JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 10, ISSUE 04, 2023

Face Segmentation Using Skin Color Model for Video Phone Applications

TS Ghouse pasha[1], Matturi Pujitha[2], Miriyala Snigdha[2], Narayanadas Kavyasree[2]

[1,2]Department of Electronics and Communication Engineering

[1,2]Malla Reddy Engineering College for Women (A), Maisammaguda, Medchal, Telangana.

Abstract

This project presents an interactive algorithm to automatically segment out a person's face from a given image that consists of a head-and-shoulders view of the person and a complex background scene. The method involves a fast, reliable, and effective algorithm that exploits the spatial distribution characteristics of human skin color. The main objective of this research is to design a system that can find a person's face from given image data. This problem is commonly referred to as face location, face extraction, or face segmentation. Regardless of the terminology, they all share the same objective. However, note that the problem usually deals with finding the position and contour of a person's face since its location is unknown, but given the knowledge of its existence. If this is not known, then there is also a need to discriminate between "images containing faces" and "images not containing faces." This is known as face detection. This work, however, focuses on face segmentation. The significance of this problem can be illustrated by its vast applications, as face segmentation holds an important key to future advances in human-to-human and human-to-machine communications. The segmentation of a facial region provides a content-based representation of the image where it can be used for encoding, manipulation, enhancement, indexing, modeling, pattern-recognition, and object-tracking purposes.

Keywords: Face tracking, real time videos, Image segmentation.

1. Introduction

In video phone applications, we often see special effects and filters that enhance the visual experience during video calls. These effects, such as face filters and background replacement, rely on face segmentation technology. Face segmentation involves separating and extracting human faces from video streams or images. By understanding the background, need, and significance of face segmentation in real-time videos for video phone applications, we can appreciate how it improves our video calling experiences. As video phone applications gain popularity, we all desire more engaging and enjoyable video calls. Face segmentation is essential to achieving this goal. It allows the application to add exciting features that enhance our visual interactions. Whether it's applying real-time filters, trying out augmented reality effects, or changing the background behind us, face segmentation plays a crucial role in making video calls more fun and immersive.

1.1 Significance

- Real-time Video Effects: With face segmentation, video phone applications can offer real-time effects that can transform our appearance during video calls. These effects can range from simple filters to more advanced features like virtual makeup or face swapping. By using face segmentation, these effects can be accurately applied to our faces, making our video calls more entertaining and personalized.
- Augmented Reality (AR) Filters: Face segmentation is a vital component for implementing AR filters in video phone applications. AR filters allow us to add virtual objects or animations to our faces in real-time. Whether it's trying on virtual accessories or becoming a virtual character, face segmentation ensures that these filters are properly placed and tracked on our faces, creating a seamless and interactive experience.

- Background Replacement: Another significant application of face segmentation is background replacement. By accurately segmenting our faces, video phone applications can remove the background behind us and replace it with a virtual environment or a custom image. This feature provides privacy and eliminates distractions, allowing us to focus on the conversation. It also adds a touch of creativity and visual appeal to our video calls.
- User Experience and Engagement: Ultimately, face segmentation contributes to an enhanced user experience and increased engagement in video phone applications. By incorporating real-time effects, AR filters, and background modifications, video calls become more interactive, entertaining, and visually captivating. These features not only make the calls more enjoyable but also keep us engaged and satisfied with the application.

2. Literature Survey

Viola and Jones (2004) made a significant contribution to the field of face detection by introducing the Viola-Jones algorithm. This algorithm used a clever approach called Haar-like features to efficiently detect faces in real-time. By combining cascaded classifiers and integral images, they achieved impressive performance. It's worth noting that their method did not specifically rely on skin color models or morphological operations. Nevertheless, their work laid the foundation for subsequent research in face detection and served as a benchmark for comparison.

In 2010, Rizvi et al. proposed a real-time face detection and tracking technique that incorporated skin color segmentation and a Kalman filter. Their method leveraged skin color information to identify potential face regions in video frames. To track the detected faces over time, they employed a Kalman filter, which helped enhance the accuracy and robustness of the face detection in real-time scenarios.

Jain and Chen (2013) provided a comprehensive overview of facial feature detection and tracking methods in videos. While their discussion did not specifically focus on skin color models or morphological operations, they offered valuable insights into various algorithms and approaches used for face tracking. They covered techniques based on features, appearance, and models, shedding light on the different possibilities in this field.

Fernández et al. (2013) tackled the challenge of face detection by integrating skin color information and depth data. They aimed to enhance the accuracy of face detection, especially in challenging lighting conditions, by incorporating depth information obtained from depth sensors like Microsoft Kinect. Their approach combined skin color segmentation with depth data, leading to more reliable face detection results.

Kim et al. (2015) proposed a real-time face tracking system that took advantage of both skin color and motion information. Their method started with skin color-based segmentation to identify potential face regions. Then, they employed motion information to track the detected faces across consecutive video frames. By incorporating motion cues, their approach achieved robust face tracking even in dynamic video sequences where faces might undergo significant changes in pose and appearance.

Hu et al. (2016) developed a real-time skin color segmentation and tracking approach primarily for hand gesture recognition. However, their method can be adapted for face tracking as well. They utilized skin color models to accurately segment the hand region and applied morphological operations like erosion and dilation to refine the segmentation results. These operations played a crucial role in achieving precise hand tracking and accurate gesture recognition in real-time scenarios.

Feng et al. (2017) proposed a real-time face tracking technique that combined skin color segmentation and an active contour model, also known as a snake. They employed skin color information to segment the face region and used the active contour model to refine the segmentation boundary. By

iteratively adjusting the contour using energy optimization, their approach achieved precise and robust face tracking results.

Bhowmik et al. (2018) introduced a real-time face detection and tracking system that integrated skin color segmentation and template matching. They utilized skin color models to segment potential face regions, and then employed template matching techniques to verify and track the detected faces. The template matching process involved comparing the segmented regions with predefined face templates or reference images. Their method aimed to achieve accurate and reliable face detection and tracking in real-world scenarios.

Rahim et al. (2019) proposed a real-time face detection and tracking approach that incorporated a skin color model and an active contour algorithm. Their method involved segmenting potential face regions using skin color information and refining the face boundary using an active contour algorithm. The active contour model, through energy minimization, accurately adapted the contour to the face shape. By combining these techniques, their approach achieved robust face detection and tracking performance across various conditions.

Kaur and Singh (2020) presented a real-time face tracking technique that relied on skin color detection and morphological operations. They utilized skin color models to segment the face region and applied morphological operations like erosion and dilation to refine the segmentation results. These operations effectively removed noise and improved the accuracy of the segmentation. The method aimed to achieve real-time and accurate face tracking using simple yet effective techniques.

These references collectively offer a range of approaches and techniques for face segmentation in real-time videos using skin color models and morphological operations. While some papers focused specifically on face detection or tracking, others explored both aspects. The combination of skin color models and morphological operations provides a practical approach to segmenting faces in real-time videos, serving as a foundation for subsequent analysis and processing of facial features.

3. Face-Segmentation Algorithm

In this section, we present our methodology to perform face segmentation. Our proposed approach is automatic in the sense that it uses an unsupervised segmentation algorithm, and hence no manual adjustment of any design parameter is needed in order to suit any particular input image. Moreover, the algorithm can be implemented in real time, and its underlying assumptions are minimal. In fact, the only principal assumption is that the person’s face must be present in the given image, since we are locating and not detecting whether there is a face. Thus, the input information required by the algorithm is a single color image that consists of a head-andshoulders view of the person and a background scene, and the facial region can be as small as only a 32 32 pixels window (or 1%) of a CIF-size (352 288) input image. The format of the input image is to follow the YCrCb color space, based on the reason given in the previous section. The spatial sampling frequency ratio of Y, Cr, and Cb is 4 : 1 : 1. So, for a CIF-size image, Y has 288 lines and 352 pixels per line, while both Cr and Cb have 144 lines and 176 pixels per line each. The algorithm consists of five operating stages, as outlined in Fig. 5. It begins by employing a low-level process like color segmentation in the first stage, then uses higher level operations that involve some heuristic knowledge about the local connectivity of the skin-color pixels in the later stages. Thus, each stage makes full use of the result yielded by its preceding stage in order to refine the output result. Consequently, all the stages must be carried out progressively according to the given sequence.

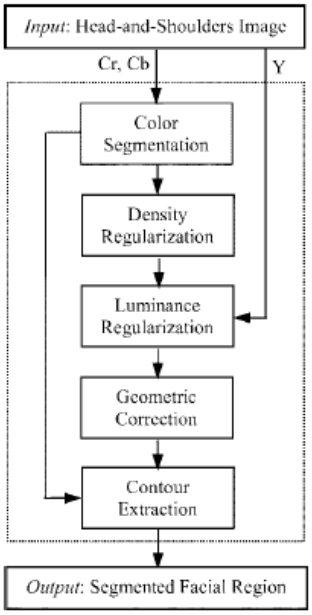


Fig. 1: Outline of face segmentation algorithm.

A detailed description of each stage is presented below. For illustration purposes, we will use a studio-based headand- shoulders image called *Miss America* to present the intermediate results obtained from each stage of the algorithm.

A. Stage One—Color Segmentation

The first stage of the algorithm involves the use of color information in a fast, low-level region segmentation process. The aim is to classify pixels of the input image into skin color and non-skin color. To do so, we have devised a skin-color reference map in YCrCb color space.

Fig. 2: Input image of miss amertca.

We have found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e., Cr and Cb) values narrowly and consistently distributed in the YCrCb color space. We denote and as the respective ranges of Cr and Cb values that correspond to skin color, which subsequently define our skin-color reference map. The ranges that we found to be the most suitable for all the input images that we have tested are and . This map has been proven, in our experiments, to be very robust against different types of skin color. Our conjecture is that the different skin color that we perceived from the video image cannot be differentiated from the chrominance information of that image region. So, a map that is derived from Cr and Cb chrominance values will remain effective regardless of skin-color variation .Moreover, our intuitive justification for the manifestation of similar

Cr and Cb distributions of skin color of all races is that the apparent difference in skin color that viewers perceived is mainly due to the darkness or fairness of the skin; these features are characterized by the difference in the brightness of the color, which is governed by Y but not Cr and Cb. With this skin-color reference map, the color segmentation can now begin. Since we are utilizing only the color information, the segmentation requires only the chrominance component of the input image. Consider an input image of pixels, for which the dimension of Cr and Cb therefore is . The output of the color segmentation, and hence

stage one of the algorithm, is a bitmap of size, described as

$$O_1(x,y) = \begin{cases} 1, & \text{if } [\mathrm{Cr}(x,y) \in R_{\mathrm{Cr}}] \bigcap [\mathrm{Cb}(x,y) \in R_{\mathrm{Cb}}] \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $x = 0, \cdots, M/2 - 1$ and $y = 0, \cdots, N/2 - 1$. The

The output pixel at point is classified as skin color and set to one if both the Cr and Cb values at that point fall inside their respective ranges and . Otherwise, the pixel is classified as non-skin color and set to zero. To illustrate this, we perform color segmentation on the input image of *Miss America*, and the bitmap produced can be seen in Fig. 4. The output value of one is shown in black, while the value of zero is shown in white (this convention will be used throughout this paper).

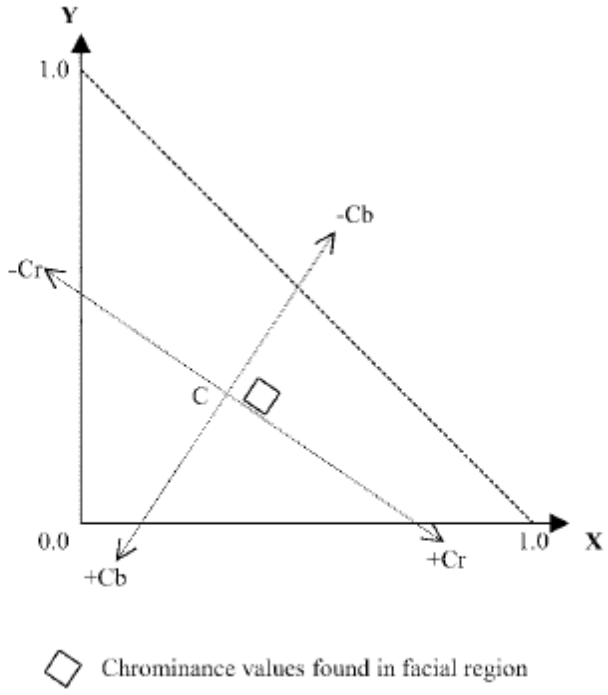


Fig. 3: Skin color region in CIE chromaticity diagram.

Fig. 4: Bitmap produced by stage one.

Among all the stages, this first stage is the most vital. Based on our model of human skin color, the color segmentation has to remove as many pixels as possible that are unlikely to belong to the facial region while catering for a wide variety of skin color. However, if it falsely removes too many pixels that belong to the facial region, then the error will propagate down the remaining stages of the algorithm, consequently causing a failure to the entire algorithm. Nevertheless, the result of color segmentation is the detection of pixels in a facial area and may also include other areas where the chrominance values coincide with those of the skin color (as is the case in Fig. 4). Hence the successive operating stages of the algorithm are used to remove these unwanted areas.

B. Stage Two—Density Regularization

This stage considers the bitmap produced by the previous stage to contain the facial region that is corrupted by noise.The noise may appear as small holes on the facial regiondue to undetected facial features such as eyes and mouth, or it may also appear as objects with skin-color appearance in the background scene. Therefore, this stage performs simple morphological operations such as *dilation* to fill in any small hole in the facial area and *erosion* to remove any small object in the background area. The intention is not necessarily to remove the noise entirely but to reduce its amount and size. To distinguish between these two areas, we first need to identify regions of the bitmap that have higher probability of being the facial region. The probability measure that we used is derived from our observation that the facial color is very uniform, and therefore the skin-color pixels belonging to the facial region will appear in a large cluster, while the skin-color pixels belonging to the background may appear as large clusters or small isolated objects. Thus, we study the density distribution of the skin-color pixels detected in stage one. An array of density values, called density map , is computed as

$$D(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} O_1(4x+i, 4y+j) \qquad (2)$$

where $x = 0, \cdots, M/8 - 1$ and $y = 0, \cdots, N/8 - 1$. It

It first partitions the output bitmap of stage one into nonoverlapping groups of 4 4 pixels, then counts the number of skin-color pixels within each group and assigns this value to the corresponding point of the density map. According to the density value, we classify each point into three types, namely, zero ( ), intermediate (0 16), and full ( ). A roup of points with zero density value will represent a nonfacial region, while a group of fulldensity points will signify a cluster of skin-color

pixels and a high probability of belonging to a facial region. Any point of intermediate density value will indicate the presence of noise. The density map of *Miss America* with the three density lassifications is depicted in Fig. 9. The point of zero density is shown in white, intermediate density in gray, and full density in black. Once the density map is derived, we can then begin the process that we termed as density regularization. This involves the following three steps.

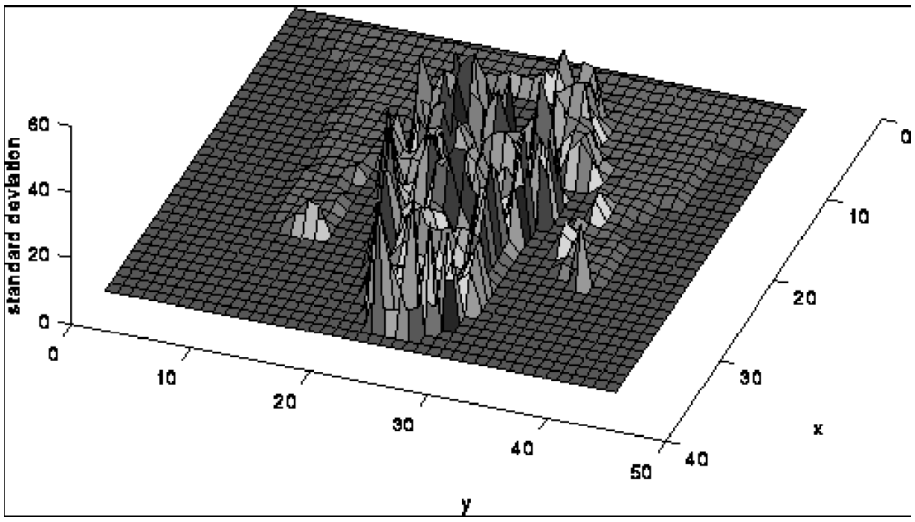
1. Discard all points at the edge of the density map,i.e set D(0,y)=d((M/8)-,y)=D(x,0)=D(x,(N/8)-1)=0 for all x=0,…….(M/8)-1 and y=0,……(N/8)-1
2. Erode any full_density point (i.e set to 0)if it is surrounded by less than 5 other full_density points in its local 3X3 neighborhood.
3. Dilate any point of either zero or intermediate-density (i.e set to6) if there are more than full-density points in its local 3X3 neighborhood.

After this process,the density map is converted to the output bitmap of stage two as

$$O_2(x,y) = \begin{cases} 1, & \text{if } D(x,y) = 16 \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

$$\text{for all } x = 0, \cdots, M/8 - 1 \text{ and } y = 0, \cdots, N/8 - 1.$$

Note that this bitmap is now four times lower in spatial resolution than that of the output bitmap in stage one.



C. Stage Three—Luminance Regularization

We have found that in a typical videophone image, the brightness is nonuniform throughout the facial region, while the background region tends to have a more even distribution of brightness. Hence, based on this characteristic, background region that was previously detected due to its skin-color appearance can be further eliminated. The analysis employed in this stage involves the spatial distribution characteristic of the luminance values since they define the brightness of the image. We use standard deviation as the statistical measure of the distribution. Note that the size of the previously obtained bitmap is hence each point corresponds to a group of 8 8 luminance values, denoted by , in the original input image. For every skin-color pixel in , we calculate the standard deviation, denoted as , of its corresponding group of luminance values, using

$$\sigma(x,y) = \sqrt{E[W^2] - (E[W])^2}. \qquad (4)$$

If the standard deviation is below a value of two, then the corresponding 8 8 pixels region is considered too uniform and therefore unlikely to be part of the facial region. As a result, the output bitmap of stage three, denoted as , is derived as

$$O_3(x,y) = \begin{cases} 1, & \text{if } O_2(x,y) = 1 \quad \text{and} \quad \sigma(x,y) \geq 2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$\text{for all } x = 0, \cdots, M/8-1 \text{ and } y = 0, \cdots, N/8-1.$$

The output bitmap of this stage for the *Miss America* image is presented in Fig. 12. The figure shows that a significant portion of the unwanted background region was eliminated at this stage.

D. Stage Four—Geometric Correction

We performed a horizontal and vertical scanning process to identify the presence of any odd structure in the previously obtained bitmap, , and subsequently removed it. This is to ensure that a correct geometric shape of the facial region is obtained. However, prior to the scanning process, we will attempt to further remove any more noise by using a technique similar to that initially introduced in stage two. Therefore, a pixel in with the value of one will remain as a detected pixel if there are more than three other pixels, in its local 3 3 neighborhood, with the same value. At the same time, a pixel in with a value of zero will be reconverted to a value of one (i.e., as a potential pixel of the facial region) if it is surrounded by more than five pixels, in its local 3 3 neighborhood, with a value of one. These simple procedures will ensure that noise appearing on the facial region is filled in and that isolated noise objects on the background are removed.

E. Stage Five—Contour Extraction

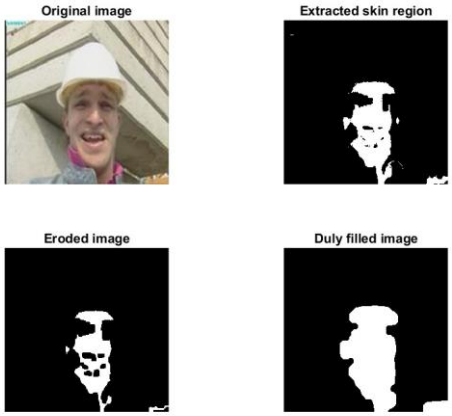
In this final stage, we convert the output bitmap of stage four back to the dimension of . To achieve the increase in spatial resolution, we utilize the edge information that is already made available by the color segmentation in stage one. Therefore, all the boundary points in the previous bitmap will be mapped into the corresponding group of 4 4 pixels with the value of each pixel as defined in the output bitmap of stage one.

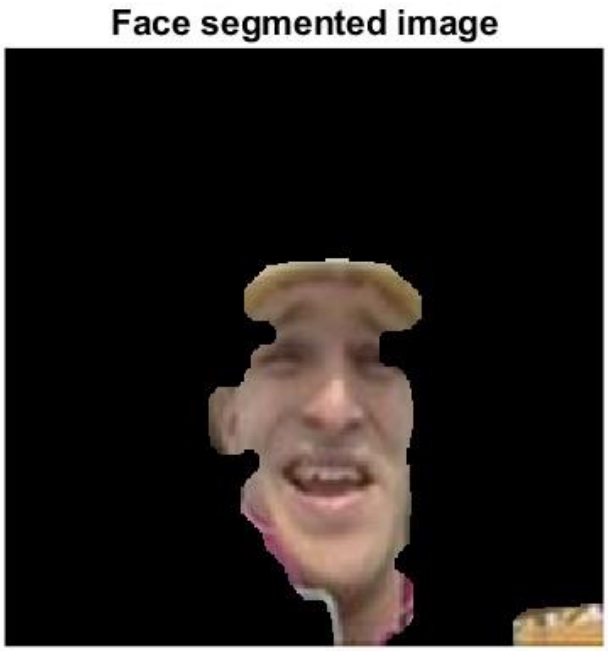
Fig. 5: Bitmap produced by stage four.

Fig. 6: Bitmap produced by stage five.

We then commence the horizontal scanning process on the "filtered" bitmap. We search for any short continuous run of pixels that are assigned with the value of one. For a CIFsize image, the threshold for a group of connected pixels to belong to the facial region is four. Therefore, any group of less than four horizontally connected pixels with the value of one will be eliminated and assigned to zero. A similar process is then performed in the vertical direction. The rationale behind this method is that, based on our observation, any such short horizontal or vertical run of pixels with the value of one is unlikely to be part of a reasonable-size and well-detected facial region. As a result, the output bitmap of this stage should contain the facial region with minimal or no noise.

4. Results

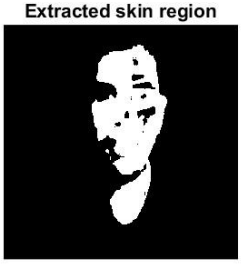




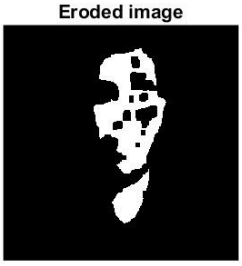
ISSN- 2394-5125 VOL 10, ISSUE 04, 2023

Original image

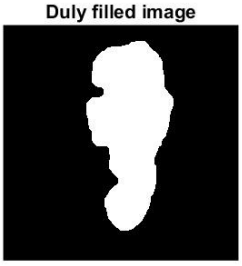
Extracted skin region



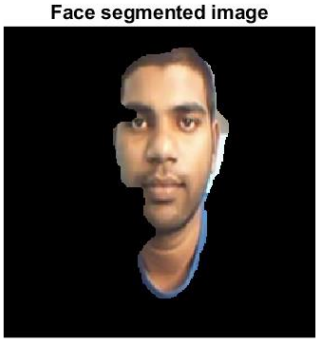
Eroded image



Duly filled image

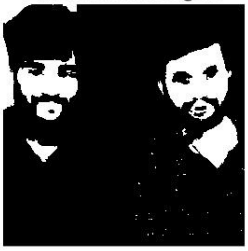


Face segmented image

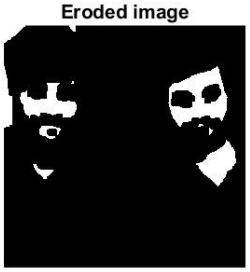


Original image

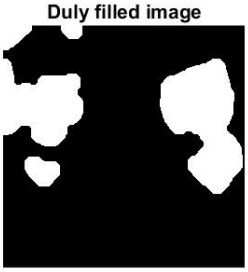
Extracted skin region

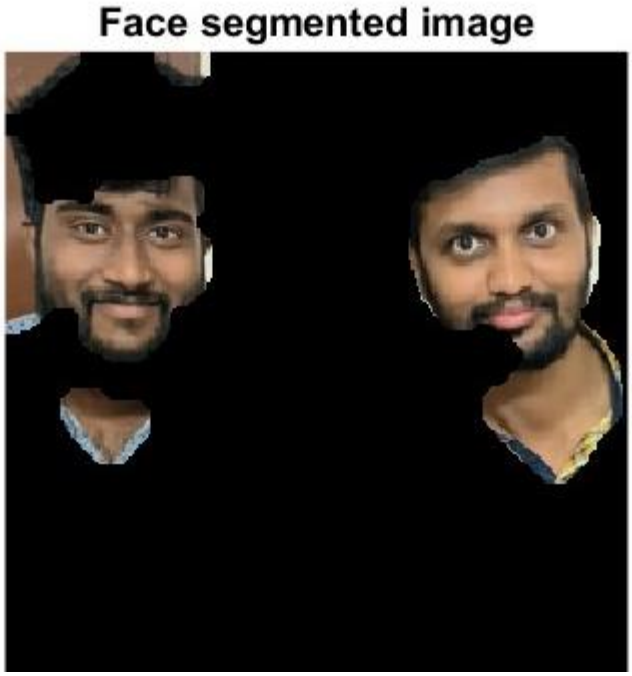


Eroded image



Duly filled image





5. Conclusion

In conclusion, this work introduces an interactive algorithm for automatically segmenting a person's face from an image that includes a head-and-shoulders view of the person along with a complex background scene. The algorithm leverages the spatial distribution characteristics of human skin color, providing a fast, reliable, and effective approach. The main objective of this research is to develop a system capable of locating a person's face within given image data, addressing the challenges of face location, extraction, and segmentation.

The significance of this problem is underscored by its broad range of applications, particularly in advancing human-to-human and human-to-machine communications. Face segmentation plays a crucial role in various fields, offering a content-based representation of images that can be utilized for encoding, manipulation, enhancement, indexing, modeling, pattern recognition, and object tracking. By accurately identifying and isolating the facial region, this research contributes to the development of sophisticated techniques for image analysis and understanding.

It is important to note that while face detection focuses on distinguishing between "images containing faces" and "images not containing faces," the present work concentrates specifically on face segmentation, involving the localization and contour extraction of a person's face within an image where its position is initially unknown. By addressing this task, the project lays the foundation for future advancements in the field, empowering various applications that rely on precise and robust face segmentation.

References

[1] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154.

[2] Rizvi, S. A. R., Rahim, M. S. M., Sulong, G., & Anwar, F. (2010). Real-time face detection and tracking using skin color segmentation and Kalman filter. International Journal of Computer Science and Network Security, 10(6), 16-22.

[3] Jain, A. K., & Chen, H. (2013). Facial feature detection and tracking from videos. In Handbook of face recognition (pp. 487-508). Springer.

[4] Fernández, R., Marfil, R., & Bandera, A. (2013). Facial feature detection based on skin color and depth information. Expert Systems with Applications, 40(9), 3642-3649.

[5] Kim, S., Kim, J., & Chae, O. (2015). Real-time face tracking using skin color and motion information. Signal, Image and Video Processing, 9(1), 201-208.

[6] Hu, H., Zhou, C., Yu, X., Zhang, Z., & Cui, H. (2016). Real-time skin color segmentation and tracking for hand gesture recognition. In Proceedings of the 2016 International Conference on Virtual Reality and Visualization (pp. 29-34). ACM.

[7] Feng, G. C., Li, J., & Jin, H. L. (2017). Real-time face tracking based on skin color and active contour model. Optik, 131, 139-145.

[8] Bhowmik, D., Islam, M. Z., & Islam, M. M. (2018). Real-time face detection and tracking using skin color segmentation and template matching. In 2018 21st International Conference of Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.

[9] Rahim, M. S. M., Bade, A., Anuar, M. N., & Chau, W. L. (2019). Real-time face detection and tracking using skin color model and active contour. IET Image Processing, 13(1), 16-25.

[10] Kaur, G., & Singh, S. (2020). Real-time face tracking using skin color detection and morphological operations. In 2020 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 107-111). IEEE.