Integrating Intensity and Texture Differences for Robust Change Detection

Liyuan Li, Student Member, IEEE, and Maylor K. H. Leung, Member, IEEE

Abstract—We propose a novel technique for robust change detection based upon the integration of intensity and texture differences between two frames. A new accurate texture difference measure based on the relations between gradient vectors is proposed. The mathematical analysis shows that the measure is robust with respect to noise and illumination changes. Two ways to integrate the intensity and texture differences have been developed. The first combines the two measures adaptively according to the weightage of texture evidence, while the second does it optimally with additional constraint of smoothness. The parameters of the algorithm are selected automatically based on a statistic analysis. An algorithm is developed for fast implementation. The computational complexity analysis indicates that the proposed technique can run in real-time. The experiment results are evaluated both visually and quantitatively. They show that by exploiting both intensity and texture differences for change detection, one can obtain much better segmentation results than using the intensity or structure difference alone.

Index Terms—Change detection, difference pictures, integration, motion detection, segmentation, subtraction, video analysis.

I. INTRODUCTION

Finding moving objects in image sequences is one of the most important tasks in computer vision. Common applications for motion analysis include object tracking [1], security surveillance [2], [3], intelligent user interfaces [4], and traffic flow analysis [5]. In all these applications, the first fundamental problem encountered is motion segmentation which extracts moving objects from the scene. Most motion segmentation techniques for dynamic-scene analysis are based on the detection of changes in a frame sequence [6]. Change detection reduces the amount of data for further processing. Because of its importance in the preprocessing of dynamic scene analysis, many techniques have been developed for robust change detection.

The existing change detection techniques can be classified into two categories: pixel-based and region-based methods. The most intuitive technique to detect change is simple differencing followed by thresholding. Change at a pixel is detected if the difference in gray levels of the corresponding pixels in two images exceeds a preset threshold. More robust methods adaptively select the threshold based on the noise estimation of difference picture [7] or thresholds at each pixel based on the gray level distributions of the background points [4], [8]. Change detection at pixel level requires less computational cost since only one pixel is considered at a time. But it is very sensitive to noise and illumination changes since it does not exploit local structural information. Hence, it is not suitable for applications in complex environments. To make change detection more robust, intensity characteristics of regions at the same location in two images are compared using statistical approaches. A straightforward solution is hypothesis testing to check if statistics of the two corresponding regions come from the same intensity distribution. The likelihood ratio test taken by Nagel [9] is one example of such approach. The performance of the likelihood ratio test can be improved by using quadratic surfaces to approximate the intensity values of pixels belonging to the regions. Hsu et al. [10] proposed a second-order bivariate polynomial in the pixel coordinates to model the gray value variation in the regions. A difference metric is then defined to determine if changes occur in corresponding regions of two images. Approaches based on the comparison of intensity distributions at region level for change detection work well in noisy environments, but they are sensitive to illumination changes since they still only employ intensity difference measures. Recently, region-based illumination independent statistics have been proposed. Skistad and Jin [11] presented a Shading Model (SM) which uses the ratio of the intensities recorded in the corresponding regions of two images to detect changes. The basic idea behind the SM method is that a change in the physical surface in the region will make the intensity ratios vary inconsistently in the corresponding regions of two images. Liu et al. [12] defined circular shift moments (CSM) to describe the illumination-independent characteristics of a region. Change detection is performed as a hypothesis testing based on the characteristics of the defined moments. These methods can detect changes accurately under time-varying illumination since they are developed based on shading or reflection models [13]. However, they all employ ratio of intensities or sums of intensities to describe structural characteristics of a region. This means they may perform poorly over the dark regions in images due to the denominator of the ratio becoming insignificant for these regions. In addition, the illumination independency of the measures is obtained only on the simple diffuse reflection model.

The robustness of region-based difference measure is based on its dissimilar representations of different local structural features, or texture. Background texture remains relatively stable with respect to noise and illumination changes unless it is cov-
erected by moving objects or abrupt lighting change occurs. On the other hand, for a background region covered by a moving object, even though the gray level distributions of the foreground and background regions may be similar, the texture features of the two regions are usually different. From this point of view, a region-based difference measure based on texture representation will be more accurate for change detection than those based on local statistics. However, for the homogeneous regions of foreground and background, the texture difference measure will become less valid than the simple intensity difference. There are usually many such regions in man-made objects. Therefore, it is desirable to integrate the intensity and texture differences adaptively for robust change detection.

In this paper, a novel texture difference measure is proposed. It is derived from the relations of local gradient vectors to compute the difference of texture patterns. Mathematical analysis indicates that the measure is robust with respect to noise and illumination changes based on a more general illumination model. Two methods to integrate intensity and texture differences have been developed. One is the adaptive weighted sum of two difference measures, and the other is an optimal integration by minimizing an energy function with additional constraint of smoothness. Real-time implementation of the new technique is also proposed. Experimental results indicate that the proposed methods can detect changes much more accurately than those based on intensity or structure difference only.

The rest of this paper is organized as follows. In Section II, a novel texture difference measure is proposed. Section III describes two methods to integrate the differences of intensity and texture. In Section IV, the experiment results and comparisons of the proposed technique with the pixel-based and region-based methods are presented. Finally, this paper is concluded in Section V with discussions.

II. TEXTURE DIFFERENCE MEASURE

Texture is an important feature for image representation. It represents the spatial arrangement of pixel gray levels in a region [14]. There are many approaches developed to describe the texture feature [15], such as the co-occurrence matrix, Fourier power spectrum, Markov random field, and Gabor filters. Being developed for image segmentation and classification, they aggregate pixels based on the similarities of local structures within a neighborhood. These similarities are extracted from statistics of gray level arrangement or spectrum feature over a great number of samples. This goes against the requirement for real time change detection to compare two small regions. For this purpose, a simple and efficient texture difference measure is derived here.

A good texture difference measure should be able to represent the difference between two local spatial gray level arrangements accurately. Since the gradient value is a good measure to describe how the gray level changes within a neighborhood and it is less sensitive to light changes, it can be used to derive an accurate local texture difference measure. Let \( p = (x, y) \) be a point in an image plane, then the \( i \)th frame and its gradient vector can be represented as \( f_i(p) \) and \( f_i^g(p) = (f_i^x(p), f_i^y(p)) \) with \( f_i^x(p) = \nabla_x f_i(p) \) and \( f_i^y(p) = \nabla_y f_i(p) \). Here the partial derivatives are generated using the Sobel operator. At a point \( p \), the cross-correlation of gradient vectors of two frames can be defined as

\[
C_{12}(p) = f_1^g(p) \cdot f_2^g(p) = \| f_1^g(p) \| \cdot \| f_2^g(p) \| \cos \theta
\]

where \( \theta \) is the angle between the vectors. Similarly, the auto-correlation of a gradient vector can be defined as

\[
C_{ii} = f_i^g(p) \cdot f_i^g(p) = \| f_i^g(p) \|^2, \quad i = 1, 2.
\]

Obviously, we have

\[
C_{11}(p) + C_{22}(p) \geq 2 \| f_1^g(p) \| \cdot \| f_2^g(p) \| \geq 2C_{12}(p).
\]

If there is no change within the neighborhood of a point \( p \) in the two frames, the local texture features of this region in the two frames would be similar, i.e., there are no great differences in length and direction of \( f_1^g(p) \) and \( f_2^g(p) \). Hence, one has \( C_{11}(p) + C_{22}(p) \approx 2C_{12}(p) \). On the other hand, if there is a change in the corresponding regions of two frames, there would usually be large differences between the local textures of the two frames since they are from surfaces of different objects. In this case, the gradient vectors \( f_1^g(p) \) and \( f_2^g(p) \) would be different in both length and direction. Then \( 2C_{12}(p) \) would become much smaller than \( C_{11}(p) + C_{22}(p) \). Let’s define a measure of the neighborhood gradient difference as

\[
R_t(p) = 1 - \frac{\sum_{u \in \mathcal{M}_p} 2C_{12}(u)}{\sum_{u \in \mathcal{M}_p} (C_{11}(u) + C_{22}(u))}
\]

where \( \mathcal{M}_p \) denotes the \( 5 \times 5 \) neighborhood centered at \( p \). Since \( \| \theta \| > 90^\circ \) would indicate much difference between the vectors with \( C_{12} < 0 \), we normalize \( R_t(p) \) by setting \( R_t(p) = 1 \) as the maximum value. The robustness of \( R_t \) with respect to illumination changes and noise can be illustrated as follows.

Illumination Factor: According to the general illumination model [17], the light from a point is composed of three parts, which are the ambient light, diffuse reflection, and specular reflection. The intensity of a pixel in an image frame is represented as

\[
I = I_{amb} + k_{diff} I_L \cos \theta + k_{spec} I_L \cos^\phi
\]

where \( k_{diff} \) and \( k_{spec} \) are the reflection coefficients, \( I_{amb} \) the ambient-light intensity, \( I_L \) the intensity of light source, \( \theta \) the angle of incidence between the light-source direction and the normal of the surface, and \( \phi \) the viewing angle relative to the specular reflection direction. To visualize the effect of illumination change on a particular background pixel \( p \), one can treat \( \theta \) and \( \phi \) as constants from one frame to the other [11], [12] and rewrite the intensity equation as

\[
f_i(p) = I_{amb}^{(i)} + f_L^{(i)} (k_{diff}^{(i)}(p) + k_{spec}^{(i)}(p)).
\]

Consequently, one gets

\[
f_2(p) = k_f f_1(p) + a
\]
with

$$k_I = \frac{I^{(2)}_L}{I^{(1)}_L}, \quad \alpha = f^{(2)}_{amb} - k_I f^{(1)}_{amb}$$  \hspace{1cm} (8)$$

where $k_I$ and $\alpha$ can be assumed constants within a small image region between two frames under the assumption of the smoothness of local illumination [11], [12], [16]. Equation (7) becomes the same as the general linear brightness model proposed by Negahdaripour [16] and it is more general than the models used in [11], [12]. Since the gradient values are calculated by using Sobel operator, it can be shown that

$$f^{(s)}_g(p) = k_I f^{(s)}_I(p), \quad s = x, y.$$  \hspace{1cm} (9)$$

This equation reveals that the gradient is invariant to the ambient illumination changes. Substituting (9) into (1), (2), and (4), one obtains

$$R_t(p) \approx 1 - \frac{2k_I}{1+k_I^2} \left( \frac{1-k_I}{1+k_I^2} \right)^2.$$  \hspace{1cm} (10)$$

When the intensity of light-source does not change too much, one has $k_I \approx 1$, and $R_t(p) \approx 0$. In practice, even though the light-source intensity decreases to 26.8% ($k_I = 0.268$) or increases to 373.2% ($k_I = 3.732$) of the original level, $R_t(p)$ is still below 0.5 as shown in Fig. 1. Hence, it can be concluded that the gradient difference measure is robust to illumination changes.

Noisy Environment: Image noise is usually modeled as an additive white Gaussian with zero mean and variance of $\sigma_n^2$. Moreover, the additive noise is assumed to be independent of brightness. Hence, the noise in the partial derivatives of $f_g^x(p)$ and $f_g^y(p)$ which are obtained by applying the Sobel operator would follow a Gaussian distribution of $N(0, \sqrt{12}\sigma)$. The partial derivative images can be represented as

$$f_{g}^{(s)}(p) = f_{g}^{(s)}(p) + n_{g}^{(s)}(p), \quad s = x, y$$  \hspace{1cm} (11)$$

where $f_{g}^{(s)}(p)$ is the partial derivative of the original noise-free image and $n_{g}^{(s)}(p)$ the additive white Gaussian noise. Substituting (1), (2), and (11) to (4) with statistic analysis, one can get

$$R_t(p) \approx \frac{SNR^2}{1 + SNR^2} \cdot \overline{R}_t(p) + \frac{1}{1 + SNR^2}$$  \hspace{1cm} (12)$$

where $\overline{R}_t(p)$ is the measure obtained with noise-free data, $SNR$ is the signal-noise-ratio which is defined as $(\sigma_T/\sigma_n)$ with

$$\sigma_n^2 = E(n_{g}^2)^2,$$  \hspace{1cm} (13)$$

$$\sigma_T^2 = \frac{1}{4} \sum_{i=1}^{2} \sum_{s=x,y} E(f_{g}^{(s)})^2,$$

where $\sigma_T^2$ is the average of texture signal power. The relations of $R_t(p)$ and $\overline{R}_t(p)$ with respect to different SNR values are plotted in Fig. 2. The ideal relation is a straight line from (0, 0) to (1, 1). It can be observed that as long as there are significant texture features in the region $\mathcal{M}_p$, one can obtain $R_t(p) \approx \overline{R}_t(p)$. Furthermore, if there is significant structure difference between frames in the region $\mathcal{M}_p$ ($R_t(p) \approx 1$), one has $R_t(p) \approx \overline{R}_t(p)$ even when SNR is small. These features indicate that the gradient difference measure is robust to noise for texture regions.

On the other hand, if the neighborhoods of a point $(p)$ in both frames are homogeneous with low SNR, $R_t(p)$ becomes invalid. This can also be concluded from (4) since the denominator of the second term would be small. In this situation, one can not classify a point with certainty based only on the local gradient difference. To tackle this, a validity weight, $w(p)$, for gradient difference at each point is computed. Let

$$w(p) = \begin{cases} 1, & \text{if } g_t(p) > 2T_w \pm 1 \sigma_n \sigma_T, \\ 0, & \text{otherwise}. \end{cases}$$  \hspace{1cm} (15)$$

where $T_w$ is a parameter based on image noise distribution that will be determined later. Consequently, one can define

$$d_t(p) = w(p) \cdot R_t(p)$$  \hspace{1cm} (16)$$

as the texture difference between two frames.

III. INTEGRATION OF TEXTURE AND INTENSITY DIFFERENCES

The texture and intensity differences can complement each other. They are two different views to the difference between two frames. A better change detection can therefore be achieved by integrating the information from these two sources.
A. Intensity Difference

Let \( d(p) \) be the intensity difference of two frames. Before integration, the intensity difference measure should be normalized into the same range of \([0, 1]\) as the texture difference measure. This is done by applying a simple slope function defined as

\[
d_t(p) = \begin{cases} 
1, & \text{if } |d(p)| > 2T \\
|d(p)|/(2T), & \text{otherwise.}
\end{cases}
\]  

(17)

The functions used in (15) and (17) can be considered as fuzzilization of two types of difference measures.

The parameters \( T \) in (17) and \( T_w \) in (15) should be properly chosen to cope with image noise. Since the image noise can be modeled as Gaussian noise following \( N(0, \sigma_d) \), the noise in the intensity difference image has a Gaussian distribution \( N(0, \sigma_d) \) with \( \sigma_d = \sqrt{2}\sigma \). The variance of noise \( \sigma_d \) may be estimated by the Least Median of Squares (LMedS) method [7]. However, due to the effect of illumination changes, the shifts of brightness values for unchanged regions should be compensated to compute \( \sigma_d \). For this purpose, a modified algorithm based on the LMedS method is proposed here. First, the shift of brightness value at each point is calculated as

\[
\bar{d}(p) = \frac{1}{M} \sum_{u \in M_p} d(u). 
\]  

(18)

Then, the difference image in which brightness shifts have been compensated becomes

\[
d'(p) = d(p) - \bar{d}(p).
\]  

(19)

Applying the LMedS method to image \( d'(p) \), one obtains \( \tilde{\sigma}_d \), the estimation of \( \sigma_d \). After this, the mean shift of brightness values for unchanged regions is estimated as

\[
\bar{d}_s = \frac{1}{|N_s|} \sum_{p \in N_s} d(p) 
\]  

(20)

with

\[N_s = \left\{ p : (|d'(p)| < 2\tilde{\sigma}_d) \land (|d(p)| < T_{20\%}) \right\}
\]  

(21)

where \( T_{20\%} \) is the median of \(|\bar{d}(p)|\) and \(|N_s|\) is the number of points in the set \( N_s \). Obviously, most points of \( N_s \) come from unchanged regions. In addition, the average operation over \( N_s \) can eliminate the effect of a few outliers (foreground points) in \( N_s \). In practice, if there is a global illumination change, \( |\bar{d}_s| \) will be large, but if there are only local illumination changes such as shadows, the value of \( |\bar{d}_s| \) will be small. From these estimations, the parameter \( T \) can be chosen as \( |\bar{d}_s| + 3\tilde{\sigma}_d \). From the analysis in the last section, it is known that if an image only contains noise (on plain background), one gets

\[
\sqrt{\frac{1}{M} \sum_{u \in M_p} C_i(u)} \approx \sqrt{2}\sigma_s = \sqrt{2k}\sigma \approx 2\sqrt{3}\sigma_d.
\]  

Since the weight \( w(p) \) is used to represent the significance of texture features within the neighborhood, \( T_w > 2\sqrt{3}\sigma_d \) is at least required. On the other hand, due to the reliability of texture difference measure, it is undesirable to chose a large \( T_w \). In this work, \( T_w = 1.5 \cdot 2\sqrt{3}\sigma_d \approx 5\sigma_d \) is chosen empirically.

B. Weighted Integration (WI)

Due to noise and illumination changes, the texture difference \( d_t(p) \) is regarded as more reliable and robust than the simple intensity difference \( d_i(p) \). Hence, one should depend on \( d_t(p) \) only if the corresponding region has no texture. This can be accomplished by using \( w(p) \) [see (15)] to create an adaptive weighted sum as

\[
d_i(p) = w_i(p) \cdot d_i(p) + w_t(p) \cdot d_t(p) 
\]  

(23)

where \( w_i(p) = w(p) \) and \( w_t(p) = 1 - w(p) \). \( d_i(p) \) would be within the range of \([0, 1]\) and the changes can be detected by thresholding \( d_i(p) \) at mid-point (0.5), which can be considered as a defuzzilization process.

C. Minimized Energy Integration (MEI)

Combining different measures, constraints, and high level knowledge about the underlying task can obtain better segmentation. Various techniques have been proposed for this purpose and the most popular one is energy function minimization [18]–[20]. The energy functions can be computed with respect to the continuous or discrete variables. These methods can produce global optimal results of segmentation and are robust to noise and errors from low level detection measures. Since the minimization (maximization) of energy functions with respect to discrete variables usually employs a simulated annealing procedure for global optimization, which requires vast amounts of computation [20], this work applies the error energy minimization with respect to the continuous variables to integrate the intensity and texture differences.

To obtain the error energy of the difference measures, it is assumed that there is an ideal elastic measure surface \( d_i^r(p) \) attracted by the difference measures with error energy as

\[
E = \sum_p \left( c_\alpha \phi^2(p) + \beta U(p) \right) 
\]  

(24)

where

\[
\phi^2(p) = w_i(p)(d_i^r(p) - d_i(p))^2 + w_t(p)(d_t^r(p) - d_t(p))^2 
\]  

(25)

represents the drag energy from both \( d_i(p) \) and \( d_t(p) \), and

\[
U(p) = \frac{1}{N_N} \sum_{u \in N} [d_t^r(p) - d_t^r(u)]^2 
\]  

(26)

represents the smoothness constraint. \( N \) is the immediate eight-connected neighborhood of \( p \) and \( N_N = 8 \).

To obtain the ideal difference measure surface, let \( \partial E/\partial d_i^r(p) = 0 \), and we will have

\[
d_i^r(p) = K_\alpha [w_i(p)d_i(p) + w_t(p)d_t(p)] + K_3 \frac{\beta}{N_N} \sum_{u \in N} d_t^r(u) 
\]  

(27)
where
\[ K_\alpha = \frac{\alpha}{\alpha + 2\beta}, \quad K_\beta = \frac{2\beta}{\alpha + 2\beta}. \] (28)

Equation (27) can be rewritten as
\[ d_{it}^{(k+1)}(p) = K_\alpha \cdot d_t(p) + K_\beta \cdot \frac{1}{N_N} \sum_{u \in N} d_{it}^{(k-1)}(u). \] (29)

Then, let \( d_{it}^{(0)}(p) = d_t(p) \) and the error be
\[ \Delta(k) = \sum_p \left| d_{it}^{(k)}(p) - d_{it}^{(k-1)}(p) \right| \] (30)

the approximation of \( d_{it}^{(k)}(p) \) can be obtained by iteratively computing \( d_{it}^{(k)}(p) \) until \( \Delta(k) < \varepsilon \) or \( k > I_{\text{max}} \), where \( I_{\text{max}} \) is a predetermined maximum number of iterations, and \( \varepsilon \) is the upper error limit for an acceptable solution.

The convergence of the algorithm can be discussed as follows. According to (29), there is
\[ d_{it}^{(k)}(p) = d_{it}^{(0)}(p) + \frac{K_\beta}{N_N} \sum_{u \in N} \left( d_{it}^{(k-1)}(u) - d_{it}^{(k-2)}(u) \right). \] (31)

From (30) and (31), it is easy to obtain
\[ \frac{\Delta(k)}{\Delta(k-1)} \leq \frac{2\beta}{\alpha + 2\beta} \leq 1 \] (32)

or \( \Delta(k) \leq \gamma \Delta(0) \) with \( 0 < \gamma < 1 \) when \( \alpha > 0 \) and \( \beta > 0 \). Hence, \( d_{it}^{(k)}(p) \) converges quickly to \( d_t(p) \). For simplicity, one can set \( \alpha + 2\beta = 1 \). If \( \beta \) is set small, \( d_{it}^{(k)}(p) \) will converge quickly to the equilibrium state with rough surface. On the other hand, if \( \beta \) is set large, \( d_{it}^{(k)}(p) \) will converge slowly, but it can generate a smooth surface. Good segmentation can be achieved if one selects \( 2\beta/\alpha \approx 10 \) based on the work in [19]. The last step thresholds \( d_{it}^{(k)}(p) \) at 0.5 to extract the moving objects. In this paper, the parameters are set as \( \alpha = 0.1, 2\beta = 0.9, I_{\text{max}} = 300 \), and \( \varepsilon = 0.1 \).

D. Computational Complexity

It is observed that a large amount of computation for \( d_t(p) \) is spent on the sums over the local region \( M_p \) [see (4) and (18)]. Obviously, there are many repetitions of computation if we calculate the sum separately at each point. An iterative algorithm is designed here to eliminate the repetitive computation as much as possible. Suppose \( f(x, y) \) is an image, \( M_p \) is a \( N \times N \) local region of \( p = (x, y) \), then the local sum on \( M_p \) is
\[ \tilde{f}(x, y) = \sum_{i=-N/2}^{N/2} \sum_{j=-N/2}^{N/2} f(x + i, y + j). \] (33)

It can be rewritten as
\[ \tilde{f}(x, y) = \tilde{f}(x-1, y) - s_c \left( x - \frac{N}{2} - 1, y \right) + s_c \left( x + \frac{N}{2}, y \right) \] (34)

where \( s_c(x, y) \) is defined as
\[ s_c(x, y) = \sum_{j=-N/2}^{N/2} f(x + i, y + j) \] (35)

where \( s_c(x, y) \) can also be calculated iteratively as
\[ s_c(x, y) = s_c(x, y - 1) - f \left( x, y - \frac{N}{2} - 1 \right) + f \left( x, y + \frac{N}{2} \right). \] (36)

Hence, four additions are enough to compute each \( \tilde{f}(x, y) \) if the \( s_c(x, y) \) image is calculated first.

The computational cost to obtain \( d_t(p) \) at each point is evaluated in Fig. 3 where the number in the parenthesis indicates the operations in each step. In this figure, the cost of calculating local sums in (14) is omitted since they have been done in (4). From this evaluation, the total computational cost to get \( d_t(p) \) for frames with size of \( M_p \times N_p \) is about \( 65M_fN_f \) operations. In contrast, the CSM method [12] will require \( 100M_fN_f \) operations for a high resolution of change detection result.

With the precomputed \( d_t(p) \), (29) can be rewritten as
\[ d_{it}^{(k)}(p) = d_{it}^{(k-1)}(0) + K_b \sum_{u \in N} d_{it}^{(k-1)}(u) \] (37)

where \( d_{it}^{(k)}(p) = K_\alpha \cdot d_t(p) \) and \( K_b = K_\beta/N_N \). Each iteration of (37) will cost \( 10M_fN_f \) operations. Hence, if 50 iterations are required for convergence, the total number of operations to obtain \( d_{it}^{(k)}(p) \) is \( 65M_fN_f \). If the hardware of parallel architecture is employed, only two operations are needed at each point in one iteration.

IV. EXPERIMENTAL RESULTS

The proposed technique has been evaluated on images of many real scenes with superior results over existing methods. In this section, some experimental results will be given to demonstrate the performance of the technique with respect to illumination changes and the presence of different noise levels. The pixel-based adaptive intensity difference (AID) [7] and the region-based SM method [11], similar to the SCSM method [12], were chosen for comparison. In the experiments, the threshold for SM was chosen manually to get the best result for each scene. Since MEI has the effect of smoothness over WI’s results, a morphological operator with size of 3 \times 3 was applied to the WIs results for comparison (MWI). Besides visual comparison, the results were also evaluated quantitatively in terms of
false negative number and rate: the number and portion of foreground pixels that are missed;
• false positive number and rate: the number and portion of background pixels that are marked as foreground.
The “ground truth” foreground pixels were marked by hand. The errors are summarized in Tables I–III while the segmentation results are displayed from Figs. 4–6 with layout as: frames (a, b), “ground truth” (c), and the results of AID (d), SM (e), WI (f), MWI (g), and MEI (h).

The first example (Fig. 4) from an indoor scene captured a person walking in front of a book shelf with image size of 192 × 144 pixels. Shadow of the person was cast on the shelf. In the lower region of the left shelf, about 50 gray levels of intensity decrease was observed. The estimated noise level was \( \hat{\sigma}_d = 11.86 \) and the mean shift of brightness \( \bar{\Delta}_s \) was \(-3.17\). From the segmentation results, it can be seen that the proposed technique produced more foreground points and less mis-classified background points than the AID and SM methods did. Examining the figures in Table I, one can observe the numerical support to this result.

The second example (Fig. 5) experimented on a scene of a bright T-shirt covering a glassy surface with the same image size as the first example. The noise level was estimated as \( \hat{\sigma}_d = 5.9 \), while the mean shift of brightness was \( \bar{\Delta}_s = 7.72 \). A closer observation revealed that the brightness of the white wall had increased about 25 gray levels. Comparing the results of AID and SM, one can observe that, in the homogeneous regions of both background and foreground, the intensity difference measure works well, while in the texture regions of background or foreground, the structure measure distinguish the difference better. Integrating these two types of difference measures can achieve better segmentation. The results from MWI and MEI supported this argument.

The third example was taken from a corridor scene where two people were walking toward an elevator (Fig. 6). In this example, the head of the man on the right was positioned in...
front of a dark region of the background. Much of it did not show up in the result of AID. In the result of SM, the interior parts of the human bodies were not detected. But in the results of WI and MEI, both the regions of human heads and bodies were segmented quite well, and the shadows on the frame of lift door (about 55 gray level decreases) were reduced and separated.

The behaviors of the proposed techniques with respect to the parameters were also investigated. Different values of $T$ and $T_w$ were also tested. $T$ had been selected as $|\hat{d}_a| + K_T \hat{d}_d$ with $K_T$ being 2, 3, 4, and 5, and $T_w$ as $K_{T_w} \hat{d}_d$ with $K_{T_w}$ being 4, 5, and 6. In general, the performance of the proposed techniques are not too sensitive to $K_T$ and $K_{T_w}$. Obviously, the lesser the $K_T$ and $K_{T_w}$, the more the points are marked. Different normalization functions used in (15) and (17), such as the step function and the sine function, were also experimented with. The performances of the sine and slope functions were similar but better than other functions such as the step function. However, the slope function is simpler in calculation.

V. CONCLUSION

In this work, the integration of intensity and texture differences for robust change detection has been investigated. A new accurate texture difference measure based upon the relation of two gradient vectors was derived. The mathematical analysis shows that the measure is robust with respect to noise and illumination changes based on a more general illumination model. Two methods to integrate the differences of intensity and texture, i.e., the adaptive weighting combination and the optimal integration by minimizing an error energy function, have been proposed. An analysis for automatic selection of parameters was also presented. The computational complexity of the proposed techniques was analyzed and real-time implementation was developed. Both the visual and quantitative evaluations of experiment results showed that impressive improvement of change detection has been achieved by integrating intensity and texture differences. Hence, the proposed technique is more accurate and robust for change detection in complex environments.

To get a meaningful segmentation, the future task would be toward recognizing relevant and irrelevant changes in the scene. This can be achieved by exploiting not only the low-level difference measure but also the high-level knowledge about 3-D objects, 3-D perspective projections, and usual events in a scene.

ACKNOWLEDGMENT

The authors would like to thank CGIT from NTU for providing the experiment facilities.
REFERENCES


Liyuan Li (S’99) received the B.Sc. degree in 1985 and the M.Sc. degree in 1988 in automatic control from Southeast University of China. Since 1999, he has been pursuing the Ph.D. degree in the School of Computer Engineering of Nanyang Technological University, Singapore.

From 1988 to 1999, he was on the faculty of Southeast University, where he was an Assistant Lecturer (1988–1990), Lecturer (1990–1994), and an Associate Professor (1995–1999). Since September of 2001, he has been a Research Staff Member with the Kent Ridge Digital Labs, Singapore. His research interests include computer vision, image processing, pattern recognition, machine learning, and artificial intelligence.

Mr. Li is a member of the IEEE Signal Processing Society and IEEE Computer Society.

Maylor K. H. Leung (M’92) received the B.Sc. degree in physics from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1979, and the B.Sc., M.Sc., and Ph.D. degrees in computer science from the University of Saskatchewan, Saskatoon, SK, Canada, in 1983, 1985, and 1992, respectively.

Currently he is an Associate Professor with Nanyang Technological University, Singapore. His research interest is in the area of computer vision, pattern recognition, and image processing. His particular interest is in improving security using visual information.