2029

Face Detection of Video Image Sequence Based on Human Characteristics^①

Yansong Liu^{1,*}, Yaopei Zhao^{1,*} and Zhenqiang Mi²

¹Department of Scientific Research, Shan Dong Management University, Jinan, 250357, Shan Dong Province, China ²School of Computer and Communication Engineering, University of Science and Technology Beijing, 100083, China

Abstract: The study uses DSP as the core and develops agile tracking and surveillance system. It includes overall design of the system, construction of the hardware and compilation of the software. The system hardware consists of DC, color camera, cradle head, SEED-VPM642 DSP development board or pan-tilt-zoom control system. The paper analyzes common algorithms of motion target detection. According to the characteristics of the acquired video images and the real-time requirements of the system, based on adjacent frame difference, the paper proposes continuous three-frame image difference method to detect moving objects and uses centroid algorithm as the core algorithm of human motion tracking. The paper uses the character of human complexion and applies face detection algorithm based on complexion, which can rapidly and accurately get the cleat face information of moving people. And the study selects the sum of the absolute value of gray difference as automatic focusing evaluation function of the design. Optimizing common focusing mountain-climb searching algorithm makes the system focus rapid and reliable, which can achieve the objective of automatic detection and trace on moving body and human head location.

Keywords: Machine vision, Frame difference method, Fuzzy clustering, Kalman filtering, Cost function.

1. INTRODUCTION

Detection and tracing of moving human body is an important subject in visual tracking field. In real life, a large number of significant visual information is included in the motion. Visual monitoring system consists of multiple cameras connecting to a group of television monitors. It demands the monitoring personnel to keep monitoring the screen to acquire the video information, which can get the corresponding conclusions and make the corresponding decisions through artificial perception and judgment. But it is an onerous job for the monitoring personnel to monitor many television monitors. Especially when there are a lot of monitory points, it is nearly impossible for the monitoring personnel to monitor completely. So the demand of people on intelligence of monitoring system is becoming increasingly urgent.

The major monitoring object of intelligent monitoring system is human, so it is important for intelligent monitoring system to study human motion detection and tracing. Based on fully studying recognition and tracking algorithm, the paper proposes a solution of human motion detection and tracing. With the aid of the power of digital signal processing technology, the subject uses DSP chip TMS320DM642 with high speed and high performance in TI Company as the core component to establish an agile tracing and surveillance system based on the characteristics of human body.

Face detection technology based on complexion segmentation is attracting people's attention. Complexion is important for face and is independent of detail features of face, and it is applicable for rotation and the variety of facial expression. Complexion has relative stability and is different from the color of most background objects. So there is advantage to use skin color features for face detection [1-3].

Skin color feature is described by skin color model in a certain color space. Skin color Gaussian model is established in YCbCr space, the colors of skin are clustered to separate the complexion and non-complexion area of images. And histogram statistics method is used to determine face area, which can separate face from the images. The paper firstly improves human motion detection and tracing algorithm and proposes face detection algorithm based on image difference and skin color detection for improving the intelligence of face detection [4, 5]. The following is an explanation on innovation. Firstly, the paper makes deep study on common algorithms, analyzes and summarizes the advantages and disadvantages of various algorithms through theoretical analysis and empirical verification. The paper improves adjacent frame difference method and proposes moving target detection method of adjacent three-frame image difference method, which can accurately and rapidly detect the objects. The subject studies common methods of image denoising in which the method combining morphological filter and scale filter is used to eliminate the noise of the images. The design uses the method based on projection to locate the moving objects, which can complete correct segmentation of motion target and prepares for the follow-up human target tracking. In YCbCr space, clustering features of complexion are studied, and nonlinear illumination compensation method is used

^{*}Address correspondence to these authors at the Department of Scientific Research, Shan Dong Management University, Jinan, 250357, Shan Dong Province, China; E-mail: mizq@ustb.edu.cn

to preprocess the images, after which similarity computation and binary segmentation of images are made. Histogram statistics method is used to determine face area, which realizes human head location of video images [6].

2. FACIAL COLOR PROCESSING

According to computer color theory, there are different expressions for a color, which forms different color spaces, and each color space has its background and application area. Skin color of human face has its own features and has different expressions in different color spaces, which makes the computer have different recognition capability and processing effect on skin color in different color spaces.

(1) Color space model

Tone, saturation and brightness can differentiate the features of colors. Tone means the color of light, and the light with different wavelengths has different colors. The tone of luminous objects depends on distribution characteristics of radiation spectrum. The tone of non-luminous objects is determined by absorption, reflection, transmission and the features of lighting source. Saturation means the depth and shades of color. The depth of saturation relates to the proportion of white color. The more the white color is added to pure color, the lower the saturation will be. So saturation reflects the degree of white color diluting some color. Brightness is the degree of shade of the light felt by human eyes. The less the energy of light wave is, the smaller the brightness is. The tone and saturation of color indicates the depth of color and is called chromaticity [7].

Computer graphics color space includes:

RGB model

RGB model is the most common color model. CIE selects red color (wavelength=700.0nm), green color (wavelength=546.1mn) and blue color (wavelength=435.8nm) as three primary colors of color system, which is RGB color coordinate system. Camera system usually applies RGB coordinate system and the color of images is expressed by the color coordinate system. In RGB model, the color expressed by human face is influenced by illumination. For the same point, the corresponding value (R',G',B') and (R,G,B) in different brightness is proportional.

$$\frac{R'}{R} = \frac{G'}{G} = \frac{B'}{B}$$

It means that the same point has different colors under the condition of different illumination, and only brightness is different. RGB color space can be translated into gray images directly through linear variation:

$$Gray = 0.299 \times R + 0.587 \times G + 0.114 \times B$$

(2) HSV model

HSV color model is a color system model proposed by Munseu and is often used by artists. For HSV, H means tone, S means saturation, and V means brightness. The format reflects the way of people observing color, and is conductive to image processing. For the processing of color information, the advantage of the format is that it separates brightness V and two parameters reflecting the features of color, and separates tone H and saturation S. When the features of color of objects, we need to know the clustering features of it in some color space, and the clustering feature is embodied in the features of color and is influenced by illumination variation. The direct influence of illumination variation on color of objects is luminance component V. So extracting luminance component from color and only using the tone and saturation which reflect the essential features of color for clustering analysis can have better effect, which is the reason why HSV format is often used for the study of color image processing and computer vision [8-10].

(3) YCbCr model

It is color representation model of CCIR601 coding scheme with quality standard of studio as the objective, and it is widely applied to color display of TV. The color system used for JPEG is the system. Y is the brightness information which is easy to be changed. Cb and Cr mean the chromaticity of blue and red, and they are stable under the condition of different illuminations. The advantage of the space is to separate the illumination component and chromaticity component, which reduces the relevance of both. The conversion formula of color space from RGB to YCbCr is:

$$\begin{pmatrix} Y\\Cb\\Cr \end{pmatrix} = \begin{pmatrix} 16\\128\\128 \end{pmatrix} + \begin{pmatrix} 65.481&128.553&24.966\\-37.797&-74.207&112.00\\112.00&-93.786&-18.214 \end{pmatrix} \begin{pmatrix} R\\G\\B \end{pmatrix}$$

(4) Rgb color model

Rgb model is normalized RGB model and the conversion formula is:

$$r = \frac{R}{R+G+B}$$
$$g = \frac{G}{R+G+B}$$
$$b = \frac{B}{R+G+B}$$

Because r+g+b=1, (r,g) space is only used, which realizes the conversion from three-dimensional space to twodimensional space. As RGB model is very sensitive to illumination, brightness information is not important for the construction of color model.

2.1. Space Selection and Modeling of Face Skin Color

Skin color is an important characteristic of face. There are studies indicating that the skin colors of people with different races, different ages and different sexes look different, but the difference is mainly manifested as illumination. In chromaticity space of removing illumination, the skin-color distribution of different people has clustering. Face can be separated from the background based on the clustering.



(a) Skin color in Cb-Cr space projection



(b) Skin color in YCbCr space

Fig. (1). Clustering feature of skin color in YCbCr.

In order to separate face area from complicated background, skin color model which is applicable to different skin colors and different brightness needs to be used. And ideal color component which is used for skin segmentation should make skin and non-skin area relatively concentrated in histogram of color component, each accounts for different distribution ranges.

Yong 21 studied clustering condition of skin color in various color spaces. The results indicate that skin color has the clustering feature in YCbCr space, as shown in Fig. (1).

Skin color model means that algebraic form, analytical form or look-up form are used to represent which pixels' color belong to skin color or to express the color of a pixel and similarity degree of skin color. Using skin color model to detect skin color is divided into two steps, establishment of model and application of model [11]. Establishment of model means to make statistical analysis on many skin color pixel sets and determine the parameters of model. Application of model means to use the established skin color model to judge if the inputted pixel or area is skin color, or get the similarity degree.

Gaussian model achieves a skin color probability graph of skin color by calculating continuous data information constructed by probability values of pixels and completes the confirmation on skin color according to the numerical, which overcomes the disadvantages of geometric model. And there is no need to consider the problem that non-skin color sample is difficult to be extracted from neural network model.

Two-dimensional Gaussian function used in the algorithm is:

$$P(C_r, C_h) = \exp[-0.5(x - M)^T C^{-1}(x - M)]$$

X is the value of sample pixel in YCbCr color space, M is sample mean of skin color in YCbCr space, and C is variance matrix of skin-color similarity model.



(a) The original color image



(**b**) Image after illumination compensation

Fig. (2). Non-linear light compensation

 $x = [Cb, Cr]^T$

$$M = E(x)$$

 $C = E((x - M)(x - M)^T)$

C. Realization process of face area detection

In the process of face detection, firstly, the inputted images are used for illumination compensation. Then, the similarity of images is calculated. The skin color area is separated from the background by threshold segmentations, which forms binary image. The third step is to use histogram statistical method to determine four sides of face and abandon non-face area, which detects face finally.

D. Light Compensation

In the study of face detection, most detected objects are digital images or digital image sequence of video flow which are collected by image capture devices. So acquisition conditions, especially light condition including the direction of light source, shade and color has great influence on the effect of images, which affects accuracy of face in color images. Therefore, it is necessary to make illumination compensation on images before face detection [12].

The system uses a new method which makes compensation on different illumination conditions to solve normal detection of face in color images under complicated illumination environment. The method fuses the solutions based on space under the condition of different illuminations which includes using color images illumination histogram balance to solve excessive brightness and shade of images, and using non-linear conversion to change the contrast of images for removing the shadows. The methods are as follows.

(1) Rectification of over-bright and over-dark area

If the brightness value of pixels of the original image is f(x,y), the total number of pixels of the original image is N, there are L gray levels (L=256), the gray value of the k level is rk, the frequency of gray level rk appearing is nk, and Pr(rk) is the probability of the k gray level appearing:

$$P_{r}(r_{k}) = n_{k} / N, 0 \le r_{k} \le k, k = 0, 1, 2, \dots, L-1$$
$$s_{k+1} \sum_{i=0}^{k} P_{r}(r_{k}), m_{k} = \sum_{i=0}^{k} n_{i}$$

When $S_k <5\%$ and $S_{k+1} >5\%$, if $m_k >100$ and f(x, y) approximates to gray value 0, B=f(x,y), which means that the brightness of all pixels in images is ranked from lower to higher, and the pixel of the previous 5% is taken. If the number mk of these pixels is enough such as it is greater than 100, the maximal brightness value of these pixels is B.

When Sk<95% and Sk+1>95%, if mk>100 and f(x,y) approximates to gray value 255, E=f(x,y), which means that the brightness of all pixels in images is ranked from higher to lower, and the pixel of the previous 5% is taken. If the number mk of these pixels is enough such as it is greater than 100, the minimal brightness value of these pixels is E.

(2) Rectification of middle gray area:

According to visual sensing model of human, in order to make the converted image soft and have clear level, and achieve the best visual effect, the histogram of the converted image g(x,y) should match with visual response curve of simulation vision of human. Literature [2] indicates that the brightness felt by human is logarithmic function of the incident light intensity of eyes. The transformation form is as follows:

$$g(x,y) = \frac{G_L - G_0}{\ln n - \ln m} [\ln f(x,y) - \ln m] + G_0$$

In the formula, g(x,y) is pixel brightness value of the transformed image, m and n mean the max gray and the minimal gray of the image before transformation, GL and G0 express the max gray and the minimal gray of the image after transformation.

E. Similarity Calculation

Gaussian clustering model of skin color in YCbCr space has been constructed. Two-dimensional Gaussian function formula 6 can be used to get the similarity of each point, which means that the possibility of the point belongs to skin color.

It is common that the similarity is between [0,1]. As Gaussian function in formula 6 has no coefficient, the value achieved by calculation is not completely in the interval, and the similarity needs to be calculated for unitization processing.

The variable is M, the similarity value of the first pixel initializes M, M=P(1,1), to judge the whole image. If P (i,j)>M, M=P(i,j) continues to make the judgment on the next point. After judging the whole image, the similarity of each point is divided by M to be as new similarity value, P (i,j)=P(i,j)/M. And the similarity values of all points in image are normalized between [0, 1]. We can see from the process that the similarity value is not similar to the empirical value, and it has relationship with images, which indicates that the algorithm can use similarity to segment skin color and non-skin color under different situation. And the method is adaptive.

In order to observe the distribution of similarity, the range of similarity is linearized to [0,255] from [0, 1]. And a threshold D can be gained through lots of practices and it can judge if the present pixel point is skin-color area.

$$D(C_r, C_b) = \begin{cases} 255 & P(C_r, C_b) > D\\ 0 & \text{else} \end{cases}$$

D (Cr,Cb) means binary image of skin color after segmentation. Experimental results are shown in Fig. (4). Fig. (4c) is light compensating results of Fig. (4a). Fig. (4b) is the segmented results of skin color in Fig. (4a). Fig. (4d) is the segmented results of skin color in Fig. (4c). The experiment indicates that mixing Gaussian skin color model effectively improves the clustering feature of skin color, and skin-color segmentation has good effect [13].

III. Image Smoothing Filter Technology

The objective of image smoothing filter is to eliminate noise. The existing smoothing filter technology used for digital image processing includes neighboring average filter, low-pass filter and median filter of frequency domain. It is inevitable for low-pass filter technology to make image details lost while filtering noise, which makes image fuzzy. With the expansion of low-pass filtering convolution mask, the image is more and more fuzzy. As a non-linear filtering technique, median filter not only can effectively filter impulse interference and image scanning noise, but also can overcome the disadvantage that the above filter makes image detail fuzzy. Meanwhile, filtering technology based on mathematical morphology is applied to the system and has good performance of eliminating noise.

A. Neighboring average filter

Neighboring average method is an algorithm of processing local spatial domain. If an image f(x, y) is the array of N*N and the image after smoothing is g(x, y), the gray level of each pixel is determined by the mean value of gray level of neighboring several pixels included in (x,y). And we can use the following formula to get smoothing image:

$$g(x, y) = \frac{1}{S} \sum_{(i, j \in A)} f(x, y)$$

In the formula, x,y=0, 1, 2,...,N-1, A is the neighboring center coordinate set of ((x, y) point which doesn't include point (x, y). S is the total number of coordinate points in A.



(a) Dark before compensation



(b) Skin color



(c) After compensation



(d) Skin-color segmentation

Fig. (3). Face detection effect.



Fig. (4). Common windows of median filter.

The algorithm is easy. The disadvantages of it is that it makes images fuzzy while reducing noise, especially at the edge and details, the greater the neighboring domain is, the greater the fuzziness is.

B. Median filter

Median filter is a non-linear signal processing method. Median filter not only can solve the problem that linear filter makes image detail fuzzy, but also has good effect on filter impulse interference and image scanning noise [14].

Median filter means that the gray median of the neighboring pixel is used to replace the value of the pixel. Median filter not only is a common non-linear filtering method, but also is the most common preprocessing technology of image processing technique. It has good inhibition effect on interference impulse and image noise.

Under one-dimensional form, median filter is a sliding window with an odd number of pixels. After sequencing, the window pixel sequence is:

$$\left\{F_{1-V},...,F_{1-1},F_{1},F_{1+1},...,F_{1+V}\right\}$$

V=(L-1)/2,L is the length of window, F1 is median filter output of window pixel. It is marked as:

$$G_i = Med \left\{ F_{1-V}, ..., F_1, ..., F_{1+V} \right\}$$

In the formula, $Med\{\cdot\}$ means to take window median. If the length of a window is 5, and the gray level of pixel is 20, 10, 30, 15, 25}, Gi=Med {10, 15, 20, 25, 30} =20.

The pixel whose gray level is 30 is pepper-salt noise which is eliminated after median filter.

It is easy to promote one-dimensional median filter to two-dimensional median filter, and the pixels of the window are ranked, which generates single two-dimensional data sequence $\{Fjk\}$ which is similar to one-dimensional data sequence. Two-dimensional median filter output G (j,k) is:

$$G(j,k) = Med\{F_{ik}\}$$

Generally speaking, two-dimensional median filter can inhibit noise better compared with one-dimensional median filter. The window of two-dimensional median filter has many shapes such as linear, square, cross, circle and diamond, as shown in Fig. (5). The windows with different shapes have different filtering effect. We should choose the shape according to the content of image and different requirements. Seen from the past experience, square or circle windows are applicable to object images which have longer



Fig. (6). Erosion.

outline. But crossing windows are suitable for images which have spire angel.

There are many flexible methods of using median filter to eliminate noise. One method is to use small-scale window firstly, and then enlarging the size of window. The other method is to use alternatively one-dimensional filter and two-dimensional filter. Median filter has many important features, for example,

(1) It has no influence on discrete step signals and ramp signals. Discrete impulse with length less than the length of window is smoothed. And the top of trigonometric function receives planarization.

(2) After median filter, signal spectrum is invariant.

C. Mathematical morphology filter

Mathematical morphology is based on set theory. It is a powerful means to analyze and describe geometry. In recent years, mathematical morphology has been widely applied to digital image processing and machine vision. Eliminating noise is a filtering problem. But it is a morphological analysis problem considered from the difference of form between noise and information. The objective of introducing mathematical morphology is to smooth the edge of binary image gained by inter-frame difference and to eliminate the influence of noise on skin color of face after space transformation [15].

Mathematical morphology includes four important operators, dilation, erosion, opening and closing.

Dilation

Dilation means using vector addition to combine two sets. Two operators of vector addition come from X and B, which can take any possible combination. The formula is:

 $X \oplus B = \{ p \in \varepsilon^2 : p = x + b, x \in X \boxplus b \in B \}$

In the formula, X is binary image, and B is structural element. The calculation results of dilation are shown in Fig. (6).

Dilation is used to fill in small cavity and narrow gaps of objects. It enlarges the size of objects. If it is necessary to keep the original size of objects, dilation needs to be combined with erosion.

(2) Erosion

Erosion means to use vector subtraction on set elements and combine two sets. Erosion is dual operation of dilation. Erosion and dilation are not reversible calculation. The formula is:

$X \bullet B = \{ p \in \mathcal{E}^2 : x + b \in X \}$, Every b $\in B$

In the formula, X is binary image, and B is structural element. The erosion results are shown in Fig. (7).

Erosion can be used to simplify the structure of objects. The objects which only have a pixel width need to be removed, which can divide complicated objects into several simple parts and can eliminate small noise points.

(3) Opening and closing

Erosion and dilation are not reversible operation. If an image is for erosion firstly, and then it is for dilation, we



(a) Image after illumination compensation



(b) Image after skin-color segmentation



(c) Operation of erosion firstly



(d) Operation after dilation

Fig. (7). Comparison of morphological filter.

can't get the original image. The result is that the image is easier than the original image.

Dilation after erosion is called opening operation: $X \circ B = (X \in B) \oplus B$

Erosion after dilation is called closing operation:

 $X \bullet B = (X \oplus B) \odot B$

Opening operation of each structural element is used to eliminate the details which are less than structural elements, and the local shape of objects keeps invariant. But closing operation is used to connect neighboring objects, fill in the cavity and narrow gaps, which makes the edge of objects smooth.

The difference between erosion, dilation and opening, closing is that opening and closing are not invariant for the translation of structural elements. Opening and closing change incrementally. Opening is a reverse expansion $(X \circ B \subseteq X)$ and closing is positive expansion $(X \subseteq X \bullet B)$. Opening and closing are dualistic transformation, which is the same to dilation and erosion.

The other important feature is that the result of repeatedly using opening and closing is idempotent,

 $X \circ B = (X \circ B) \circ B$

 $X \bullet B = (X \bullet B) \bullet B$

D. Comparison of filtering effect

Morphological filtering dilation and erosion is respectively applied to a binary image with the size of 245*285. The operation results indicate that mathematical morphological filtering effectively filters the noise of images, and keeps the content of interesting area. As shown in Fig. (8), there are effect figures of images after illumination compensation, face detection and morphological filter.

CONCLUSION

The paper studies face detection algorithm of color video images under complicated background. The algorithm uses basic characteristics of face images to divide face detection location into three steps, illumination compensation pretreatment, segmentation of skin color area, and using binary image morphological filter. The design and realization of face detection algorithm based on skin color includes illumination compensation, skin color modeling and morphological filter. The experiment indicates that the algorithm not only can realize face detection function in YCbCr under complicated environment, but also has great significance.

For the results of human motion detection, the face detection algorithm can make face detection on motion area, which can determine general human head position. And it can control the zoom and focus of lens, which gets clear human head information.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

Received: June 16, 2015

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- M. H. Yang, and N. Ahuja, "Detecting human faces in color images," In : *Proc IEEE Conference on Image Processing*, Chicago, PP.127-139.1998.
- [2] P. J. Phillips, H. Moon, A. S. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE TPAMI*, vol. 22, no. 10, PP. 1090-1104, 2000.
- [3] J. Buhmann, M. Lades, and F. Beckman, "Silicon Retina Object Recognition. Technical Report No 8596-CS, Bonn," Institute for Informatics University Bonn 1993.
- [4] D. Reisfeld, and Y. Yeshurun, "Robust detection of facial features by generalized symmetry," In: *Proceedings of International Conference on Pattern. Recognition, The Hague Netherlands*, PP.117-120, 1992.
- [5] J. Miao, B. C. Yin, K. Q. Wang, L. Shenb, and X. Chen, "A hierarchical multiscale and multiangle system for human face detection in a complex background using gavity-center template," *Pattern Recognition*, vol. 32, no. 7, PP.1237-1248,1999.
- [6] H. F. Chen, P. N. Belhumeur, and D. W. Jacobs, "In Search of Illumination Invariants," In: *IEEE Conf. on Computer Vision and Pattern Recognition*, PP. 254-261, 2000.
- [7] Texas Instruments, TMS320DM642 video/imaging fixed point digital signal processor. Texas Instruments Incorporated, 2003.
- [8] TMS320DM642 Technical Overview, Application Report, SPRU615, 2002.
- [9] TMS320C64x DSP Video Port/VCXO Interpolated Control (VIC) Port Reference Guide, Literature Number: SPRU629D, 2005.
- J. Segen, "A camera-based system for tracking people in real time," In: Proceedings of the 13th International Conference on Pattern Recognition, vol.3, pp. 63-67, 1996.
- [11] KaraulovaI, P. Hall, and A. Marshall, "A hierarchical model of dynamic for tracking people with a single video camera," In: *British Machine Vision Conference*, Bristol, UK, pp. 352-361, 2000.
- [12] S. Ju, M. Black, and Y. Yaccob, A parameterized model of articulated image motion," In: *Proc IEEE International Conference on Automatic Face and Gesture Recognition*, USA, pp. 838-844, 1996.
- [13] S. Wachter, and H. H. Nagel, "Tracking persons in monocular image sequences," Computer Vision and Image Understanding, vol. 74, no.3, pp. 174-192, 1999.
- J. Segen, "A camera-based system for tracking people in real time," In: Proceedings of the 13th International Conference on Pattern Recognition, vol. 3, pp. 63-67, 1996.
- [15] R. Polana, and R. Nelson, "Low level recognition of human motion (or how to get your man without finding his body parts)," In: *Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 77-82, 1994.

Revised: August 10, 2015

Accepted: September 19, 2015

© Liu et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the (https://creativecommons.org/licenses/by/4.0/legalcode), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.