

# 行政院國家科學委員會專題研究計畫 期中進度報告

## 多媒體影音高階處理、傳輸及設計--總計畫:多媒體影音高 階處理、傳輸及設計(2/3) 期中進度報告(精簡版)

計畫類別：整合型  
計畫編號：NSC 95-2221-E-002-188-  
執行期間：95年08月01日至96年07月31日  
執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：貝蘇章  
共同主持人：陳良基、李枝宏、馮世邁、貝蘇章

報告附件：國外研究心得報告

處理方式：期中報告不提供公開查詢

中 華 民 國 96 年 06 月 14 日

**多媒體影音高階處理、傳輸及設計--總計畫 (2/3)**  
**Multimedia Audio/Video High Level Processing,**  
**Transmission and Design (II)**

**計畫編號: NSC 95-2221-E-002-188**

**執行期限: 95 年 8 月 1 日至 96 年 7 月 31 日**

**主持人: 貝蘇章 台灣大學電機系教授**

**摘要**

現今所有的數位影像與視訊幾乎都轉成壓縮格式而進行資料儲存。因此針對壓縮格式的資料進行移動物體的抽取，將可以有效減少運算處理，本論文提出的方法可以有效率區別前景與背景抽出移動物件，並對照明的改變與雜訊較不敏感及影響。

**ABSTRACT**

Research of image segmentation has been studied for many years. Image segmentation techniques are important but difficult in many image processing topics, such as object recognition and content-based image retrieval. In order to solve those problems, a successful image segmentation method is essential for splitting an image into meaningful regions, such as the discrimination between foreground and background of the segmentation of moving object and constant background. Then, make again these fundamental units (such as pixels or blocks) into further significant processing.

The task of Segmentation in video sequence based on DCT domain is dividing the video sequence into frames, and dividing 8X8 pixels element block each frame. Then passes through a continually string to calculate after determining this block is belongs to the foreground or the background. The main framework is to transfer each 8X8 block into 2 dimension-DCT domain in order to get the information of frequency domain, and then utilize the relation in the identical position block to calculate the threshold of background and foreground. This method has the quite good performance of catching moving object and also show to be low sensitive to illumination change and to noise.

**1. INTRODUCTION**

Segmentation in Video sequences based on DCT domain is an algorithm of catching moving object without background in gray level intensity. With many obstacles, accurate foreground segmentation is a difficult task due to such factors as illumination variation, occlusion, background movements, noise and

so forth. We present an adaptive transform domain approach for foreground segmentation in video sequence; a set of Discrete Cosine Transform based features is employed to develop the correlation between the spatial and temporal video sequences. We sustain an adaptive background model and make a decision based on the distance between the features of the current frame and those of the background model. Additional higher level processing is employed to deal with the variations of the environment and to improve the accuracy of illumination change and noise. It also overcomes many common difficulties of segmentation such as moving background objects or multiple moving objects with steady background. This algorithm also can perform in real time with fast calculation.

**2. RELATED WORK**

Video object segmentation and tracking is relevant to information extraction in many multimedia applications such as video surveillance, traffic monitoring, and security monitoring. No matter in temporal domain or in spatial domain, many approaches have been proposed to segment the foreground moving objects in video sequence. Several methods such as Ref [1]-[6] are more complicated to make the algorithm vigorous to a change in illumination or in the background adaptive background modeling approaches.

For spatial domain methods, some efficient and reliable segmentation algorithms in spatial domain that aim at providing object segmentation in real time have been proposed. In Ref [1], the basic ideal is to detect moving objects by object boundary rather than pixel difference. The algorithm consists of a reliable adaptive background model and fast edge based detection schemes. Kim and Hwang[2] developed a novel automatic object segmentation algorithm based on inter frame edge change detection that incorporates spatial edge information into the motion detection stage and overcomes the disadvantages of conventional change detection methods. Kalman filtering based methods Ref [3] recover slowly to sudden lighting change. These methods are robust to illumination change however; all

these methods based on spatial domain have more computation in than temporal domain with block-calculation so that we have to find more efficient ways to compute segmentation in video sequence.

For temporal domain, Bescos and Menendez [4] have proposed a method which focuses on a module that detects the number of people that crosses an entrance to a big store, so that the approach is oriented to be hardware implemented in order to allow for low cost real time operation. In Ref [5], the authors provided a background subtraction algorithm that uses multiple competing Hidden- Markov-Models(HMM) over small neighborhoods to maintain a locally valid background model in all situations. The use the DCT coefficients of JPEG encoded images directly to minimize computation and to use local information in a principled way. Region level processing is reduced to the minimum so that the extracted information that goes to higher level processing is unbiased.

All above approaches use the intensity, color information of the pixels or temporal domain data, and are susceptible to sudden lighting illumination changes. Recently, efforts have been made to incorporate other illumination robust features for scene modeling. The intensity and texture information can be integrated for change detection, with the texture based decision taken over a small neighborhood [6]. The color and gradient information can be fused and robust results have been reported [7]. However the computation of these algorithms is often too intensive for real time implementation.

### 3. THE PROPOSED METHOD

First of the all, each frame is divided into 8x8 block according Ref [12]. Two features, fDC and fAC of each block, are determined from the 2D-DCT coefficients and each of them is modeled as independent Gaussian distribution with different mean  $\mu$  and variance  $\sigma^2$ . These means  $\mu$  and variances  $\sigma^2$  are estimated initially from a first few frames, possibly with some moving object, and then they are updated depending on the scenery evolution.

Segmentation of a new frame is obtained initially by calculating the thresholds of these feature values ( fDC and fAC ) which contain illumination change detection and texture difference detection, then refined the result using size filtering, merging blocks and filtering. Based on the final segmentation of the current frame, the parameters of means  $\mu$  and variances  $\sigma^2$  are updated according to the segmentation result of the next frame.

With 2D-DCT domain, block based features incorporate local neighborhood information such as illumination change and texture difference between neighbor blocks, so the algorithm using these features is less sensitive to shadow, reflection of mirror-like

wall, lighting illumination change, noise, small vibration of background movement and small scenery changes than those using pixel value. Furthermore, the calculation of block base is much lower than that of pixel base, so the algorithm of DCT based feature is faster to become real time.

#### 3.1. The features of dct domain

We compute two features fDC and fAC which are derived from the 2D-DCT domain coefficients of each 8x8 block within each frame.

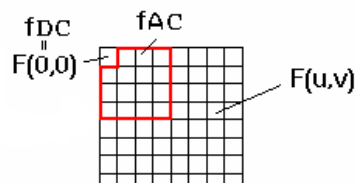


Fig 2. Definition of fDC and fAC

$$fDC = DCT(0,0) \dots \dots \dots Eq(3)$$

$$fAC = \sum_{i=0}^3 \sum_{j=0}^3 (i^2 + j^2) abs(DCT(i, j)) \dots \dots Eq(4)$$

Where  $DCT(i,j)$  is the  $(i,j)^{th}$  coefficient of 2D-DCT domain.

The DC coefficient fDC is simply the DC coefficient without any frequency component, and according to the Eq(2), fDC is the average gray level intensity of the block and sensitive to large illumination changes. It is computationally easier to implement and more efficient to regard the 2D-DCT as a set of basis functions which given a known input array size (8 x 8) can be pre-computed and stored. This involves simply computing values for a convolution mask (8 x8 window) that get applied (average values x pixel the window overlap with image apply window across all rows/columns of block).

The AC coefficient fAC is a weighted sum of the lower frequency AC component without any DC component. Much of the signal energy lies at low frequencies; these appearing in the upper left corner of the 2D-DCT and the high frequency coefficients are not used because they are more sensitive to noise and they relate to fine detail that are more susceptible to small illumination changes so that the AC coefficient fAC has a great stress on texture differences between neighbor blocks.

In Eq(4), the weight  $(i^2 + j^2)$  de-emphasizes the very low frequency components, which often correspond to smooth transition of intensity such as shadow, reflection of mirror-like wall and small vibration of background movement. The emphasis of middle frequency can help fAC concentrate on texture difference between blocks and avoid shadows being taken as part of foreground object.

### 3.3. The feature models of dct domain

An important step of segmentation is the determination of background model. Various approaches have been proposed for modeling video in both spatial and temporal domain, as well as modeling video in derived from video. The choice of a good model must balance various factors, including what quantities are to be modeled, computation requirement, speed of parameter adaptation and how it affects the final performance of the particular application.

We compute features fDC and fAC of each block, instead of the pixel value. As mention before, fDC emphasizes on large illumination changes and fAC is sensitive to texture difference between neighbor block. We assume that each model of fDC and fAC is a single non stationary Gaussian random process. That means each fDC's and fAC's distribution varies with time  $t$  and the means and variances of the parameters change every single frame.

$$f_{DC}(t) \sim N(\mathbf{m}_{DC}(t), \mathbf{S}_{DC}^2(t)) \dots\dots \text{Eq(5)}$$

$$f_{AC}(t) \sim N(\mathbf{m}_{AC}(t), \mathbf{S}_{AC}^2(t)) \dots\dots \text{Eq(6)}$$

So, we have to update those means and variances every single frame and extensive experiment was carried out using video captures under different conditions. Although the use of mixture of Gaussian distribution has shown a better fit than single Gaussian mode, we don't have to care much about the small fitness by adding models in the probability distribution function because the presence of foreground objects always bring significant difference from the background distribution. Moreover, Mixture of Gaussian distribution is slow to adapt to fast changing sequence and needs more computation and memory.

We discovered that the use of a single non-stationary Gaussian model as Eq(5) and Eq(6) with a flexible policy allowing for fast and slow adaptation of parameters can accurately model the evolution of these features and ultimately can lead to accurate foreground segmentation under challenging conditions, including repetitive background movement.

### 3.4. Determination of initial mean and variance

In Eq(5) and Eq(6), we assume that each model of fDC and fAC is a single non stationary Gaussian random process with time variant means and variances. We'll train first  $N=15\sim 20$  frames to calculate fDC and fAC of each block. These parameters of blocks which are in the same position of frames can be connected as a set of sample either fDC or fAC, like XDC and XAC in Eq(7,8,9,10).

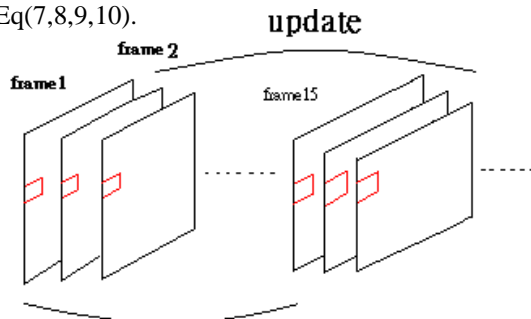


Fig 3 parameter training and update

$$\mathbf{m}_{0,DC} = MED(X_{DC}) \dots\dots\dots \text{Eq(7)}$$

$$\mathbf{S}_{DC} = 1.5 \times \left(1 + \frac{5}{(N-1)}\right) \cdot MED(ABS(X_{DC} - MED(X_{DC}))) \dots\dots\dots \text{Eq(8)}$$

$$\mathbf{m}_{0,AC} = MED(X_{AC}) \dots\dots\dots \text{Eq(9)}$$

$$\mathbf{S}_{AC} = 1.5 \times \left(1 + \frac{5}{(N-1)}\right) \cdot MED(ABS(X_{AC} - MED(X_{AC}))) \dots\dots\dots \text{Eq(10)}$$

Where MED donated the median operation and ABS means the absolute value. After determining the initial mean and variance, we will move to the next section to update each mean and variance which are time variant as non stationary Gaussian distribution.

### 3.5. Update background model

Before updating these parameters of each block, we assume that the DC&AC means and variances of the block which is at frame  $t$  are updated only if the block at frame  $t-1$  is labeled as background. Since we haven't described the method of segmentation, we assume that all blocks at zero-th frame are labeled as background in order to update all parameters at 1st frame. After initial segmentation, we'll update parameters of background model for the next frame  $t$ , where  $t > 1$ .

$$\mathbf{m}_{t,DC} = (1 - \mathbf{a}_{m,DC}) \cdot \mathbf{m}_{t-1,DC} + \mathbf{a}_{m,DC} \cdot f_{DC}(t) \dots\dots \text{Eq(11)}$$

$$\mathbf{m}_{t,AC} = (1 - \mathbf{a}_{m,AC}) \cdot \mathbf{m}_{t-1,AC} + \mathbf{a}_{m,AC} \cdot f_{AC}(t) \dots\dots \text{Eq(12)}$$

$$\mathbf{S}_{t,DC}^2 = (1 - \mathbf{a}_{s,DC}) \cdot \mathbf{S}_{t-1,DC}^2 + \mathbf{a}_{s,DC} \cdot (f_{DC}(t) - \mathbf{m}_{t,DC})^2 \dots\dots\dots \text{Eq(13)}$$

$$\mathbf{S}_{t,AC}^2 = (1 - \mathbf{a}_{s,AC}) \cdot \mathbf{S}_{t-1,AC}^2 + \mathbf{a}_{s,AC} \cdot (f_{AC}(t) - \mathbf{m}_{t,AC})^2 \dots\dots\dots \text{Eq(14)}$$

The DC (or AC) mean of the block at  $t$ -th frame has a portion of the DC (or AC) mean of the block at  $t-1$ -th frame and a portion of fDC (or fAC) of the block at current frame. The update calculations of DC and AC variances have the same meaning as that of mean.

Where  $\mathbf{a}_m$  and  $\mathbf{a}_s$  are called as learning parameter which control the learning update speed and we will discuss these learning parameter in the section of 3.8.

### 3.6. Segmentation

After determining these four parameters:  $m_{DC}$ ,  $m_{AC}$ ,  $S^2_{DC}$  and  $S^2_{AC}$  for all blocks of a frame, initial segmentation can be accomplished by computing the thresholds and classifying a block as foreground or background. The block at t-th frame is label as foreground if it satisfied both Eq(15) and Eq(16). Where TDC and TAC are thresholds and ABS means absolute value.

$$ABS(f_{DC}(t) - m_{t,DC}) > T_{DC} \quad \text{Eq(15)}$$

and

$$ABS(f_{AC}(t) - m_{t,AC}) > T_{AC} \quad \text{Eq(16)}$$

Eq(15) is the detection of great illumination change between the current block and the last block. There is a large difference in the DC value of the intensity between a moving foreground block and the background block, since  $f_{DC}(t)$  means the average value of block in gray level intensity and  $m_{t,DC}$  has a portion of last block's  $f_{DC}(t-1)$ . WE call those blocks satisfying Eq(15) "DC foreground blocks".

The use of Eq(16) is because the presence of the edge of the foreground objects and the different texture between the foreground and the background will lead to a large difference between the AC values of the current block and those of the background model block. The texture difference with small illumination change may occurs in shadow, reflection of mirror-like wall and small vibration of background movement, since  $m_{t,AC}$  also has a portion of last block's  $f_{AC}(t-1)$  if the last block is labeled as background. WE call those blocks satisfying Eq(16) "AC foreground blocks".

The use of these two threshold operation is equivalent to use both intensity and texture information, and will likely produce more vigorous and reliable segmentation results. The block at t-th frame is label as foreground if it is both DC block and AC block and we will move to next section to calculation the threshold of DC and AC value.

### 3.7. Threshold calculation

To handle the effect of illumination and texture change, we use for the threshold a value derived from both the sample mean and the sample standard deviation. Eq(17) is the threshold of DC value to detect a large illumination change and Eq(18) is the threshold of AC value to determine small illumination variation with texture difference between blocks.

$$T_{DC} = d_{DC} \times m_{t,DC} + g_{DC} \times S_{t,DC} \quad \text{Eq(17)}$$

$$T_{AC} = d_{AC} \times m_{t,AC} + g_{AC} \times S_{t,AC} \quad \text{Eq(18)}$$

Where,  $d_{AC}$ ,  $g_{DC}$  and  $g_{AC}$  are parameters, whose value depends on how much lighting change is expected to be handled here. The values of  $d$  and  $g$  are sensitive to the speed of illumination change and velocity of object movement.

$$0.5 < d < 1.5 \text{ and } 0.8 \leq g \leq 3$$

The above equation is only the estimation values of  $d$  and  $g$ . When training is not allowed or when re-training for variance is taken at the time of severe environment change, i.e. when segmentation should be done without an estimate of the variance, we simple raise the value of  $d$  and  $g$ .

### 3.6. Learning parameters

As mention at section 3.5, both these tow equation below:

$$m_{t,DC} = (1 - a_{m,DC}) \cdot m_{t-1,DC} + a_{m,DC} \cdot f_{DC}(t) \quad \text{Eq(11)}$$

$$m_{t,AC} = (1 - a_{m,AC}) \cdot m_{t-1,AC} + a_{m,AC} \cdot f_{AC}(t) \quad \text{Eq(12)}$$

Those above equations have  $a_m$  which is called learning parameter. There is a tradeoff between stability and adaptation speed when choosing a learning parameter. When the scene is slowly varying, a small  $a_m$  is preferred as it avoids the impact of outliers. When there is a change of either the camera or the background environment, a large  $a_m$  is preferred in order to quickly arrive at a new background model. For the parameter  $a_m$  that controls mean value adaptation, we set a range  $[a_{\max}, a_{\min}]$  for it and use the scheme as below:

$$a_m = \begin{cases} 0 & \text{foreground block} \\ a_{\min} & \text{background block} \end{cases} \quad \text{Eq(19)}$$

$$a_{\max} \quad \text{fast update}$$

$$0 \leq a_m \leq 1$$

Apparently, when there is a foreground block at (t-1)th frame, we won't update the DC and AC parameters of the block at t-th frame, so  $a_m$  is zero when foreground block is applied. If there is a background block at (t-1)th frame, we will assign  $a_m$  as  $a_{\min}$  whose value depends on the overall situation of the video sequence. When fast update occurs, we assign  $a_m$  as  $a_{\max}$  which has grater value than  $a_{\min}$ .

Fast update occurs when there is a large illumination change or hasty moving foreground object is applied in the video sequence. The blocks which are in the position at different frames change so quickly



that we need larger learning parameter  $a_m$  to keep the DC and AC means up with the speedy illumination change or fast moving object.

For the parameter  $a_s$  that controls variance update in Eq(13) and Eq(14),

$$S_{t,DC}^2 = (1 - a_{s,DC}) \cdot S_{t-1,DC}^2 + a_{s,DC} (f_{DC}(t) - m_{t,DC})^2 \quad \dots \quad \text{Eq(13)}$$

$$S_{t,DC}^2 = (1 - a_{s,DC}) \cdot S_{t-1,DC}^2 + a_{s,DC} (f_{DC}(t) - m_{t,DC})^2 \quad \dots \quad \text{Eq(14)}$$

At the time of severe change of the environment, it is not fair to update the variance as in the above equations, for the distribution is in the fact quite different from the original one before change occurs, and the mean value used in the variance update equation may be very distinct from the real value due to the delay of its own adaptation, so  $a_s$  does not take effect when fate update is applied to  $a_m$ .

### 3.9. Merge

After segmentation discussed above, there could be a few blocks which should belong to foreground blocks but counted as the background blocks shown in Fig 4 and we call these blocks: Undetermined block. Those undetermined blocks may occur due to those reasons:

1. The inaccuracy of the update mean and variance which vary with the illumination changes and the estimation of mean and variance is incorrect
2. Moving object which is too large or in slow motion results in similar feature and small distinct between blocks of frame  $t$  and  $t-1$ .
3. Object with similar background raises the difficulties of segmentation.
4. When great background movement occurs, we need larger thresholds both  $T_{DC}$  and  $T_{AC}$  to segment the foreground object. These thresholds may cause the damage of integrity after segmentation.

In order to take this problem in account, we present methods in section 3.9 and 3.10 to repair the result of segmentation.

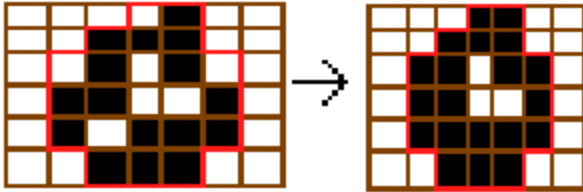


Fig 4. Block condition after segmentation and merge.

In Fig 4, red line represents the margin of the moving object and black blocks represent foreground label. We compute four features to describe DCT components in different frequency. The DCT coefficients have several different spatial features since

DCT coefficients provide the frequency domain information in the block. Eq(20,21,22,23) show the high frequency DCT coefficients.

$$f_0 = |DCT(0,0)| \quad \dots \quad \text{Eq(20)}$$

$$f_1 = \sqrt{\sum_{j=1}^{N-1} |DCT(0,j)|^2} \quad \dots \quad \text{Eq(21)}$$

$$f_2 = \sqrt{\sum_{i=1}^{N-1} |DCT(i,0)|^2} \quad \dots \quad \text{Eq(22)}$$

$$f_3 = \sqrt{\sum_{i=1}^{N-1} \sum_{j=1}^{N-1} |DCT(i,j)|^2} \quad \dots \quad \text{Eq(23)}$$

$$f_m^{spatial} = \sum_{i=0}^3 c_i \times f_i \quad \dots \quad \text{Eq(24)}$$

In Eq(21),  $f_1$  describes the horizontal high frequency.  $f_2$  describes the vertical high frequency and  $f_3$  describes the mixed horizontal and vertical high frequencies for  $1 \leq (i, j) \leq N-1$ . In Eq(24), we sum these features by multiplying a weighting vector  $\vec{C} = (c_0, c_1, c_2, c_3)$  to calculate  $f_m^{spatial}$  at the block  $m$ . The vector is usually set to  $\vec{C} = (1,1,1,1)$ .

The proposed spatial merge method checks if the spatial feature vectors of two neighboring blocks are similar according to Eq(25). If they are similar, they are merged into a region and the region growing of current region proceeds until there is no similar block.

$$|f_m^{spatial} - f_n^{spatial}| \leq T^{spatial} \quad \dots \quad \text{Eq(25)}$$

where  $n \in N(m) = \text{Neighborhood of block } m$

When the luminance value ranges between 0 and 255, the luminance difference smaller than 4~5 can be considered as similar block and the difference between similar blocks in DCT domain usually can be found in horizontal or vertical frequency. Since the spatial feature vector in Eq(24) has four feature components in DCT 8x8 block, the threshold in Eq(25) is set to  $T^{spatial} = 15 \times N$ ,  $N=8$ . As a result, several homogeneous blocks are obtained. There is still another method in section 3.10 to refine the result of segmentation.

### 3.10. Size filtering

The detection above those sections can find out the great part of foreground moving object's edge. However, when the moving object is too large or foreground with slow motion, it may occur that several blocks can not be detected as foreground block, especially the interior blocks. The interior block at frame  $t$  is too similar to the block at frame  $t-1$  in DC and AC components due to the

large moving object with slow motion and tardy illumination change.

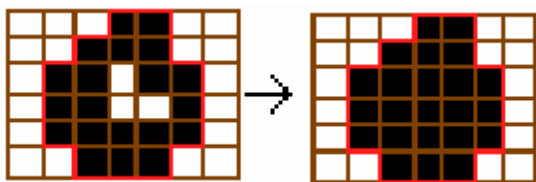


Fig 5. Size filtering

We use 8 neighbor connected component to refine the result of the segmentation. If a small background region is surrounded by foreground blocks, we simply fill those background blocks to become foreground blocks.

#### 4. EXPERIMENT RESULT

There are three typical experiments applied in DCT method in foreground segmentation:

1. A man walked with highly reflective wall in the background and persisting global illumination change.
2. Two persons cross paths at the entrance of a store.
3. A man walked with cluttered background of trees and swaying branches on a windy day.

These three representative works are different types from one another. Each video sample has particular circumstance for foreground segmentation on DCT domain and the source of every video sample comes from the video database online or is made by myself. We'll discuss each of them at the sections below.

##### 4.1. Demo 1

The first sequence is a feature of shadow in highly reflective wall in the background. A man walked with highly reflective wall in the background and persisting global illumination change. The mirror like background shows a considerable number of moving shadow from objects at a distance from the wall. In addition, the background contains a large area looking much like the skin tone in intensity image. The method of segmentation in DCT domain result in very few false negatives and almost no false positives. Several sample frames are shown in the sets of Fig 6.



Fig 6-1. Frame 1st, background of DEMO 1. The right region is input video sample and the left region is the result.



Shadow Object  
Fig 6-2. Frame 20th.



Fig 6-3. Frame 27th.



Fig 6-4. Frame 38th.

As can be seen from the result, the foreground segmentation is almost perfect with the background moving shadow and the lighting condition kept changing while the foreground object kept moving forward. If the thresholds are raised higher, there will be more false negatives; if they are lower, a lot of false positives. While the simultaneous modeling of block based features leads to satisfactory result, with all the parameters kept unchanged, without any adaptation.

However, there are still several drawbacks of this DCT method. In Fig 6-4, it is clear to observe that the result has rough edge in the segmentation because calculation is based on 8x8-pixels block. The DCT method has more rugged edge than traditional pixel method and the solution may be reducing the block size to be 6x6 or 4x4.

##### 4.2. Demo 2

The second video sequence which comes from the video database online at "<http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>" presented that two persons cross paths at the entrance of a store. The main idea of this sequence is to observe the adaptation speed of learning parameter update





Fig 7-1. Initial frame of video sequence, the right region is input video sample and the left region is the result.



Fig 7-2. Frame 4th. Two persons began to cross on the path. The DCT method only detected the movement of the left person.



Fig7-3. Frame 14th. Two persons began to separate.



Fig7-4. Frame 27th.

The result of segmentation is not good enough at the beginning of this video sequence, only two persons' legs were detected. After updating several frames, the result became better and better. This is because that the adaptation of mean and variance parameters can keep up with the variation of illumination change with proper values of learning parameters( $a_m$  and  $a_s$ ) as long as the foreground object keeps moving.

Nevertheless, there is another shortcoming of this DCT method. In Fig7-3, 7-4 and 7-5, only two persons' legs were detected, because the two walking persons wearing clothes with similar color as that of the background. After several frames, in Fig8-6, the result showed the left walking person's whole body but still can not show the upper body of the right person. In Fig8-7,8-8 and 8-9, the upper body of the right person was detected until he walked through the background which has much different color with the clothes.

### 4.3. Demo 3

The third video sequence contains cluttered background of trees, swaying branches and a person who wore clothes with the color similar to the left background. At the beginning, computer can not detect the block of moving object due to the too much similar color of the walking person to the left background. However, after several frames updating, computer can easily detected the movement of the walking people without these blocks of swaying branches.



Fig 8-1. Frame 2nd. A person wore clothes with the color similar to the left background



Fig 8-2. Frame 15th. A walking person with swaying branches on the top of the picture.



Fig 8-3. Frame 24th. After several frame updating, the computer can easily detect the block of the walking without the block of swaying branches.



Fig 8-4. Frame 32rd

It can be seen clearly that this DCT segmentation can also handle the tiny vibration such as swaying branches. Because the weight ( $i^2 + j^2$ ) de-emphasizes the very low frequency components which often correspond to smooth transition of intensity such as shadow, reflection of mirror-like wall and small vibration of background movement presented in DEMO 3.



## 5. CONCLUSION AND FUTURE WORK

The method of segmentation in DCT domain for video sequence presents a simple way to separate the moving foreground object from solid background. It uses two features derived from the DCT domain and has low computation because the calculation is based on 8x8 pixels block so that it can be computed almost real time. A single Gaussian distribution is used to model the evolution of the two features in each block. An updating process is designed to handle effectively the changing background. The method is also able to successfully handle the problems of gradual global and local illumination change, moved background objects and small vibration background movement.

The drawbacks of DCT method are undeniable. According to the result, it has rougher edge than pixel method due to the block based computation. This may be overcome by using 4x4 or 6x6 pixels-block to get smoother margin. There is another disadvantage to be discovered that when the color of the foreground moving object is very similar to that of background, it is hard to segment the foreground from the video sequence.

This method may be used on the security monitor or robot vision. As in DEMO 2, this video sequence actually come from the security monitor record video sequence of a department store. It also has utility on robot vision in order to extract the movement of foreground object.

## 6. REFERENCES

- [1] Jiashu Zhang, Like Zhang and Heng Ming Tai, "Efficient Video Object Segmentation using Adaptive Background Registration and Edge Based Change Detection Techniques" Volume 2, 27-30 June 2004 Page(s):1467 - 1470 Vol.2 IEEE Trans
- [2] C. Kim and J Hwang, "Video object extraction for object oriented applications", Journal of VLSI signal processing, vol29, pp.7-21, 2001
- [3] C Ridder, O Munkelt et al, "Adaptive Background estimation and foreground detection using kalman filtering", Proc of Intl Conf On Recent Advances in Mechatronics(ICRAM), pp193-199, 1995.
- [4] Jesus Bescos, Jose M Menendez and Narciso Garcia "DCT based segmentation applied to a scalable zenithal people counter" IEEE 2003
- [5] Mathieu Lamarre, James J Clark, "Background subtraction using competing models in the block DCT domain" Volume 1, 11-15 Aug. 2002 Page(s):299 - 302 vol.1 Digital Object Identifier 10.1109/ICPR.2002.1044695 IEEE CNF
- [6] L, Li and M Lenug, "Integrating intensity and texture differences for robust change detection" IEEE Trans. On image processing, Vol 11, No2, pp105-112 2002
- [7] O. Javed, K Shafique and M Shah, "A Hierarchical Approach to robust background subtraction using color and gradient information", Proc. Workshop on Motion and Video Computing, 2002, pp22-27
- [8] Jie Wei, "Image segmentation using situational DCT descriptors", Volume 1, 7-10 Oct. 2001 Page(s):738 - 741 vol.1 Digital Object Identifier 10.1109/ICIP.2001.959151 IEEE CNF
- [9] Rahul Shankar Mathur, Punit Chandra and Dr. Sumana Gupta, "A low bit rate video coder using content based automatic segmentation and shape adaptive DCT" Volume 2, 19-22 Jan. 2000 Page(s):119 - 122 vol.1 IEEE CNF
- [10] Salkmann Ji, Hyun Wook Park, "Region based Video segmentation using DCT coefficients" Volume 2, 24-28 Oct. 1999 Page(s):150 - 154 vol.2 Digital Object Identifier 10.1109/ICIP.1999.822873 IEEE Trans 1999
- [11] Salkmann Ji, Hyun Wook Park, "Moving-object Segmentation with adaptive sprite for DCT- based video coder" Volume 3, 7-10 Oct. 2001 Page(s):566 - 569 vol.3 Digital Object Identifier 10.1109/ICIP.2001.958177 IEEE Trans 2001
- [12] Juhua Zhu, Stuart C. Schwartz, Bede Liu, "A transform Domain Approach to Real-Time Foreground Segmentation in Video Sequence" Volume 2, March 18-23, 2005 Page(s):685 - 688 IEEE Trans 2005.

## 出席國際學術會議心得報告

計畫編號	NSC 95-2221-E-002-188
計畫名稱	多媒體影音高階處理、傳輸及設計--總計畫:多媒體影音高階處理、傳輸及設計(2/3)
出國人員姓名 服務機關及職稱	貝蘇章 國立台灣大學電機系 教授
會議時間地點	95/10/06~95/10/12 美國亞特蘭大市
會議名稱	( 中文 )2006 國際影像處理會議 ( 英文 )2006 IEEE Int'l Conference on Image Processing
發表論文題目	( 中文 ) 權重式統計分離之彩色影像處理法 ( 英文 ) Color image enhancement using weighted histogram separation. (ICIP)

一、參加會議經過

二、與會心得

## 報告內容:

- 一、 出席 2006 國際影像處理會議
- 二、 參加會議經過及與會心得
- 三、 大會的重點及發展趨勢
- 四、 攜回資料名稱
- 五、 結論

## 出席國際會議報告

『2006 國際影像處理會議』(2006 IEEE Int'l Conference on Image Processing)假美國亞特蘭大市盛大舉行，筆者應邀出席該會，發表論文一篇，茲將參加會議的心得分述於下：

- 一、本會共有一仟伍佰多位來自世界各地五十多個國家的學者、專家與會，約有 1200 篇論文在大會發表，內容著重在訊號處理、數位影像處理，影像壓縮與視訊、及多媒體等是個內容非常豐富及具學術權威地位的重要會議。
- 二、此次大會的主辦單位為美國國際電機電子工程師學會、訊號處理分會協辦(IEEE Signal Processing Society)，其目的想藉此會交換各國最新「影像處理」科技，並促進學術交流。
- 三、亞特蘭大的會議地點為城市旅館國際會議中心設備優良，交通方便，是個理想的會議場所。
- 四、兩個會議的頭一天均有短期課程班(Tutorials)，讓與會者有再接受新科技之機會。大會的學術論文發表方式分成 12 個場地同時進行，另外有個較大的場地以海報方式舉行，並有書展及影像系統及產品展示，海報方式能與原作者親自相互討論及交換心得，效果十分良好。

- 五、大會在開幕式邀請三位專家，發表學術專題演講，內容精采，受益良多。
- 六、大會並在星期二晚上舉行晚宴，並安排音樂演奏及歌唱舞蹈表演，以宴與會嘉賓，同時進行 IEEE 訊號處理學會的最佳論文頒獎。
- 七、本人此次承蒙「國科會專題計劃」補助旅費，得以順利成行，在此深表謝意，並攜回大會論文集之光碟一片，引進及吸收不少新知及科技發展，供國內參考，並達到學術交流的目的。

台大電機系教授

貝蘇章 敬上 95/10/20