Robust background subtraction in HSV color space

Ming Zhao Jiajun Bu Chun Chen
School of Computer Science, Zhejiang University, Hangzhou, 310027, P.R.China
Contact: {bjj, chenc}@cs.zju.edu.cn

ABSTRACT

In the new MPEG-4 video coding standard, automatic video object segmentation plays a key role in supporting object-oriented coding and enabling content-based functionalities. Background subtraction is one of the basic automatic video object segmentation methods. But various environmental illumination conditions often make it hard to work. A robust background subtraction method is presented in this paper. A statistical background model is first setup in this algorithm. Then the hypothesis testing is applied to the following frames to segment the video objects. The HSV color model is used and its color components are efficiently analyzed and treated separately so that the proposed algorithm can adapt to different environmental illumination conditions. Shadows are detected and a new background update algorithm is also presented based on the observation that the illumination changes are temporal and will not influence all the following frames. All of them contribute to the robustness of the method. The experimental results show that the proposed background subtraction method can automatically segment video objects robustly and accurately in various illuminating environments.

Keywords: background subtraction, video object segmentation, hypothesis testing, HSV color space, MPEG-4

1. INTRODUCTION

With the development of video coding technologies, it has evolved from pixel and frame based coding schemes such as MPEG-1 and MPEG-2 towards object based coding such as MPEG-4. The new features of the MPEG-4 coding standard include enhanced coding efficiency; content based interactivity and great error robustness for a large range of bit rates [1]. However, in spite of this advanced coding concept, the MPEG-4 visual coding mechanism always assumes that the video content to be coded is well represented in terms of video objects (VOs). So how to get the video objects, i.e. video objects segmentation, is key to MPEG-4 and its features.

The approaches to the video object segmentation problem are divided into automatic and semi-automatic segmentation of video objects. This paper will address the area of automatic video object segmentation. The basic approach to automatic segmentation is change-detection. It uses the luminance, color, texture or shape changes between frames to detect and segment the video objects. Change-detection can be further divided into adjacent frame subtraction and background subtraction. Though adjacent frame subtraction can adapt to luminance change with ease, it is hard to acquire the whole body of the video object due to the various movement of the video object. Edge information can be used to restore the body, but it is time consuming with low accuracy. However, background subtraction can solve these problems. It is to extract video objects from the static background scene. The idea of background subtraction is to subtract the current frame from a reference image, which is acquired from a static background during a period of time. The subtraction leaves only non-stationary or new objects, which include the video objects' entire silhouette region. However, a static background is necessary for background subtraction. Because background subtraction is fast and can produce good results, it is widely considered [2]-[5]. The basic steps of background subtraction are (1) Setup the background model or the reference image from a certain number of static background frames (with no video objects). (2) Subtract the current frame from the background model to get the video objects. (3) Post processing is used to acquire the accurate objects.

Horprasert et al. [2] proposed a background subtraction method using RGB color space. A background model basing on RBG color space is setup to separate brightness from chromaticity. During background modeling, the mean and variance of each pixel's color are calculated. Then for each pixel of the current frame, the brightness distortion and color distortion from the mean color are obtained. By these results, the current frame is divided into background, foreground, shadow and highlight region. This method can only adapt to little illumination changes, and only work in good environmental illumination conditions. The Pfinder [3] system in MIT used a method based on YUV color space.

It performed well only with little gradual illumination changes. If the luminance changes a lot, the result is not so good. Francois et al. [4] presented an HSV color space based background subtraction technique. It can produce good results. But it subtracted only the current frame from the background model. Thus, there is a lot of noise in the result. And it did not analyze the different property of each color component of the pixels and process them separately, which led to little robustness.

This paper proposes a robust background subtraction technique based on hypothesis testing using HSV color space. A statistical background is firstly setup. Then the hypothesis testing is applied to the following frames to segment the video objects. The HSV color model is used and its color components are efficiently analyzed and treated separately so that the proposed algorithm can adapt to different environmental illumination conditions. Comparing with the existing techniques, the proposed method has the following advantages: (1) Less noise and higher accuracy are achieved by means of the hypothesis testing. (2) Stability is enhanced by using the HSV color space which separates the brightness from chromaticity. (3) Robustness is guaranteed by analyzing the different properties of each pixel's color components and their statistical features, then utilizing them distinctively.

The rest of this paper is organized as follows. Section 2 introduces the background modeling based on HSV color space. And the hypothesis testing based background subtraction for video object segmentation algorithm is illustrated in section 3. Experimental results are presented in section 4 and concluding remarks are provided in section 5.

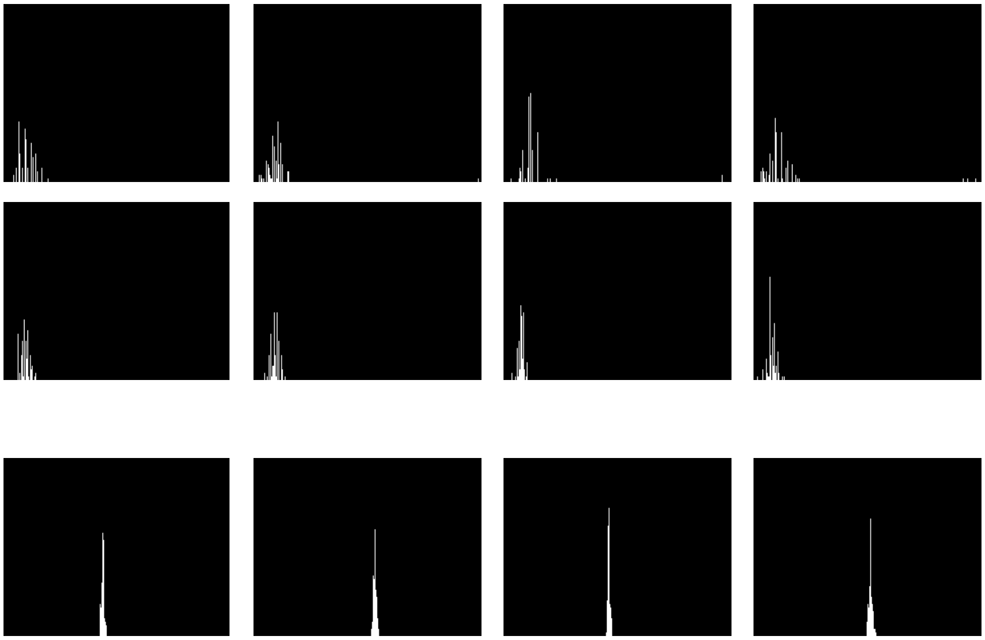


Figure 1 histograms of the H, S, V color components

2. REQUIREMENTS OF BACKGROUND SUGTRACTION FOR HSV

Most video capturing devices use RGB color space. But correlations exist among the three color components (R,G,B). All of them will change f the illumination changes. This will introduce instability. However, the HSV color space divides the color into three separate components: hue (H), saturation (S) and value (V). If the illumination changes, the hue will not change. By means of HSV color space, the algorithm will be more stable. So, in this paper, the HSV color space is used in order to separate the brightness from the chromaticity so that stability is enhanced. Though the transformation will cost some computing time, it is acceptable for current PC computing power.

2.1. Data Analysis - the properties of each color component in HSV

Figure 1 illustrates the properties of the color components in HSV of 4 pixels with coordinates (100, 120), (100, 180), (200, 120), (200, 180). Each of the pictures represents a histogram during the first 100 frames without video objects.

From the top to the bottom row, histograms of each row are those of H, S, V component respectively. And the histograms of each column are those of each pixel.

From figure 1, we can see that V is most stable during the modeling period. It approximates a normal distribution with small standard deviation. H and S are less stable than V. They may vary a lot, depending on their locations and the environmental conditions. Sometimes, they do not approximate a normal distribution, or they even vary with a large standard deviation.

After video object enter the scene, the environmental illumination will change. Though these changes may be too small to be noticed by our eyes, they will change H, S, and V to different degrees. We assume that the color (R, G, B) for a given pixel is changed to $(R+d, G+d, B+d)$ due to the luminance change. In the transformation of RGB to HSV, V is equal to the maximum of R, G, and B, i.e. $V = \max(R, G, B)$. So, V will change to the degree of d. H is proportional to $(X - Y)/(Max - Min)$, where X and Y are two of R, G and B, Max and Min are the maximum and minimum of R, G and B. Therefore, H will not change with R, G, and B. As S is equal to $(Max - Min)/Max$, it change with the degree of $\frac{Max-Min}{Max} - \frac{Max-Min}{Max+d} = \frac{Max-Min}{Max(Max+d)} \bullet d$, which is usually far smaller than d.

It is clear that H, S, V have their advantages and disadvantages respectively. We can conclude them as follows:

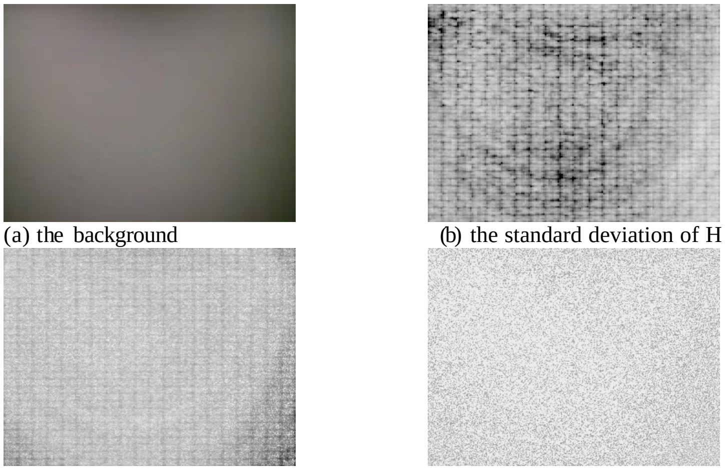
- H, S: their distributions vary a lot. Sometimes, they do not approximate a normal distribution or they vary with a large standard deviation. But H does not change with the illumination, and S only changes a little. Thereby, for pixels with stable distribution in H and/or S, H and/or S should be firstly considered for the background subtraction. But at first, we must make sure whether they are stable or not. If they not, they should be ignored. For this purpose, a background model should be setup. They will be discussed in section 2.2 and 2.3.
- V: its distribution is most stable. It usually approximates a normal distribution with a small standard deviation. But it changes with the illumination. In order that V could be efficiently used, background update must be enforced. This will be discussed in section 2.4. Because H, S are not sensitive to illumination change, they do not need background update.

2.2. Background model

As stated in section 1, the first step of background subtraction is to setup the background model or the reference image. The background model has three functions. The first is to acquire the varying features of H,S,V for each pixel in the background. The background subtraction decisions are based on these features. The second is to find out whether the variances of H,S are stable or not, as is discussed formerly in this section. The third is to subtract the current frame from this background model or reference image to obtain the video objects or the foreground. To achieve these purposes, a statistical model is used to model the background. Due to the white noise of the camera, Gaussian Normal Distribution can represent the distribution of each color component of the background's pixels. So the task of background modeling is to acquire the mean and standard deviation of the color components of the background's pixels. Such background model contains the normal variation range of the background. When the video objects enter the background, the variation of some pixels will overstep the normal variation range. So by detecting the variation of the pixels in the background model, the video objects can be segmented.

The first N frames, which are static background with no video objects, in the video stream are used to model the statistical background. These frames are denoted as B^i_{XK}, where $i \in [1..N]$ is the frame number, $X = (x, y)$ is the pixels' coordinates, and $K \in \{H, S, V\}$ is one of the color components. They are a sample complying with Gaussian Normal distribution. The sample mean of them is $\overline{B}^N_{XK} = \frac{1}{N}\sum_{i=1}^{N} B^i_{XK}$, and the sample standard deviation is $S^N_{XK} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(B^i_{XK} - \overline{B}^N_{XK})}$. Due to the fact that sample mean and sample standard deviation are the unbiased estimate of popular mean and standard deviation, they can be used as the mean and standard deviation of the color components of the background's pixels, i.e. $m_{XK}^{N} = \overline{B}_{XK}^{N}$, $s_{XK}^{N} = S_{XK}^{N}$. So for each pixel $X = (x, y)$ in the background, its mean and standard deviation of the color component $K \in \{H, S, V\}$ can be acquired by computing the samp le mean and sample standard deviation of the first N frames. All the mean $m_{XK}^{N} = \overline{B}_{XK}^{N}$ and standard deviation $s_{XK}^{N} = S_{XK}^{N}$ comprise statistical the background model.

2.3. Stability testing for H, S


(a) the background (b) the standard deviation of H

(c) the standard deviation of S (d) the standard deviation of V

Fig 2 .Illustration of the properties of the color components of HSV color space. (a) is the background . (b),(c),(d) are the standard deviation of H,S,V separately. The darker, the larger.

As discussed in section 2.1, the color components of the HSV color space have different stabilities. Not all of them are stable for all the pixels in the background. In general, V is most stable. It usually approximates a normal distribution with small standard deviation. H and S are less stab le than V. They may vary a lot, depending on their locations and the environmental conditions. Sometimes, they do not approximate a normal distribution, or they even vary with a large standard deviation. Figure 2 illustrates the standard deviations of H, S, and V for a blank background. Standard deviation is used to measure the stability. We can see that H and S are not stable. If one of them varies intensely enough and is used to detect the video objects, the result is likely erroneous. Fortunately, the V component is usually stable, as stated in section 2.1.

So, the stability must be checked for H and S for each pixel in the background. For this purpose, two methods are used in this paper. Only those which pass both the tests can be used for the background subtraction.

(1) Test of goodness-of-fit to a normal distribution. We have known that their distributions do not always approximate a normal distribution. On occasions that they do not approximate a normal distribution, we can not trust them because we do not know their distributions, and they will not be used for the background subtraction. In this paper, the test of goodness-of-fit is based on the sample skewness and kurtosis. For a true normal distribution, the sample skewness should be near 0 and the sample kurtosis should be near 3. The test determines whether the sample skewness and kurtosis are unusually different than their expected values, as measured by a chi-square statistic.

(2) Test of the standard deviation for H and S Though the distributions of H and/or S can approximate a normal distribution, they may vary a lot. In such cases, we can not trust them either. If any of them of a pixel is bigger than a threshold, i.e. $s_{XK} > T_{XK} (K \in \{H, S\})$, it will not be used for the background subtraction of this pixel to segment video objects.

2.4. Background update for V

As stated in section 2.1, the background should be updated in order that V could be efficiently used. The environmental luminance usually changes every now and then and the movement of the video objects could also change the local luminance in the background. And V will change with them. If the background model is always the same, some background can be detected as video objects when V is used for the background subtraction. So the background model should be updated over the time in order to adapt to such changes. Literatures [3][4] recursively updated their background models using a simple linear filter $m_t = ay + (1-a)m_{t-1}$, where a is the learning rate and y is the current color value. C.Riddle et al. [5] used the Kalman filter to adaptively estimate the background. Unfortunately, these methods are not suitable for the update of the static background, because the background does not change and only the luminance changes. So the original background model should not be modified. And what's more, the luminance changes are usually temporal and will not influence all the following frames. So another disadvantage of these methods is that they retain all the past changes in the background model. Therefore, these methods will distort the background model so much after some time that the background model cannot be used any more. A new algorithm is proposed in this paper to resolve such problems. This algorithm has two advantages: (1) the original background model will not be changed. (2) The background update only retains the latest changes of the luminance.

Let m_{XV} denote the latest mean for V, which is also considered as the latest background model. Only m_{XV} will be updated, for H, S do not sensitive to luminance change. So the original background model m_{XK}^{N} does not need updating. When background modeling finishes, m_{XV} is the same as m_{XV}^{N}. And let F_{XV} denote the current frame for V The proposed algorithm works as follows:

(1) Dilate the mask, which can be obtained by the background subtraction, containing video objects and their shadows thrice using mathematic morphology. The background subtraction will be discussed in section 3.

(2) For pixels $X = (x, y)$ belonging to the dilated mask, m_{XV} is updated with the difference which is linear interpolated. Let X_L denote the first pixel, which does not belong to the dilated mask, on the left of X. X_R, X_T and X_B are on the right, top and bottom in this way. Hence, $m_{XV} = \frac{1}{2}(I_H + I_V) + m_{XV}$, where

$$I_H = \frac{1}{|X_L - X_R|}\Big(|X_L - X|(F_{X_R V} - m_{XV}) + |X_R - X|(F_{X_L V} - m_{XV})\Big)$$ and

$$I_V = \frac{1}{|X_T - X_B|}\Big(|X_B - X|(F_{X_T V} - m_{XV}) + |X_T - X|(F_{X_B V} - m_{XV})\Big).$$

(3) For pixels $X = (x, y)$ not belonging to the dilated mask, m_{XV} is updated with F_{XV}, that is $m_{XV} = \mathrm{F}_{XV}$.

The proposed algorithm takes the temporariness of the luminance change into consideration, for it only update the latest background model with the current frame. The dilation of the mask ensures that the video objects or the foreground will not be updated as background. And the interpolation for the covered background makes it adapt to the luminance change. All these make the background update more robust.

3. HYPOTHESIS TESTING BASED BACKGROUND SUBTRACTION

In this paper, the hypothesis testing is used for background subtraction. The advantage of hypothesis testing is that it can retrain noise. Shadows also need considering for background subtract. At last, post-processing is applied to refine the results.

3.1. Hypothesis testing for video object segmentation

When the statistical background model has been setup, it can be used to test whether the pixels of the following frames belong to background or not. If not, they should belong to the video objects. The hypothesis testing is used for this purpose.

Let m_{XK}^{m} denote the mean of the color component $K \in \{H, S, V\}$ of the pixels at $X = (x, y)$ in certain continuous m frames. If the pixels are part of the background, m_{XK}^{m} should not differ a lot from m_{XK}^{N}. So the hypothesis about the mean is written as:

$H_0 : m_{XK}^m = m_{XK}^N$, $H_1 : m_{XK}^m \neq m_{XK}^N$

If the evidence level is a_{XK} , the rejection region is $| m_{XK}^m - m_{XK}^N | \geq z_{a_{XK}/2} s_{XK}^N / \sqrt{m}$, that is , if $m_{XK}^m \geq m_{XK}^N + z_{a_{XK}/2} s_{XK}^N / \sqrt{m}$ or $m_{XK}^m \leq m_{XK}^N - z_{a_{XK}/2} s_{XK}^N / \sqrt{m}$, the null hypothesis H_0 is rejected and alternative hypothesis H_1 is accepted, which means that there is too much difference between m_{XK}^m and m_{XK}^N to believe they are the same and it is rational to think that this is brought about by the video objects, i.e. at this time the pixels belong to the video objects or the foreground. The segmentation result is stored in a binary mask, in which 1 indicates the video object and 0 indicates the background.

When the background update is considered, m_{XV}^N will be replaced with m_{XV} for all the above discussion.

3.2. Shadow detection

The video objects could lay shadows on the background. But shadows are not part of the video objects. So they should be detected and eliminated. However, shadows are much different from the background. They can be easily detected as video objects by background subtraction as described in 3.1. They must be treated specially.

By analyzing the color properties of the pixels in shadows, we find that shadows have similar chromaticity with the background but lower brightness than those of the same pixels in the background model. In RGB color space, the brightness is not separate from the chromaticity. So RGB color space is not suitable for this problem. However, the HSV color space can separate the brightness from the chromaticity easily, because the value component (V) indicates the brightness and the hue and saturation component represent the chromaticity. Therefore, the HSV color space is very suitable for this purpose. Based on the HSV color space, shadows contain the pixels satisfying the following conditions:

$$m_{XV}^m - m_{XV} \leq -z_{a_{XV}/2} s_{XV}^N / \sqrt{m} \quad (1)$$

$$| m_{XH}^m - m_{XH}^N | < z_{a_{XH}/2} s_{XH}^N / \sqrt{m} \quad (2)$$

$$| m_{XS}^m - m_{XS}^N | < z_{a_{XS}/2} s_{XS}^N / \sqrt{m} \quad (3)$$

The formula (1) means that shadows should be dark enough comparing with their corresponding pixels in the updated background model. The formulas (2) and (3) mean that shadows have similar chromaticity with the original background model.

3.3. Robust background subtraction to segment video objects

Now the background subtraction algorithm is summarized as follows.

(1)Setup the statistical background model by computing m_{XK}^N and s_{XK}^N $K \in \{H, S, V\}$, for all pixels $X = (x, y)$ in the background from the first N frames, which are static background with no video objects, in the video stream. And stability is tested for H, S for all pixels.

(2)For each pixel $X = (x, y)$ of the current frame, m_{XK}^m is calculated and the following steps are executed.

(3)If $m_{XV}^m - m_{XV} \geq z_{a_{XV}/2} s_{XV}^N / \sqrt{m}$, X belongs to the video objects and go to step 8, otherwise go to the next step.

(4)If $m_{XV}^m - m_{XV} \leq -z_{a_{XV}/2} s_{XV}^N / \sqrt{m}$, go to step 7, otherwise go to the next step.

(5)If S is stable for pixel X and $| m_{XS}^m - m_{XS}^N | \geq z_{a_{XS}/2} s_{XS}^N / \sqrt{m}$, X belongs to the video objects and go to step 8, otherwise go to the next step.

(6)If H is stable for pixel X and $| m_{XH}^m - m_{XH}^N | \geq z_{a_{XH}/2} s_{XH}^N / \sqrt{m}$, X belongs to the video objects, otherwise X belongs to the background. Go to step 8.

(7)If $| m_{XS}^m - m_{XS} | < z_{a_{XS}/2} s_{XS} / \sqrt{m}$ and $| m_{XH}^m - m_{XH} | < z_{a_{XH}/2} s_{XH} / \sqrt{m}$, X belongs to the shadows, otherwise it belongs to the video objects. Go to next step.

(8)The latest background model m_{XK} is updated.

The segmentation result is stored in a binary mask, where 0 indicates the background and 1 indicates the video objects.

3.4. Post processing

As the segmentation noise is inevitable, post processing is needed to refine the segmentation results. Noise can be small and big. So we use a block labeling algorithm to work for this purpose. It can also fill the holes inside the video objects. The algorithm proposed in literature [6] is used to efficiently find all the blocks in the binary result mask. Then small blocks can be deleted as noise. For big blocks, which is regarded as video objects, we fill up their inside regions. At last, the close operation of mathematic morphology is used to smooth the final result.

4. EXPERIMENTAL RESULTS

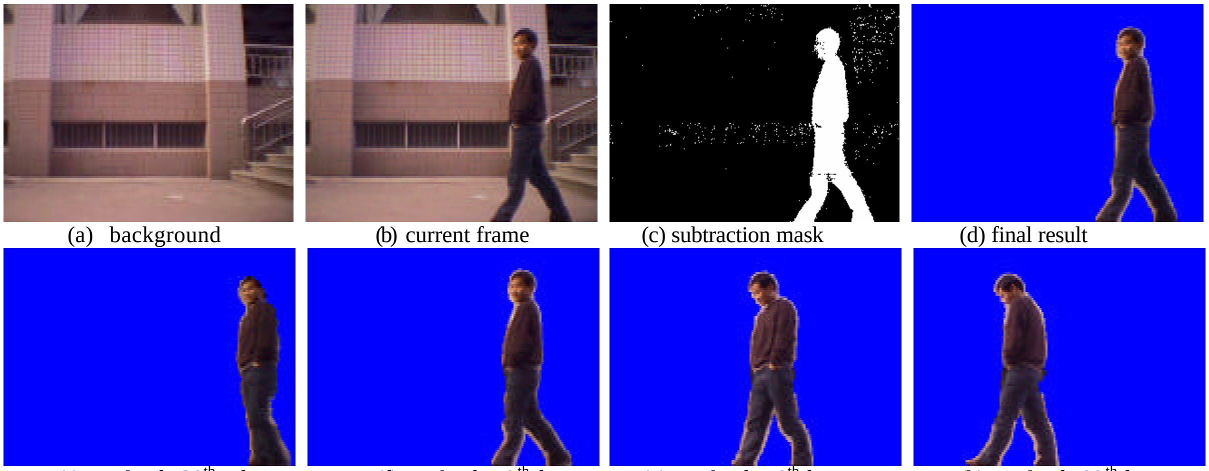
The proposed method runs at 120ms per frame of 320*240 on a Pentium-III CPU of 500. In order to illustrate the robustness of this method, we test our algorithm on different kind of videos. The first test video is an outdoor one. The illumination condition is very good in this video. There are nearly no luminance change and no shadows. The results are given in figure 3. From the results, it can be seen that there is little noise in the result of this paper and it can be eliminated to get accurate video object. The second test video is acquired indoors. In this video, the illumination condition is bad. The luminance changes as the video object moves. And there are shadows on the background. The results are given in figure 4. It can be seen that the algorithm of this paper is still robust and can produce accurate result in bad illumination conditions

5. CONCLUSIONS

A robust background subtraction method using HSV color model is presented to adapt to different environment and various illumination conditions. A statistical background model is firstly setup by calculating the means and variances of each pixel's color components from the first N frames without video objects. Then the hypothesis testing is applied to all the following frames. The null hypothesis is that the means of the color components of the M pixels with same coordinates in the M continuous frames are equal to those of the corresponding pixel in the background model. The alternative hypothesis is on the contrary. If the null hypothesis is rejected, these pixels in the current frames are detected as parts of the video objects. By this algorithm, less noise and higher accuracy can be achieved. The stabilities of each pixel's color components vary with their positions and illumination conditions. If one of the color components is not stable enough, it should not be used for the hypothesis testing, by which the method can adapt to different environment. A new background update algorithm is also presented based on the observation that the illumination changes are temporal and will not influence all the following frames. If the difference between the current pixel and that of the background model is big enough, the background model is updated. Otherwise, the background model returns to its original state. Shadow detection is considered and post processing is applied to refine the results. All the approaches contribute to the robustness of the background subtraction method. The experimental results are very satisfactory. This method can be used for object-oriented coding, content-based multimedia functionalities, video surveillance and video conference etc. The future work of this paper includes speeding up and boundary refinement.

REFERENCES

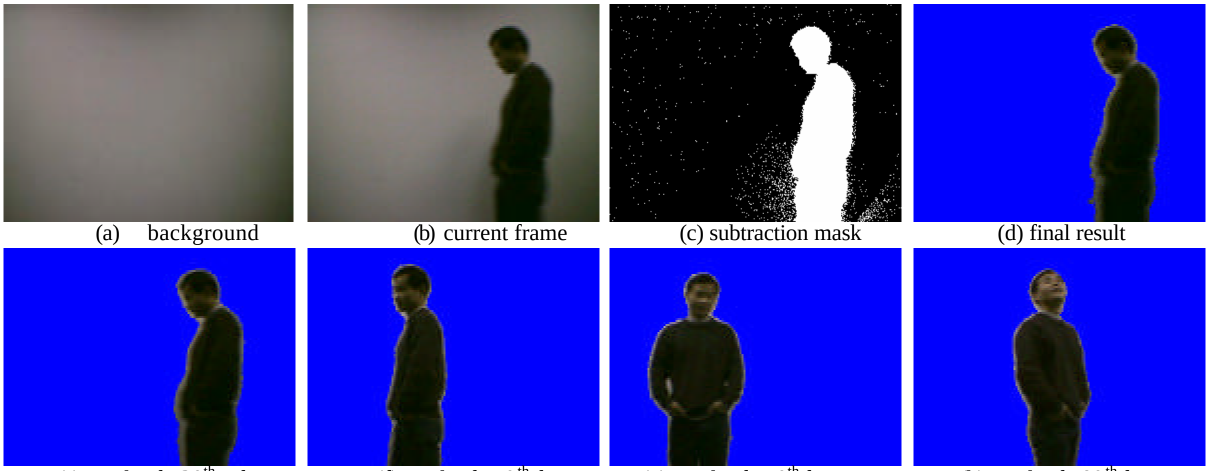
1. ISO/IEC JTC1/SC29/WG11, "Overview of the MPEG-4 Standard" , MPEG98/N2323, Dublin, July 1998
2. T. Horprasert, D. Harwood, and L.S. Davis, " A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection". Proc. IEEE ICCV'99 FRAME-RATE Workshop, Kerkyra, Greece, September 1999. http://www.eecs.lehigh.edu/FRAME/Horprasert/index.html
3. C.Wren, etc, "PFinder:Real-Time Tracking of the Human Body", IEEE Transaction on Pattern Analysis and Machine Intelligence, pp.780-785, 1997
4. A. Francois, G. Medioni, "Adaptive Color Background Modeling for Real-Time Segmentation of Video Streams", Proc. of International on Imaging Science, Systems, and Technology, pp.227-232, 1999.
5. C. Ridder, O. Munkelt, H. Kirchner, " Adaptive Background Estimation and Foreground Detection Using Kalman-filtering" , Proc. ICRAM'95, pp.193-199,1995.
6. Jae-Chang Shim, Chitra Dorai, " A Generalized Region Labeling Algorithm for Image Coding, Restoration, and Segmentation" . ICIP99, pp.46-50,1999.



(a) background (b) current frame (c) subtraction mask (d) final result

(e) result of 130th frame (f) result of 150th frame (g) result of 170th frame (h) result of 190th frame

Figure 3. the segmentation results for outdoor video stream (a) background (b) current frame (c) subtraction mask (d) final result of the current frame (e)-(h) final result of frames: 130, 150, 170, 190



(a) background (b) current frame (c) subtraction mask (d) final result

(e) result of 130th frame (f) result of 150th frame (g) result of 170th frame (h) result of 190th frame

Figure 4. the segmentation results for indoor video stream (a) background (b) current frame (c) subtraction mask (d) final result of the current frame (e)-(h) final result of frames: 130, 150, 170, 190