Research on Video Object Segmentation Algorithm

Guan Bo

Baicheng Normal College 137000

Keywords: video surveillance system; object segmentation; decoding

Abstract. The video surveillance system is a more direct monitoring means, widely used in industrial production, transportation, electricity, office buildings and residential areas. Plenty of remote monitoring systems adopt the mode to collect and compress data through local end and remote monitoring center to make decompression processing. This paper analyzed the urban traffic surveillance video streaming, and it regarded vehicles as search targets, which have characteristics of rapid movement, therefore a particular object segmentation method was put forward. The algorithm took advantage of object motion vector in the compressed domain, as well as proposed the new regional connectivity labeling algorithms.

Algorithm introduction

The algorithm will be first implemented in a video stream decoding playback software, and the motion detection module is embedded in the video stream decoding, then the intercepted macro block motion vector can be as input information for later segmentation, and these data after the rapid processing is still left to subsequent decoding steps, as follows: the control information obtained multiplexed solutions according to the levels, and video streaming is gathered in the input buffer. When reaching certain fullness, a decoder starts. According to the level of video sequence, group of pictures, images, slices, macro blocks and blocks, the decoder begins decoding in a top-down sequence, and each layer has a corresponding start tag and control information. The decoder starts to mark the sequence header code after finding the first sequence header, followed by bits, which are the image width (12bits), height (12bits), high aspect ratio (4bits), frame rate (frames/sec, 4bits), image transmission bit rate (bit/s, 1 8bits), decoding buffer large tJ, (10bits) and other initialization information, and they have predetermined number of bits and syntax descriptions, so downward to the block layer with multiplexed solutions. Variable length decoding VLD, carries on statistics for the macro block addressing, macro block type, macro block pattern, macro block motion vectors, block DCT coefficient, and formulates a corresponding variable length coding statistical table, and both sides of encoding and decoding have this statistical table. When a string of binary bit stream is received, we do not know in advance how many parts it can be divided into, and how many the number of bits for each part is. Under the condition of relatively few cases entry, referring to the relevant statistics with access to traversal methods can quickly decode, but there are more entries in some statistical table, for example the variable-length coding statistical table of block DCT coefficients has hundreds of items. Followed by the Z-positioning and inverse quantization, this step is performed on the block level. The quantized DCT coefficients received in the order will be placed in each block according to the "Z" type order, multiplied by the quantization scale at the same time, so as to complete inverse quantization. In the same frame image, quantization scale is invariant.

Inverse discrete cosine transform IDCT is to achieve the transform on the block layer from frequency domain to spatial domain. The calculation amount of inverse discrete cosine transform IDCT (inverse discrete cosine transformation) is large, which is the most time consuming link in decoding. In the motion cache and compensation, motion compensation is the reverse process of motion estimation, and P-frame is forward motion compensation with reference to the past time image; B-frame is interpolation compensation with reference to past and future images, but also allows the unidirectional reference, which are respectively forward motion compensation and

backward motion compensation.

1) This algorithm in the paper only adopts I and P frames to segment and retrieve the video object, so that each part of algorithm above proceeds the B-frame skipping operation. First two storage areas were opened up, respectively retaining decoded past or future I, P picture as references, and find predicted macro block value in the reference images by means of macro block motion vector, then restore the current macro block initial value with the addition of matching error that may exist. In this step, decoded macro block motion vector will be cached for segmentation. Concrete realization function is as follows:

void dec：cal_mv(int mvx，int mvy,int bx，int by)

(inti，J，width；

i=int(by／16)；／／Macro block number in horizontal direction

j=int(bx／1 6)；／／Macro block number in vertical direction

mv_src[i][j]=float(mvx)；／／Macro block motion vector cache in horizontal direction

mv_srcl[i][j]=float(mvy)／／Macro block motion vector cache in vertical direction

When all the information is available, we start the region segmentation of moving object on the decoded video sequence, and further carry on the object retrieval. In the region segmentation of moving object, we apply the background recovery and motion regional connectivity algorithm, and use a template matching algorithm in the search algorithm. Finally, we transform the coding sequence to the display order and display output.

Background recovery

When performing object segmentation, we first make preprocessing for video sequence, namely background recovery. Background recovery algorithm first guarantee two adjacent frames that may be possible object motion are filtered to obtain the static part, and reliable and static part will be copied to the background cache. The algorithm is as follows: YUV sequence that the decoding outputs is processed, first converted into RGB sequences, respectively seeking the frame difference of R, G, B for the two adjacent frames, and adaptive threshold is adopted for binarization. If the value is "1" indicates that the pixel has changed, if the value is "0" explains that the point has not changed. The pixel frame differences without changes satisfy the normal distribution:

$$p(FD|H_0) = \frac{1}{\sigma\sqrt{2\pi}} EXP(\frac{-FD^2}{2\sigma^2})$$

Among them, FD is frame difference, σ^2 is variance of frame difference, H0 is the hypothetical situation that the current pixel has no movement. Binarization threshold is determined by the type:

$$a = p(|FD| > THH_{0]})$$

Wherein a is significance level, TH is the binarization threshold. Since R, G, B space is adopted for background detection, if any component of the pixel including gray scale and color scale changes, then the pixel is considered to be the movement pixel, and the advantage of this detection algorithm is to remove movement pixel and reserve the static background pixel to the hilt. In the three frame difference components, if component value of a frame difference is "1", the pixel of corresponding point is considered to change. After calculating the stationary and moving points by means of the above method, it can count for the pixel of every point with the same frames in a row. If the count exceeds a certain value, the point is regarded as the background frame, and pixel values average of all the continuously constant frames in a row is considered to the background frame. After gaining the background frame, a moving object area subtracting from each decoded frame and the background frame is set as the input of the subsequent algorithms.

Motion regional connectivity algorithm

Generally, the rigid targets in space position are gathered together. Thus after obtaining I, P frames Mv, in which the connected macro blocks are combined into a area. This paper proposed a new connectivity algorithm for macro block merging. Specific approach is first to mark the macro block and then merge. The macro blocks are scanned in the order from top to bottom, left to right, and for any macro block A judged as different background, the macro blocks on the upper and left side have been already scanned and labeled macro block. The marking method of A will be determined the macro blocks on the upper and left side, divided into three cases: the current macro block is added a new sign when two adjacent macro blocks belong to the background; When only one of two macro blocks is labeled different background macro block, it and A are marked; When two macro blocks are all well marked and different background macro blocks, A and the left are labeled together. The following is a portion of implementation code for connected region labeling algorithm.

```
void mpegdec：：cal_sim(int mvx，im mvy,int bx，int by)
{
iIlt flag=l，i，dcsum=O，dcflag=1；
mvb *mvobj,*pointer,*pointerZ,*pointerl,*pointer2,*pointerj,*pointer3,*pxmaxl;
mvb *pyma xl,*pxmax2,*pymax2,*pxminl,*pyminl,*pxmin2,*pymin2;
double    angll；
int cndx，cndy,j，k，obj_num 1；
mv_obj=(mvb_obj)malloc(sizeof(mvb))；
pointer=pointerl=pointer2=pointer3=pointerj=pxmaxl=pymaxl=pxmax2=pymax2
=pxmin 1=pymin 1=pxmin2=pymin2=NULL；
mv_obj->mvx=mvx；
mv_obj->mvy=mvy；
mv_obj ⟶>by=by；
mv__obj->xmin=bx；
mv__obj->ymin2by；
mv_obj->xmax5bx；
mv_obj->ymax=by；
mv_obj->angle=0：
mv_obj->next=NULL；
mv_obj·>num_mb=0：
if((obj_num)==0)
{
mv_obj->xmax=bx；
mv_obj->ymax=by；
mv_obj->xmin=bx；
mv_obj->ymin=by；
mv_obj->num_mb=l；
obj_p[0]=mv_obj；
(obj_pum)++；
}
```

The next is regional connectivity algorithm for motion and non-background blocks, the detected backgrounds are subtracted from each frame, After the foreground region is determined according to the threshold, it can amend the connected region of the above movement with these foreground blocks to get the moving object area. A union set of moving object area subtracting from the above-mentioned moving object area and the background frame is regarded as the extraction area of

the edge of moving object. Due to the influence of noise and the algorithm itself based on a 4 * 4 block (in order to speed up the computation), the above algorithm can't be well corrected to the pixel segmentation in the border area of the moving target, providing a accurate rectangle of moving objects for the further object retrieval. We only make edge extraction for images in the rectangular box, and the algorithm used is CANNY edge extraction algorithm, then the edges of extracted objects are set as input of target tracking and object retrieval in the next step.

References

[1] Hotter　M.，Thoma R.，Image segmentation based on object oriented mapping parameter estimation，Signal Processing，15(3)：315-334，2008

[2] Musmann H．G．H6tter M．and 0stermarm J．．Object．oriented analysis-synthesis coding of moving images，Signal Processing：Image Communication，1(2)：1 17-138，2009

[3] Murray D．W．，Buxton B．E，Scene segmentation from visual motion using global optimization，IEEE Tram．Pattern Analysis and Machine Intelligence，PAMI-9(2)：220-228，1997

[4] Bouthemy P．,Francois E．．Motion segmentation and qualitative dynamic scene analysis from all image sequence，hat．Journal of Computer Vision，1 0(2)：1 57-1 82，1993