An Improved Method for Detection of Moving Objects in Image Sequences Using Statistical Hypothesis Tests

Jae Gark Choi, Munchurl Kim, Myoung Ho Lee, and Chieteuk Ahn

Electronics and Telecommunications Research Institute
161 Kajong-Dong, Yusong-Gu, Taejon, 305-350, Korea
E-mail: cjg@video.etri.re.kr

ABSTRACT

This paper presents a spatio-temporal video segmentation method. The algorithm segments each frame of video sequences captured by a static or moving camera into moving objects (foreground) and background using a statistical hypothesis test. In the proposed method, three consecutive image frames are exploited and a hypothesis testing is performed by comparing two means from two consecutive difference images, which results in a *T*-test. This hypothesis test yields change detection mask that indicates moving areas (foreground) and non-moving areas (background). Moreover, an effective method for extracting object mask from change detection mask is proposed.

1. INTRODUCTION

Image segmentation techniques are important tools for content-based image coding, manipulation of image contents, and interactive multimedia applications, etc. Segmentation of image usually divides the image contents into semantic regions that can be dealt as objects. These semantically segmented objects can be coded so that object-based manipulation of image content can be possible in interactive multimedia applications [2,3]. This finds an example, MPEG-4 being currently standardized which aims providing core techniques of object based manipulation of audio-visual data in coding frameworks [1].

For intensity change detection, intensity difference between successive image frames has been used as a cue in the findings of moving objects in image sequences through time evolution [2,3,4,6]. This is because a moving object in an image sequence usually entails intensity changes in magnitude, depending upon its locations. A local small window slides over the difference image. The sum of squared values of difference image pixel or sum of absolute difference pixel values in the local window is compared with a heuristically defined threshold [2,3]. Aach et all used a test statistic that is the variance estimate divided by the true variance. This test statistic has a χ^2 distribution and a threshold can be theoretically determined for a given level of significance [4]. However the true variance must be *a priori* known in order to compute the test statistic that is supposed to be compared with the threshold. In practice, the true variance is usually unknown and is dependent on particular types of image.

This paper proposes a detection method of intensity changes in image sequences using a statistical hypothesis test. The proposed method exploits three consecutive image frames for moving object segmentation and performs a hypothesis test made on means in observation windows of difference image frames. The test statistic used in the hypothesis test approximately follows a *t*-distribution and does not necessarily require the true variances *a priori*. Also, an effective method for extracting exact object mask from change detection mask is proposed by using spatial information.

In this paper, Section 2 describes the proposed segmentation algorithm. The simulation results are addressed in Section 4 and Section 5 concludes this paper.

2. THE SEGMENTATION ALGORITHM

In this section, we describe our spatio-temporal segmentation algorithm. It is based on statistical hypothesis test from two intensity difference frames. The block diagram is illustrates in Fig. 1. Each process is describes in the following subsections.

2.1 Global motion estimation and compensation

For the extraction of moving objects in the current frame F_k, intensity change is considered in between two consecutive frame pairs: the first frame pair of F_{k-1} and F_k; the second of F_k and F_{k+1}. Before considering the intensity change, we first perform global motion estimation on the frame pairs, and make decision as to whether global motion of a camera exists in the frame pairs. If there exists the camera motion between F_{k-1} and F_k, we predict from the frame F_{k-1} a globally motion-compensated frame F_{k-1}^{gmc} in the forward direction. Similarly, we also predict from F_{k+1} a globally motion-compensated frame F_{k+1}^{gmc} in the backward direction when the global motion exists between F_k and F_{k+1}. Now we will describe the global motion estimation (GME), GMC ON/OFF decision and GMC method in details.

We use the affine 6 parameter motion model for the global motion estimation (GME). First, we calculate the local motion vectors for each 16x16 rectangular block using BMA. Then, we estimate the 6 parameters using a linear regression method. In this process, we remove the outlier vectors and iterate the regression process until the estimates of 6 parameters converge. After the final parameters are calculated, GMC(global motion compensation) ON/OFF decision is performed. This is because the camera movement does not always exist in the sequence. We adopt the *coefficient of multiple determination* that is used for the model adequacy test in statistical analysis. If the value of *coefficient of multiple determination* is greater than the given threshold, the GMC is performed. Otherwise, in case of no global motion in the sequence, the GMC is not performed.

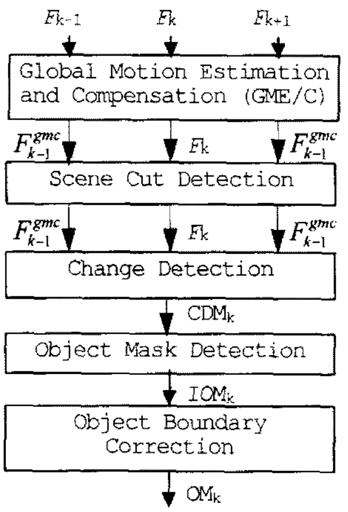


Fig. 1 Principle block diagram

2.2. Scene Change Detection

If there exists the scene chage between F_k and F_{k-1}^{gmc} or F_k and F_{k+1}^{gmc}, we reset the object mask(OM) to background. For scene change detection, mean absolute difference (MAD) of luminance components is considered between consecutive images. If the MAD is greater than a given threshold, it is considered that scene change has occurred.

2.3 Change detection based on mean comparison

In Fig. 2, let D_{k-1} be the difference image frame between F_{k-1}^{gmc} and F_k, and D_k the difference image frame between F_k and F_{k+1}^{gmc}. The difference images are usually modeled as having normal distributions [4]. For simplicity, the difference intensities X are assumed to be identically independent random variables. A set of random variables in the observation window W_1 is compared based on mean with that of the observation window W_2 at the same location. Therefore the null hypothesis made on mean is assumed that the two random samples observed in W_1 and W_2 were drawn from the same distribution, that is, the two population means μ_1 and μ_2 are the same. The hypothesis made on mean is written such that

$$H_0:\ \mu_1 = \mu_2$$
$$H_1:\ \mu_1 \neq \mu_2$$

In order to perform a hypothesis test on a parameter θ that is the mean μ, a test statistic needs to be developed which should includes the parameter assumed.

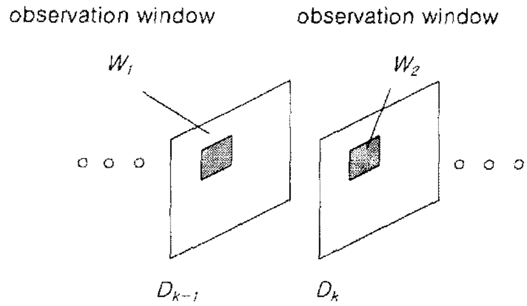


Fig. 2. Observation windows sliding over two consecutive difference image frames.

Comparing two means can make the decision as to whether a pixel under consideration belongs to foreground or background. The mean comparison can be done based on the difference between two mean estimates $\bar{X}_1$ and $\bar{X}_2$ in W_1 and W_2.

The random variable $\bar{X}_1 - \bar{X}_2$ is an unbiased estimator of $\mu_1 - \mu_2$ and has a normal distribution with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$.

The standardized form of the random variable $\overline{X}_1 - \overline{X}_2$ is written as

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \qquad (1)$$

and has a standard normal distribution. Under the null hypothesis, that is $\mu_1 = \mu_2$, the test statistic Z is reduced to $(\overline{X}_1 - \overline{X}_2)/\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}$. However the test statistic requires the two true variances σ_1^2 and σ_2^2 to be *a priori* known. In practical problems, the two true variances are usually unknown. Instead, σ_1^2 and σ_2^2 are replaced by their respective estimators $S_{n_1}^2$ and $S_{n_2}^2$ to obtain the random variable

$$U = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_1^2 / n_1 + S_2^2 / n_2}} \qquad (2)$$

Then this random variable U has an approximate t-distribution from the random variable Z having a standard normal distribution. However the number of degree of freedom must be estimated from the data in W_1 and W_2. According to the Smith-Satterthwaite [5], the number of degree of freedom df_U is given by

$$df_U \approx \frac{(S_1^2 / n_1 + S_2^2 / n_2)^2}{\frac{(S_1^2 / n_1)^2}{n_1 - 1} + \frac{(S_2^2 / n_2)^2}{n_2 - 1}} \qquad (3)$$

If the value for df_U is not an integer it is round down to the nearest integer. Under the null hypothesis for a given level of significance α, the lower and upper bounds, $-t_{\alpha/2}$ and $+t_{\alpha/2}$, for a confidence interval on $\mu_1 - \mu_2$ are obtain by

$$P[-t_{\alpha/2} < U < t_{\alpha/2}] = 1 - \alpha. \qquad (4)$$

where $t_{\alpha/2}$ is the value from which the integration of a t-distribution function to infinite is $\alpha/2$. The null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted if $U < -t_{\alpha/2}$, or $U > +t_{\alpha/2}$, otherwise the null hypothesis is accepted and the alternative hypothesis is rejected. That is, when the test statistic computed at a pixel under consideration is greater than $t_{\alpha/2}$ or less than $-t_{\alpha/2}$, the pixel is declared as foreground, otherwise background. The resulting foreground/background decision yields a change detection mask (CDM).

2.4 Moving Object Detection

The CDM consists of the moving object (MO) parts, uncovered background (UB) for the F_k, and background to be covered (BC) for the F_{k+1}, as shown in Fig. 3. We have to distinguish the MOs of the F_k from the CDM between the F_{k-1} and F_{k+1}.

We first separate the UB for the F_k from the CDM [2]. Then the MOs for the F_k are extracted by separating the BC for the F_{k+1} from the CDM. For the separation of the UB for the F_k, we first estimate the motion vector field describing the motion of moving objects from the F_k to F_{k-1} by using a differential method. We use the affine model for the motion description of moving objects. Then these displacement vectors connect the corresponding positions of MOs starting from the F_k within the changed region. All pels which are to be displaced from the changed region of the F_k to the unchanged region of F_{k-1} belong to the uncovered background (UB). The other pels within the changed region of the F_k belong to the MO. Similarly, we can extract the BC for the F_{k+1} from the CDM and then finally obtain an object mask (OM).

2.5 The object boundary correction

Finally, the object boundary is corrected by exploiting the spatial information. For this, we first form the uncertainty areas around the boundary of the IOM. We assume there are exact object boundaries within the uncertainty areas. The uncertainty areas are obtained by blowing and shrinking the IOM. The pels within a small predefined width from the boundaries of the IOM are marked as uncertainty pels. The determination of the width depends on the accuracy of the moving object detection. In the range of a relatively moderate motion, 3-pixel width will be sufficient. Then we exploit the watershed algorithm in order to assign the pels of the uncertainty areas into moving object or background. The initialization and flooding of the watershed algorithm merges the whole pels in the uncertainty areas based on the similarity between the intensity value of the considered pel and the average intensity value of the neighbor regions[7]. After the intensity-based boundary correction, some spark labels may appear along the edges in the OM. We apply the median filter over the OM. The result is the final object mask (OM_k), describing the shape of the moving objects in each image of the sequence.

3. SIMULATION RESULTS

In order to evaluate performance of the proposed segmentation method, MOTHER DAUGHTOR and HALL MONITOR image sequences with a QCIF size were used. In Fig. 4(a) – (c) and 5(a) – (c), two sets of three consecutive frames of the two sequences are shown. The test statistic proposed in Eq. (2) was computed and compared with thresholds for a given level of significance = 10^{-3} at each pixel location. A 7×7 size of an observation window was used for the computation of the test statistic. The resulting both CDMs for two image sets are shown in Fig. 4(d)and 5(d), respectively. Thanks to the object motions, it can be noticed that the intensity-changed regions that indicate the moving objects (foreground) are well distinguished from their still backgrounds. It is also noticed that some isolated spots also appear as foreground due to a large variation of intensity at some locations in Fig. 5(d). A median filter of a size of 5×5 was used to eliminate small spots in background and to fill in small regions inside the CDMs. Notice that in Fig. 5(d) there is an intensity-changed region in the right side of the right leg of the walking person in the hall. This was caused by the shadow of the person.

Due to object movements from F_{k-1} to F_k and from F_k to F_{k+1}, the resulting uncovered and covered regions coincide within the CDMs as shown in Fig. 4(e), 4(f), 5(e), and 5(f). After removal of those regions, the resulting IOMs are obtained and shown in Fig. 4 (g) and 5 (g). Fig. 4(h) and 5(h) indicate the final moving objects (OM), after object boundary was corrected by using spatial information.

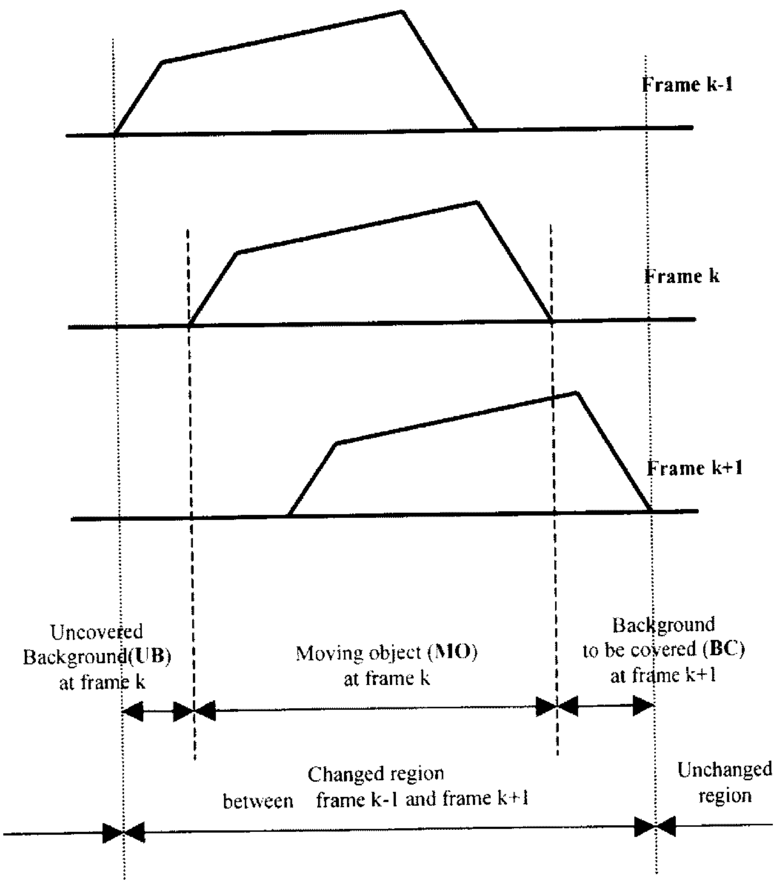


Fig. 3 The separation of changed region ino MO, UB and BC

4. CONCLUSIONS

This paper proposed an image segmentation method that utilizes statistical hypothesis testing and compared two test statistics based on segmentation performance. The test statistic following an approximate t-distribution can be used without *a priori* knowledge about the true variances, which is suitable to real world problems. As the sample size increases in the observation windows the test statistic asymptotically becomes having a normal distribution. In the case of frames containing moving objects, the test statistic showed a good capturing capability of intensity variation caused by object motion. Also the extraction and correction method of moving object from the change detection mask is proposed. It separates moving objects by utilizing the motion information and corrects the exact object boundary by utilizing the spatial information.

Acknowledgements

This work was supported by the Ministry of Information Communication of Korean Government.

References

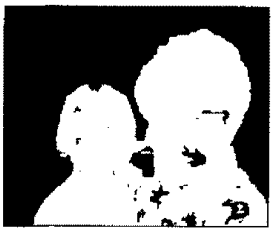
1. ISO/IEC JTC1/SC29/WG11, N1683, "Overview of the MPEG-4 standard," Sevilla, Spain, Feb., 1997.
2. Michael Hotter and Robert Thoma, "Image segmentation based on object oriented mapping parameter estimation," *Signal Processing*, vol. 15, pp. 315-334, 1988.
3. Hans G. Musmann, Michael Hotter and Jorn Ostermann, "Object-oriented analysis-synthesis coding of moving images," *Signal Processing: Image Communication*, vol. 1, pp. 117-138, 1989.
4. Til Aach and Andre Kaup, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, pp. 165-180, 1993.
5. J. S. Milton and J. C. Arnold, *Introduction to Probability and Statistics*, McGraw-Hill, New York, 1995.
6. Roland Mech and Michael Wollborn, "A noise robust method for segmentation of moving objects in video sequences," *International Conference on Acoustic, Speech and Signal*, Munich, Germany, April 1997.
7. J. G. Choi, S-. Lee, S-. Kim, "Spatio-temporal video segmentation using a joint similarity measure," *IEEE Trans. on Circuits and Systems for Video technology*, Vol.7, No. 2, pp. 279-286, April, 1997.

MOTHER_DAUGHTOR images

(a) frame #183

(b) frame #186

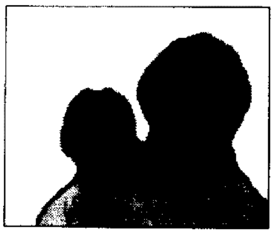
(c) frame #189



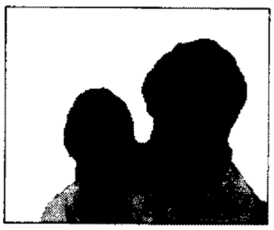
(d) CDM

(e) uncovered region

(f) covered region



(g) IOM (foreground)



(h) OM (foreground)

Fig. 4. Segmentation of moving objects in MOTHER_DAUGHTOR image sequences

HALL_MONITOR images

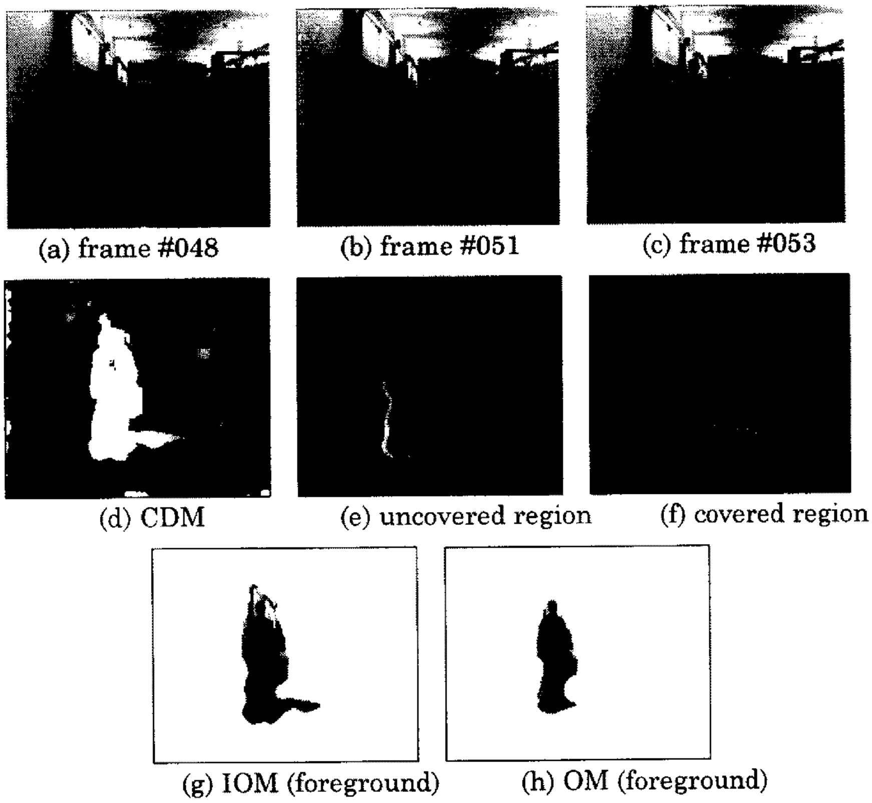


Fig. 5 Segmentation of moving objects in HALL_MONITOR image sequences