Face Recognition and Skin Extraction under a Dynamic Video

Li-Hong Juang* and Yu Fang

Fujian Provincial Academic Engineering Research Centre in Industrial Intellectual Techniques and Systems, HuaQiao University, China

*Corresponding author: Li-HongJuang, Fujian Provincial Academic Engineering Research Centrein Industrial Intellectual Techniques and Systems, Engineering College, HuaQiao University, No.269, Chenghua North Road, Fengze District Quanzhou, Fujian 362021, P.R.China, Email: lipuu@qq.com

Abstract

As society develops further, the demand for more complex information increases. With this in mind, the use of the naked eye becomes unsatisfactory, as the desired information may present itself in a manner that requires more advanced technology in order to capture it. Therefore, we need to understand how to capture and confirm what it is. For the reason, in this paper, we demonstrate first how to capture the video information, deal, and store each frame as an image then by comparing the current frame with its previous frame, finally return a histogram matching algorithm value (Bhattacharyya Coefficient) for their similarity, meanwhile determine whether it is a fast-moving object. As an object appears on a dynamic video, the first frame differential dichotomy processes the image and disregards the background. The object then appears again in a second image via the same processing while a background image is taken simultaneously. These steps allow for the corresponding regions to be analysed so that the object is identified via a collaboration of the two separate images. Afterwards, face recognition and skin pigmentation are extracted and detailed. By using Face ++ interface for face and skin recognition experiment, after some tests, in the case of a good picture resolution, the recognition rate is about 70%; the skin color extraction is one kind of relevant color extracting, when it has a well-lit background, so the effect will be better.

Keywords: Similarity; Frame differential dichotomy; Face recognition; Skin color extraction

Abbreviations: FLD: Fisher Linear Discriminator; DLA: Dynamic Link Architecture; PCA: Principal Component Analysis

Introduction

Nowadays, the human vision is more and more limited with the development trend of the world. For observation, when an object flash too fast or small in some cases, the human's eye sometimes is unable to catch these things due to its structure and characteristics [1]. Because of these limitations of human eye, however people still want to acquire and understand these information, we need to take our eyes to observe things through the computer.

Scientists and researches utilize this study to detect, determine and track the video (static background) of dynamic objects. Video object recognition and tracking for the purposes of this aspect of computer vision research is not an alien concept, and then in recent years to identify objects and real-time tracking of the object of everyone's attention more and more technology, and now are usually intelligent wisdom under the circumstances, for the city, it is hoped that more and better urban environment, crime rates drop, so they know we are not satisfied with the recorded data, extract information not satisfied, we want to tap the regularity of the content in the video, for example, through a person's posture we can see this is a good man or a bad man hovering waiting crime of opportunity there that we need to face life video recognition. Do people face security control, you need a video device on the grade. This includes human face acquisition, data extraction and expression, and so on. By these methods, it is possible to recognize facial expressions more clearly [2-4]. There by understand this technology has a broad space for development and market, and thus produced for this research and understanding of the idea.

The object recognition is a very easy thing due to its real applications related on human vision, however, it is not an easy thing for a computer system [5-6]. Since the object recognition scene is complex and would change at any time, thatwill cause some difficulties for object identification and tracking by a computer. In addition, the most important thing is the recognition accuracy will drop a lot. So for this reason, this studymainly focus on understanding and exploring a dynamic video under a static background.

The main processes of this study first is through the video analysis and background information to calculate these image in video, then re-take the image of each frame by comparing their degree of acquaintance among images in video to determine whether there has been an object can be identified, and then after these return-related information, which will be marked on the feedback moment in the image frame for easy observation. If the object is identified and recognized the

facecase, it will make a different kind of feedback information such as a color recognition or the number of individual appearing in the video.

For the above process described, this paper will analyze and use a number of different algorithms and methodologies to deal with the various stages in order to understand each algorithm and compare the rate and efficiency of operations, and then improve them after summarizing conclusions. The operating platform is selected on the PC online using Open CV and MATLAB caculation.

Methods and Experiments

The present trend of object recognition

Identifying object in a video is belonging to aspects of the image recognition, for the purpose, it needs to experience three development stages: character recognition, digital image processing and recognition, object recognition. Text recognition began to be studied from 1950, and became more and more strong after the digital image storage capacity promoted, it has a convenience and a huge advantage to usbut is easy to distort, so people began to study the digital image processing and recognition methodologies to tackle it. The main object recognition refers to the three-dimensional world of objects and the environment perception and understanding, it is an advanced computer vision scope [7]. In the image recognition mode, there are three kind of development process ways to have been identified: statistical pattern recognition, structural pattern recognition, fuzzy pattern recognition [8].

In the past ten years,the objectrecognition in video has been the study core for many researchers. Up to nowdevelopment, there is indeed a great achievement, such as domestic face ++ and HW Cloud for the study of face recognition, and there are some domestic companies for the image recognition in the text and the image recognition for studying bank card numbers and other identification strings, even the network has a lot of relevant information.

It means that the image recognition technology at home and abroad is growing fast, however, for object recognition in video, this research is not very mature, because most of the studies are not very complex at specific scene collection and analysis, and do some non-specific or random scene still existing a certain degree of difficulty, that is, an adaptive performance is poor, if the target image has a strong noise, or a larger image defects, it will has usually no desired results [9-11].

Video deal & image similarity

When identifying objects in video, we first need to deal with the video, the video needs to be converted to the corresponding frame picture output which can be learned by image processing and recognition.When one image in video is converted into each frame, because the object may not always be necessarily to appear in the video, when detecting whether it is an object, many frames without the object appearing will be also included in the detected queue, for example, if a period of 20 s video only has 1swith an object appearing, then this will be the case many unnecessary image framesare detected, and thus will have a longer run hours and be inefficient. And we need to understand how to filter out these unwanted image frames. To solve the above problem, in this research,we comparetheir acquaintance between theseimage frame to filter out some useless image frame.

Image similarity algorithm introduction

Histogram matching

Histogram matching must first be a gray-scale, then transfer it into the desired histogram. It has the following three steps: 1) the histogram of original image is equalized of gray; 2) form the desiredhistogram; 3) reverse the first step conversion [12-14] .

When comparing the acquaintancedegree uses thehistogram matching, it usually calculates Bhattacharyya distance and Bhattacharyya coefficient,where the pasteurized factor is used to measure two discrete probability distribution. The pasteurized factor is the amount of overlap of two statistical samples' myopia calculation [15-16]. Why it is used for the processing image similarity,because its effect on the image acquaintance degree is the best. These formulasare the followingEquation (1) and (2). Bhattacharyya distance:

$$DB(p,q) = -\ln(BC(p,q)) \tag{1}$$

where q and p refer to the two discrete probability distribution in the number domain, BC refers to the pasteurized factor, Bhattacharyya coefficient as follows：

$$\text{Bhattacharyya} = \sum\{i = 1/n\}\sqrt{\left(\sum ai.\sum bi\right)} \tag{2}$$

where a and b are for the two samples, n is the number of sub-blocks, ai and bi are the a and b number in section i.

We start histogram based matching approach to compare their image acquaintance by the above description. Theimage must be first gray-scale processing, here we use MATLAB fuction to transform as the following statement: I = rgb2gray (M), where M is theinput image andI is the grayscale image referredto M. Then use the statement [Count, x] = imhist (I) to read the grayscale image histogram information, where Count is the histogram data of vector grayscale image and x is the corresponding color vectorrespectively, then go through the calculation of distance and coefficientformulas. Another method instead of MATLAB for grayscale images are as follows: For grayscale images, the image is actually through the weighted average of RGB three components to acquire the heaviest gray value as the following standard formula: gray = 0.11B + 0.59G + 0.3R, where is just a one kind of weight from the perspective of human physiology. As Figure 1(a)-(b) show the original image, then got the results as shown in Figure 2 observation.
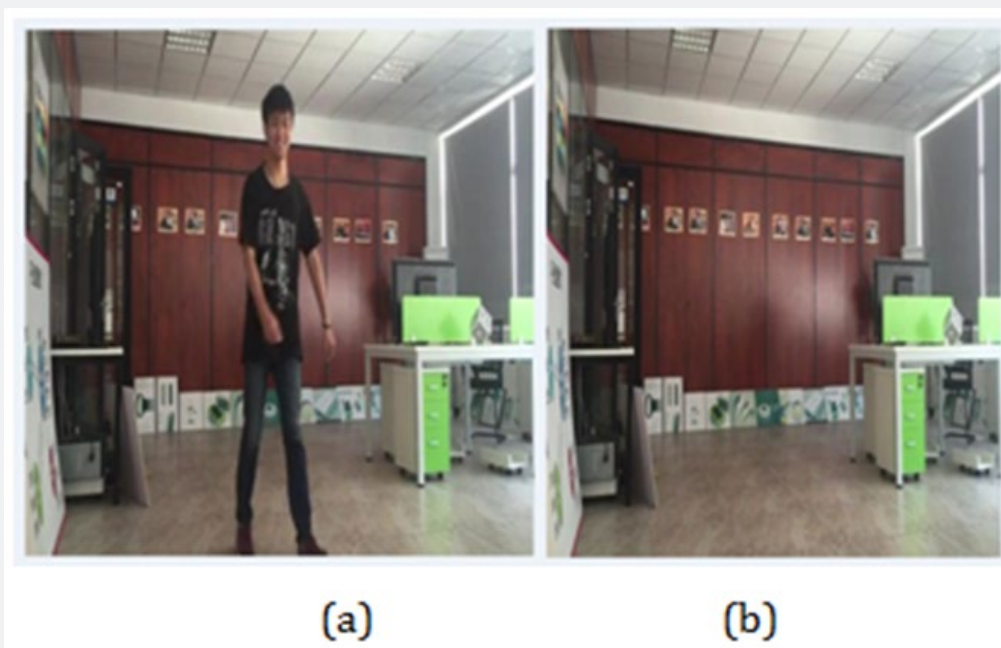
Figure 1: **(a)** Test 1 picture and **(b)** Test 2 picture.
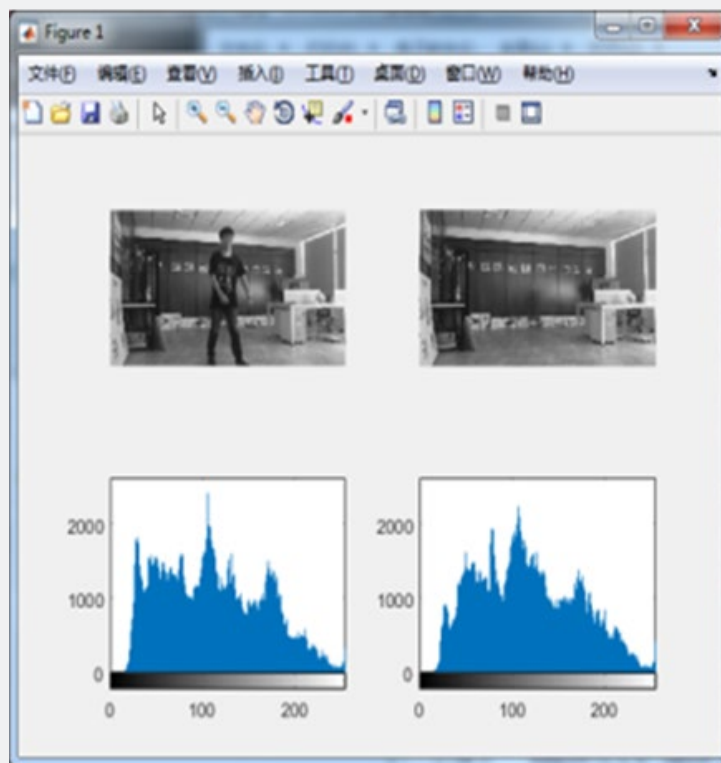

Figure 2: The results of histogram data.

As shown in these results, the hitis their respective histogram data referringthe pasteurized factor as the similarity of the two images. But the method is indeed obvious weak point, for example,one picture with above white color and below black color compared with one picture with aboveblack color and below white color, their similarity is 100, hit = 1.00, this is an apparent error.

Sift Transform Algorithm

Sift is also known as scale-invariant feature transform, which was developed by [9], and made further advance in 2004; its application includes object recognition, robot map perception and navigation, image stitching, 3D modeling, gesture recognition, video tracking and motion match; this

algorithm has been patent and ownedby the University of British Columbia.Sift algorithm consists of four steps:

a)   detection of the extreme value in scale space;

b)   use the neighborhood pixel at key points for the distribution parameters of gradient direction as each critical point in the specified direction, so that the operators pose rotational invariances;

c)   generate the sift feature vector and rotate the axis of feature point to ensure its rotational invariance;

d)   do the feature matching [17].

Perceptual hash algorithm

This is a kind of method that it will generate a 'fingerprint' string after each image is processed, and then make a comparison with it, if it is closer to the result, it would has much more similarity. Usually the first step of the algorithm is for image size shrinking to reduce the differences caused by the different image scaleand then simplifies color to gray image and compares each gray pixel again after calculating the average of gray value, If the mean greater is than or equal to the average, denoted by 1, others by 0.5, more thanassemblesas the above results, it is a so-called image 'fingerprint', finally followsup the same rule for the different images. The advantage for this method is that no matter how you deal with any image, change its size or color, its 'fingerprint' will not change.

With the above three algorithms, the latter two algorithms havea more powerful calculation and higher accuracy for these high complexity image than the first algorithm, howeverour study does not fix on the high image acquaintance, andwe just want to know whether there is an object in the video on the line, so that this study uses and implements a simple histogram matching algorithm.

Video background extraction

There are lots methodologies for video background extraction [18], for example, the time average, the multi-frame average, the codebook method [19], and the Gaussian mixture background model method [20]. Here we use to inter-frame difference method to realize it, which uses to images adding plus meaning to deal the video background extraction. Assuming the taken all image in frames are designated picture (N), where N refers to the N-th frame of video, specific implementation as follows:

For N=1：NumberofFrames

Backg=Backg+picture(N)

End

Backg=Backg/NumberofFrames

But when we have some real-time requirements for the background processing, then do as follows：

For N=1：mov.NumberofFrames

If（N<=NumberofFrames）

Backg=Backg+picture(N)

Else

Backg=Backg+picture(N)-picture(N-NumberofFrames)

End

End

Backg=Backg/NumberofFrames

The video background extractions can update real-time through the manner described above.

Mark motion objects

The object recognition is carried out by determining the differences between two images to confirm the object; it can be performed as follows:

C=picture（N）-Backg：

Figure 3: The picture（N）image in the programming.

Figure 4: Its background image.

Here it is described by two images as shown in Figure 3 & 4, and the result is shown in Figure 5. The image will be binarized after image subtraction, which uses iterative method based on the approximation value, first find the maximum and minimum gray value recorded as Rmax and Rmin, then make the threshold $T = (R\max + R\min)/2$, divide it into two groups R1 and R2 based on the mean gray value after the threshold

value, then obtain two average gray value of μ1 and μ2, finally obtain the new threshold value $T = (\mu 1 + \mu 2)/2$. Furthermore, let f (x, y) is a function of the input gray-scale image, and g (x, y) is a binary image output as follows:

$$g(x,y) = \begin{cases} 0, f(x,y) < T \\ 255, f(x,y) \geq T \end{cases} \quad (3)$$

Figure 5: The two image subtraction result.

Figure 6 shows the display after the manner described above, then extract the inside information from the image binarized result and feed back to the picture (N).
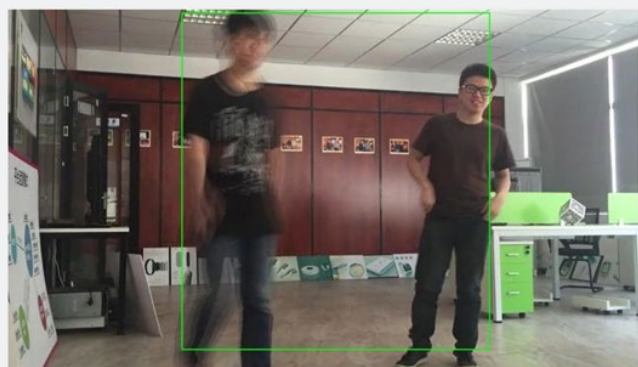

Figure 6: The final result of feeding back to the picture (N) on the display.

Mark the recognized objects

By the manner described above, we have successfully found the objects in video meantime label them, so now we want to identify items. Because each object has its own characteristics, even the same thing for basic terms that have different forms, for this reason, we will mainly fix the recognized object marks for face recognition.

Face Recognition

The researchers have a lot of attention for face recognition in the recent studies. Because the recognition from the image of a person's face is a much challenged project, the face size, orientation and posture in a image have different changes; even also some special cases, such as too bright and object occlusion cases; the above situations will affect the recognition efficiency. Here we introduce the definition of face recognition: it is to determine the location size and posture of all face in the input image if present; face recognition is a key technology of face information processing in recent years, and it has become a research subject within the field of pattern recognition and computer vision [14].

Principal components analysis

Principal Component Analysis(PCA) is for the optimal orthogonal transformation in image compression, [15-16] proposed the first PCA for face recognition application based on the Eigen face concept using the main component vector to reshape human face. The principle is based on the optimal orthogonal variation to expand network recursive manner to achieve its recognition. For this mode, it is assumed there are X categories in face gallery, and each category has Y face training images. Each image will be treated as a sample Pij (the jth people image in the ith category), where 1≤i≤X, 1≤j≤Y, and image is N * N, first the image icon will be vectorized into N^2 * 1, then the average training vector in X face image is as follows:

$$\mu = \frac{1}{X \times Y} \Sigma_{i=1}^{X} \Sigma_{j=1}^{Y} Pij \quad (4)$$

uses Pij-μ to acquire the mean difference vector eij, after that forms a N^2 * XY dimension matrix as follows:

$$A = [e_{11}, e_{12}, ..., e_{21}, ..., e_{XY}] \quad (5)$$

extract the eigenvalues and orthonormal eigenvectors of total scatter matrix AA^T through this matrix, denoted by B = AA^T, then introduce R = A^TA, and find out the corresponding eigenvalues λi and eigenvectors vi, so can obtain the feature vector:

$$\mu_i = \frac{1}{\sqrt{\lambda_i}} A v_i (i = 1, 2, ... X * Y) \quad (6)$$

Finally sort out λ_i, and select the corresponding eigenvector subspace according to requirement, thus complete the dimensionality reduction and feature extraction purposes [16], PCA uses the described above mathematics for face recognition. However, this approach has some shortcomings, such as light and size, etc., which will degrade its recognition rate.

Neural networks face recognition mode

Artificial neural network is a set of brain science, neural psychology and information science crossover studies, its algorithm is a mathematical model to simulate the human brain systems and applications; it is a lot number of processing units support each other with non-linear, adaptive information processing system [12-13]. This method can be for a low-resolution face image, partial autocorrelation function, and partial second-order matrix. The main advantages for this method are

a) simulate a person's thinking in image;

b) have a massively parallel and collaboration processing power;

c) strong self-learning ability and adaptability;

d) has a good fault tolerance;

e) have a non-linear mapping.

Elastic image matching face recognition

Elastic graph matching method is based on dynamic link architecture (DLA) method, which presents the face with a trellis sparse diagram, nodes in image use the Gabor wavelet of image position to obtain the feature vector mark called the jet, and the image margin is tagged with the distant vector. Its matching process flowchart is shown in Figure 7. Wavelet analysis is characterized by a time-frequency analysis, when the point in space at different frequencies in response to the surrounding area constitutes the point feature string, the high-frequency portion will correspond to the small details within the scope, and the low frequency part of the point will be a wide range around the scope. The wavelet transform elastic image matching algorithm takes into account both the face local detail and retains its spatial distribution, therefore, Gabor function is often used as a wavelet base function [9].

Figure 7: The matching process flowchart.
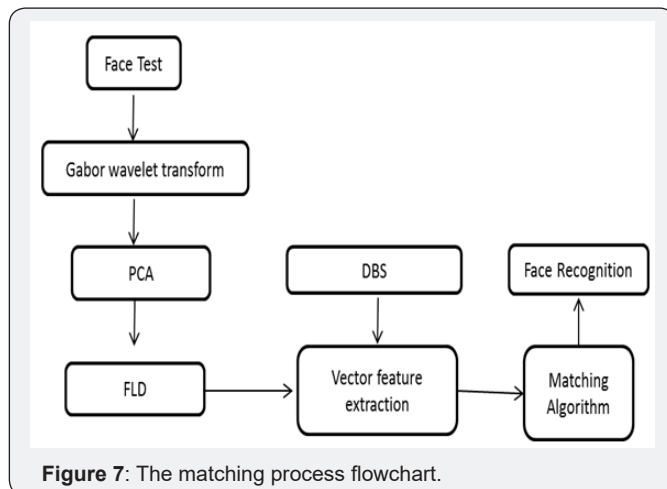
The Gabor transform is generated by an analog of the human visual system. The retinal imaging can be decomposed into a set of filter images by simulating the human visual system; each image can be decomposed to reflect these intensity changes in frequency and direction within the local area. The texture features can be obtained by a set of multi-channel Gabor filters, which actually is to design Gabor transforms, where fisher linear discriminator (FLD) refers to an improved feature face. Although this method can tolerate a certain degree of change in posture, facial expression and lighting, etc., but due to the high complexity of time and space, it is difficult to meet the requirements of large-scale real-time face recognition.

Face recognition method summary

At moment, there are many different algorithms for face recognition, not only just above described, for example, the Hidden Markov model, support vector machine (SVM) for face recognition [6], the line segment of Hausdorff distance for face recognition and human face recognition methods are based on skin color. Each method has its own advantages and disadvantages, however the present study, we tend to mix a variety of methods.

Experiment Tests

Main process

Here the main study is to implement whole operation for some complicated tests like multi-faces recognition and multi-skin color extraction under a video flow, the main process flowchart is shown in Figure 8. After each experiment, we will explore their effects.

Figure 8: The full test flowchart for face recognition and skin color extraction.
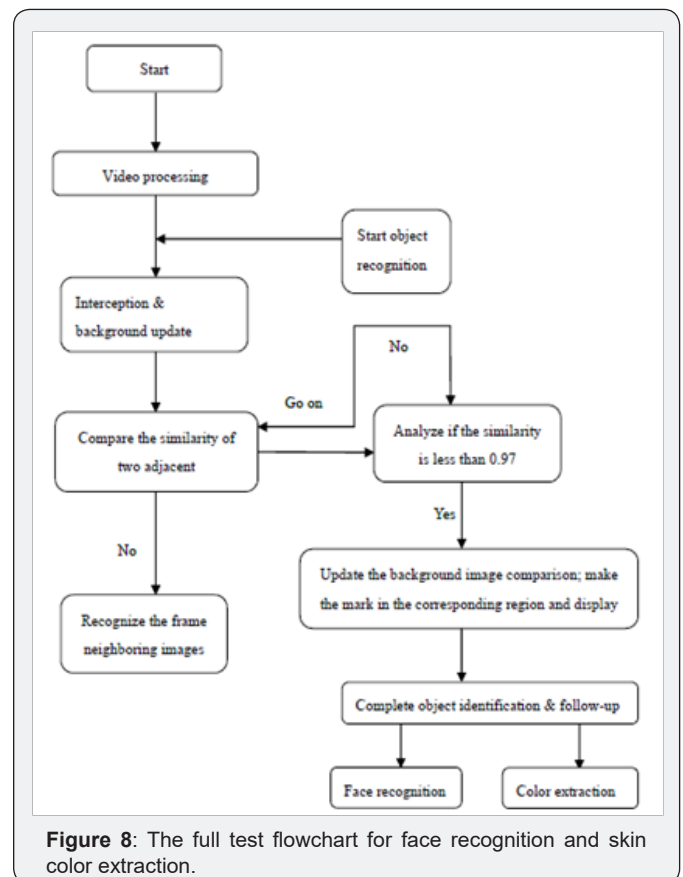
Video processing

Because this test video has no fast-speed objects, we use second speed rate for image processing in video and intercept their background by a few frames in a second. Now the video in seconds is intercepted, then their background is updated for image display in real time. In this research, the video is mainly static background, so we can directly use the background of this scene as the background, and it is updated in one second

time in several frames of video as it relates to the quality of the background image and the entire experimental results. Table 1 shows the image similarity rate for the filtered image collection after the moving object is identified.

Table 1: Image similarity comparisons for the filtered image collection in video.

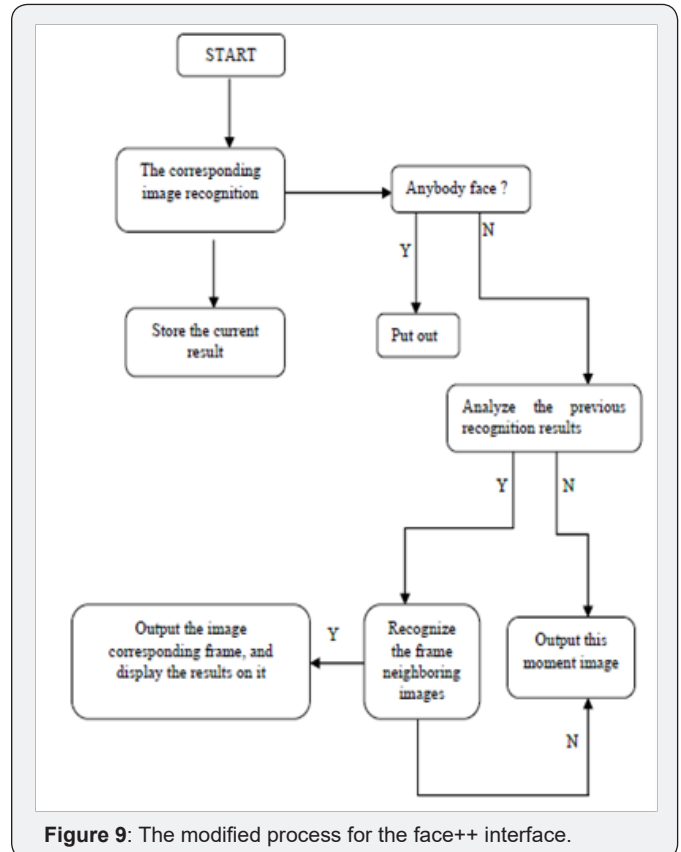| The Current Frame (The First Few Seconds of Video)	The Last Frame (The First Few Seconds of Video)	Bhattacharyya Distance (Image Similarity)
2	1	0.975782
3	2	0.973902
4	3	0.970653
5	4	0.902657
6	5	0.927859
7	6	0.951648
8	7	0.90608
9	8	0.92624
10	9	0.966241
11	10	0.955217
12	11	0.96784
13	12	0.933398
14	13	0.935977
15	14	0.927042

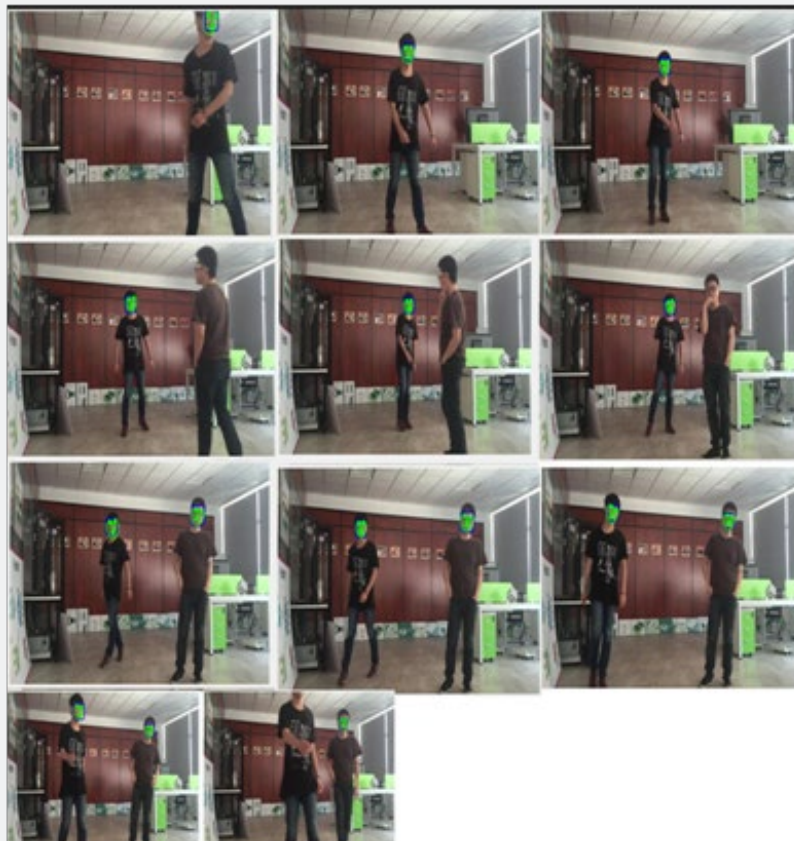**Figure 9**: The modified process for the face++ interface.

Figure 10(a): the face ++ plus recognition processing and post-processing effects,

Figure 10(b): the skin color region extraction is followed up (a).

The next recognition test is by the face ++ interface, we found that some time they still cannot be identified from these results and thus it is needed to do more processing. When it is confirmed in the previous image, in the frontal face appearing, and then if no face appears again through the process at the current image, then confirms it is in the different frame and identifies it. Furthermore, when the received human face number of Jason data from the face++ interface does not return 0, then put this frame as the current output. Therefore, we make a further process based on the modified process as Figure 9 shown and the processed output is shown in Figure 10 (a). When these faces are recognized by the above-described manner, the skin color region extraction is followed up and the results are shown in Figure 10 (b). Because the background color and the body color are somewhat similar, so some part of the background sometime will be also extracted.

The experimental tests show a more successful set of results. In general, the method described above is for the PC base, it can be seen in the study there is still some effect, although fast object recognition marked area is sometimes not very obvious and face recognition in many people's situation cannot be fully recognized and skin color is not very complete extracted. The main reason is that the software process speed cannot match with the hardware's.

Conclusion

In the study, we demonstrated some characteristics of the image operations in a video through some simple algorithms and methods and made some improvement such as a sift transform algorithm and perceptual hashing algorithm to fulfill. The face recognition rate can be promoted by increasing the image resolution, and then takes into account the speed, so the video resolution is set on 640 × 360. We can extract more background in frames by summing and averaging them, then the effect is better. For recognizing object in video, we try a running marked in real time, and face recognition and skin color extraction have been realized by extracting a video image to observe their effect and implemented some algorithms, the experimental tests show a more successful set of results.

Authors' Contribution

First demonstrate how to capture the video information, deal, and store each frame as a image then by comparing the current frame with its previous frame, and then analyze the object corresponding region marks on images then determine what the object is and meanwhile do the face recognition and color skin extraction.

Competing Interests

The authors declare that they have no competing interests.

References

1. Garcia C, Tzinitas G (1999) Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis. IEEE Transactions on Multimedia 1(3): 264-277.

2. Fox NA, Gross R, Cohn J, Reilly R (2007) Robust Biometric Person Identification using Automatic Classifier Fusion of Speech, Mouth, and Face Experts. IEEE Trans. Multimedia 9(4): 701–714.

3. Zeng Z, Tu J, Pianfetti BM, Huang TS (2008) Audio–visual Affective Expression Recognition through Multistream Fused HMM. IEEE Trans. Multimedia 10(4): 570–577.

4. Zeng X, Tu J, Liu M, Huang TS, Pianfetti B (2007) Audio-visual Affect Recognition. IEEE Trans. Multimedia 9(2):424–428.

5. Huang WM, Lee BH, Rajapakse M, Li LY (2002) Face Recognition by Incremental Learning for Robotic Interaction. In Proc. IEEE Workshop Multimedia Signal Process, pp. 280–283.

6. Wang Y, Yuan B (2000) Segmentation Method for Face Detection in Complex Background. Electronics Letter 36(3): 213-214.

7. Gao Y, Leung MKH (2002) Face Recognition Using Line Edge Map. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(6): 764-779.

8. Yang F, Paindavoine M (2003) Implementation of an RBF Neural Network on Embedded Systems: Real-Time Face Tracking and Identity Verification. IEEE Transactions on Networks 14(5): 1162-1175.

9. Hassaballah M, et al. (2010) Efficient Eye Detection Method Based on Grey Intensity Variance and Independent Components Analysis. IET Computer Vision 4(4): 261-271.

10. Lucey P, et al. (2011) Automatically Detecting Pain in Video through Facial Action Units. IEEE Transactions on Systems, Man, and Cybernetics 41(3): 664-674.

11. Tosato D, Spera M, Cristani M, Murino V, et al. (2013) Characterizing Humans on Riemannian Manifolds. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8): 1972-1983.

12. Kohli N, Yadav D, Noore A (2015) Multiple Projective Dictionary Learning to Detect Plastic Surgery for Face Verification. IEEE Access 3: 2572-2580.

13. Tan WR, Chan CS, Yogarajah P, Condell J (2010) A Fusion Approach for Efficient Human Skin Detection. IEEE Transactions on Industrial Informatics 8(1): 138-147.

14. Shoyaib M, et al. (2012) Skin Detection Using Statistics of Small Amount of Training Aata. Electronics Letters 48(2): 87-88.

15. Zafarifar B, Kerkhof TVD, de With PHN (2012) Texture-adaptive Skin Detection for TV and its Real-time Implementation on DSP and FPGA. IEEE Transactions on Consumer Electronics 58(1): 161-169.

16. Choi W, Pantofaru C, Savarese S (2013) A General Framework for Tracking Multiple People from a Moving Camera. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(7): 1577-1592.

17. Xu T, Wang Y, Zhang Z (2013) Pixel-wise Skin Colour Detection Based on Flexible Neural Tree. IET Image Processing 7(8): 751-761.

18. Korotkov K, Quintana J, Puig S, Malvehy J, Garcia R (2015) A New Total Body Scanning System for Automatic Change Detection in MultiplePigmented Skin Lesions. IEEE Transactions on Medical Imaging 34(1): 317-338.

19. Hariharan A, Adam MTP (2015) Blended Emotion Detection for Decision Support. IEEE Transactions on Human-Machine Systems 45(4): 510-517.

20. Sawicki DJ, Miziolek W (2015) Human ColourSkin Detection in CMYK ColourSpace. IET Image Processing 9(9): 751-757.