1016/j.isprsjprs.2020.07.007
12. Wang R, Li W, Zhang L. Blur image identification with ensemble convolution neural networks. Signal Processing. 2019 Feb;155:73-82. DOI: https://doi.org/10.1016/j.sigpro.2018.09.027
13. Deng J, Guo J, Ververas E, Kotsia I, Zafeiriou S. RetinaFace: single-shot multi-level face localisation in the wild. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020 Jun;5203-5212. DOI: https://doi.org/10.1109/CVPR42600.2020.00525
14. Peng H, Yu S. Beyond softmax loss: intra-concentration and inter-separability loss for classification. Neurocomputing. 2020 Dec;438:155-164. DOI: https://doi.org/10.1016/j.neucom.2020.11.030
15. Naghashi V. Co-occurrence of adjacent sparse local ternary patterns: a feature descriptor for texture and face image retrieval. Optik, 2017 Nov;157:877-889. DOI: https://doi.org/10.1016/j.ijleo.2017.11.160