Published in Proceedings of the "Nonlinear Image Processing and Pattern Analysis XII" 4304, 85-97, 2001
which should be used for any reference to this work

# Countering Illumination Variations in a Video Surveillance Environment

R. Costantini, G. Ramponi[a], J. Bracamonte, B. Piller[b], M. Ansorge, and F. Pellandini

Inst. of Microtechnology, Uni. of Neuchâtel, CH-2000 Neuchâtel, Switzerland
[a]Uni. of Trieste, I-34127 Trieste, Italy
[b]SysNova S.A., CH-9525 Lenggenwil, Switzerland
e-mail: roberto.costantini@unine.ch

## ABSTRACT

In the field of video technology for surveillance applications it is often necessary to cope with the phenomenon of illumination variations. In fact, if not compensated, such variations can falsely trigger the change detection module that detects intrusions in video surveillance systems, thus affecting their reliability. Many studies have been made to solve the change detection problem under varying illumination conditions. Most of the published methods, however, rely only on the luminance information. The algorithm proposed in this paper exploits *independently* the information of each band of the RGB color space of the video sequences, thus producing a change detection algorithm that is more robust to illumination variations. These illumination variations are globally modeled by the so-called Von Kries model (also known as *diagonal scaling model*). This model is generally used to solve the color constancy problems, where conformance to a reference image illumination has to be guaranteed, like in color image retrieval applications. The use of this model is motivated by its low computational cost and by the interest of studying the relationship between color constancy and change detection. Based on practical experiments which confirm the interest in this method, new and more robust change detection algorithms are expected to be designed. In addition, the paper proposes the use of an iterative scheme whose aim is to improve the results obtained in the change detection module, and which is independent of this module, i.e., it can be used with other change detection schemes. It will be shown that the iteration can improve the quality of the final change mask, thus permitting to obtain a more effective change detection scheme.

**Keywords:** Change detection, video-surveillance, color constancy.

## 1. INTRODUCTION

Within the field of video surveillance different situations of potential danger can occur and have to be signaled by the surveillance system. The nature of these situations can vary, but from the point of view of the surveillance system it can be simply represented as the occurrence of a *change* in an image of the video sequence with respect to an image taken as a reference. From one side this change can be due to the intrusion of an unknown person in a room/building, and so, it can be related to the presence of *motion* in a scene that, normally, is supposed to remain static. From the other side the change can derive from the presence in the current image of a new object which does not appear in the reference image and is not foreseen in that particular context, or remains too long at the same place (as for example suspicious bags left in airports or railway stations); in this case the change that should be signaled by the surveillance system is not related to motion, but to the *permanence* of something in the scene. Either it derives from motion or permanence of objects/persons in a scene, the change which is produced in the images constituting the video sequence has to be detected and signaled. This task can be accomplished by one or more *change detection* algorithms embedded in the surveillance system.

The phenomenon of the variation of illumination that could occur in a video sequence can significantly affect the performance of change detection algorithms. In fact, such a variation could be detected as a change caused by the presence of a person in the scene, thus constituting what is called a *false alarm.* Since it is preferable to have the less false alarms as possible, the illumination variations have to be taken into account and compensated.

Generally it is quite difficult to make a classification of the possible cases of illumination variations, since such variations can be very different one from the other. They can be produced both in outdoor scenes, where they can be caused by the change of the sunlight illumination caused by the movement of a cloud over the scene for example, or by the headlight of a car, and in indoor scenes, caused by the turning on of a light. Therefore, a general method to treat all the cases has not yet been found.

The method proposed in the present paper covers some significant cases, that can be representative of common real world situations. Furthermore, the proposed method permits to deepen the analysis of the relationship between two distinct fields of the image processing discipline like color constancy and change detection.

In Section 2 a description of several change detection algorithms is provided, whereas the so-called Von Kries model is presented in Section 3. Section 4 is then discussing the proposed algorithm, the corresponding experimental tests and achieved results proving their robustness being thoroughly presented in Section 5. Finally, the conclusion are drawn in Section 6.

## 2. CHANGE DETECTION ALGORITHMS

The methods which specifically detect changes due to the presence of motion in a scene are usually referred to as *motion detection* algorithms, and their task is not only to signal the change, but also to provide some supplementary information about the motion, as the displacement vectors or the velocity field. Many techniques have been developed to accomplish this task. In a great number of algorithms an estimate of the motion fields is found by computing the *optical flow* of the images forming the video sequence. Since the optical flow is defined as the apparent motion of brightness patterns in these images, it generally corresponds to the motion field, but in some cases it can bring to false estimation, like in the well-known barber pole sequence [1], where the real movement of the pole is circular, while the perceived optical flow is vertical. Excluding the special cases, like the latter one, these methods are quite effective and usually produce good results. Other methods make use of a minimization strategy of a cost function to find the displacement vectors, as it is commonly done in the motion estimation block of a typical video coder. An evaluation of some motion detection techniques can be found in [2], while in [3],[4],[5] some new algorithms are proposed. A characteristic common to all the motion detection algorithms is their computational burden, even though much work has been done to reduce it.

The present work is not interested in the problem of motion detection, since the aims of the proposed algorithm is to obtain a change detection mask and not a motion field. For this reason, the motion detection methods will not be further described.

The design of effective change detection algorithms has been of interest in many studies in the image processing field. At the moment there exist several methods that are suitable for the detection of changes in a video sequence, and that differ one from the other according to the change/noise model they use and the distinct strategies they apply. A first classification of the change detection algorithms can be found in [7]. In this work two new methods are proposed: the *Derivative Model Method* and the *Shading Model Method*. Mainly, these two methods have been designed to cope with the phenomenon of illumination variation, and it is shown that the Shading Model Method is particularly robust even to strong variations of illumination. This method is based on the model of intensity value proposed by Phong [6], where the intensity value $I_p$ of a pixel in an image is modeled as the product of the illumination $I_i$ and a so-called *shading coefficient* $S_p$:

$$I_p = I_i S_p \ . \tag{1}$$

In order to establish if a pixel in position $(i,j)$ of an image has changed, the algorithm computes the variance of the ratios of the pixel intensity values in an $N \times N$ window centered at the pixel's position:

$$\mathrm{VAR}_{\mathcal{W}}\left(\frac{I_c}{I_r}\right) = \frac{1}{N^2}\sum_{(i,j)\in\mathcal{W}}\left(\frac{I_c^{(i,j)}}{I_r^{(i,j)}} - \mu\right)^2 \tag{2}$$

where $\mathcal{W}$ represents the set of pixels in the window of size $N \times N$, $I_c^{(i,j)}$ and $I_r^{(i,j)}$ are the intensity of the pixel $(i,j)$ belonging to the current and reference images respectively, and $\mu$ is the mean value of the ratios $I_c^{(i,j)}/I_r^{(i,j)}$ within the window. If the variance of the ratios exceeds a certain threshold, then the pixel is classified as changed, otherwise it is considered as unchanged. This method exhibits good performances even with strong illumination changes. Thanks to its effectiveness, other and more recent works have been based on this method, as those proposed in [9] and [10]. The shortcoming of this method is that many division operations have to be computed, resulting in a quite high computational cost.

An alternative technique is proposed in [8], where instead of calculating the variance of the intensity ratios of the images, a change detection is performed using some *circular-shift moments* of the pixels intensity within a specified window. It is demonstrated that this method requires a smaller number of operations per pixel than the previous method, still assuring the same quality of results. The algorithm proposed in [8] belongs to the broad class of *statistical change detection* algorithms, which constitute a great portion of the change detection algorithms, both because of the relevant studies that have been carried on and for the results that they allow to obtain.
The statistical change detection algorithms are all characterized by the same processing of the information. In a first phase some assumptions are made about the statistics of the noise present in the image, then, to evaluate the possible change of a pixel, a *distance function* is calculated between a pixel in the current image and the correspondent one in the reference image, generally considering the values of the other pixels in the neighborhood, i.e., on a window centered at the pixel's position. The statistical properties of this function are then studied, in order to predict a range of values that are assumed in the case the pixel has not changed. A subsequent threshold mechanism allows to distinguish changed pixels from unchanged ones.

In [11] the functions that are proposed are the following:

$$\bar{\Delta}_1^i = \gamma \sum_{k \in \mathcal{W}_i} |I_r^k - I_c^k| \quad \text{or} \quad \bar{\Delta}_2^i = \gamma \sum_{k \in \mathcal{W}_i} (I_r^k - I_c^k)^2 \tag{3}$$

where $\mathcal{W}_i$ represents the window around the pixel $i$, $I_r^k$ and $I_c^k$ represent respectively the reference and current value of the pixels intensity in that window, and $\gamma$ is a normalization parameter adjustable to different noise levels. After the threshold step, which is based on statistical evaluations, a change mask is obtained. Since this mask can present holes and rough contours, the algorithm uses in addition a Markov Random Field refining technique, which consists of imposing the minimization of a cost function obtained from the mask by some reasonable considerations. The final result is quite good, since the change mask is accurate and well refined. However, this method has not been designed specifically for change detection in presence of illumination variations.
On the other hand in [12] the function proposed is:

$$F_i = \frac{1}{N} \sum_{k \in \mathcal{W}} (d_k - \mu_i)^4 \tag{4}$$

where $d_k = (I_r^k - I_c^k)$, and $\mu_i$ is the mean value of $d_k$ within $\mathcal{W}$. It can be easily seen that this function is the fourth-order moment of the differences between pixels intensities in the chosen window. This algorithm takes into consideration also other characteristics of the images, thus obtaining a high-order-statistic (HOS) test. The resulting detection mask, as in the previous algorithm, is further processed in order to obtain more coherent and homogeneous regions. In this case, some morphological operations are performed. The results obtained are good, but this method, like the last one presented, has not been directly designed for applications in which the illumination could suddenly vary. To cope with the problem of illumination changes, more specific algorithms have been designed.

In [13] an algorithm is proposed which applies the statistical change detection method described in [11] on the *reflectance* component of the image, i.e., the component which is called $S_p$ in Eq.1, instead of its intensity. Since the reflectance component of an image represents the inner properties of the objects contained in the image, and is thus not influenced by the illumination variation, the resulting change detection algorithm is more robust to changes of illumination. The reflectance component is extracted from the intensity image by the use of a homomorphic filter. The whole algorithm guarantees better performance with respect to the simple original change detection algorithm, since the reflectance image is less dependent on changes of illumination. Obviously, the homomorphic filter cannot *exactly* separate the reflectance and the illumination contributions that constitute the image intensity, and so, only illumination variations that do not have high spatial frequencies can be eliminated. This corresponds to the situation of a quite uniform illumination change.
In [14] and [15] a different strategy is applied to cope with the phenomenon of the illumination variation. In these two works the reference image is *updated* over the time, such that the illumination variation can be included and thus compensated. In [14] this update is done by a continuous refresh of the reference image in function of the speed of the change of illumination, while a Kalman filter is used in [15].

The methods that have been presented are all based on the intensity values of the images. If the change detection has to be performed on color images, usually described in the RGB color space of the acquisition devices, it is possible to apply these methods by working on the Y component of the video sequence, obtained by a simple color conversion

to the YUV space. But there exist also some change detection algorithms that exploit the information given by the color components and work directly on the RGB color space.
An interesting example of a method of this kind can be found in [16]. The aim of the algorithm is to separate the shadows produced by the presence of a person in a scene from the person itself, and, to some extent, to compensate for a global illumination change that could affect the scene. The concept of this algorithm is to compute a distance between the color pixels that form the current image of the scene and their *expected values*, which have been obtained by a prediction mechanism that is based on the statistical properties of the background which is fixed and constitutes the reference image. If this distance exceeds certain thresholds, which are statistically adapted frame by frame, then a change is signaled. This method applies a strategy that is similar to the one proposed in the present paper, and it ensures optimal results when dealing with shadows and global illumination variations. Anyway, the method proposed in [16] was not designed to be robust to strong illumination variations, even though the research is on progress.

The method proposed in the present paper consists of a change detection in the RGB color space and makes use of a simple model for the change of illumination that could occur in the scene. This model is called *Diagonal Scaling Model*, and will be described in the following section.

## 3. THE DIAGONAL SCALING MODEL

Let $E(\vec{x},\lambda)$ be the intensity of illumination of wavelength $\lambda$ given by a source $E$ and arriving at a position $\vec{x}$ on a surface present in a scene filmed by a video camera. The actual pixel values measured by the $k$-th sensor of the camera are:

$$\rho_k(\vec{x}) = \int_\lambda E(\vec{x},\lambda)[S_s(\vec{x},\lambda) + S_b(\vec{x},\lambda)]R_k(\lambda)d\lambda \quad \text{with } k = 1,2,3 \quad ; \tag{5}$$

where $R_k(\lambda)$ is the spectral response function of the sensor, and $S_s(\vec{x},\lambda)$ and $S_b(\vec{x},\lambda)$ are respectively the *surface* and *body* reflections (also known as specular and diffuse reflections) of the surface. Eq.5 represents the so-called *dichromatic reflectance model* [17] for the color values perceived by a video camera. As can be noticed, the perceived values depend on the sensor spectral response, the illumination, and the inner characteristic of the surface, which is represented by two factors $S_s(\vec{x},\lambda)$ and $S_b(\vec{x},\lambda)$, thus explaining the reason why the model is called dichromatic. The surface reflection is dominant in metallic surfaces, where the incoming light is reflected almost maintaining its spectral content, so the dependence on $\lambda$ is quite weak, while for dielectric/matt surfaces the body reflection represents the major contribution to the perceived color values, since the light is almost completely scattered in all directions.
If the illumination changes from $E(\vec{x},\lambda)$ to $E'(\vec{x},\lambda)$, then the perceived colors change according to the transformation:

$$\rho_k(\vec{x}) = \int_\lambda E(\vec{x},\lambda)[S_s(\vec{x},\lambda) + S_b(\vec{x},\lambda)]R_k(\lambda)d\lambda \longrightarrow \rho'_k(\vec{x}) = \int_\lambda E'(\vec{x},\lambda)[S_s(\vec{x},\lambda) + S_b(\vec{x},\lambda)]R_k(\lambda)d\lambda \quad . \tag{6}$$

The relationship between $\rho_k(\vec{x})$ and $\rho'_k(\vec{x})$ is not evident, but if some assumptions are made about the color sensors properties and the illumination sources, then it can be expressed in a simpler form.
In fact, it can be simply demonstrated that, if the spectral response of the sensors were a perfect Dirac impulse $\delta_k(\lambda)$, then the relationship between the perceived pixel colors would simply be:

$$\rho'_k(\vec{x}) = a_k(\vec{x})\rho_k(\vec{x}) \tag{7}$$

where $a_k(\vec{x}) = E'(\vec{x})/E(\vec{x})$, i.e., the color response that we have under the illuminant $E'(\vec{x},\lambda)$ would be simply obtained by the previous one with a separated scaling of each color component. This can be expressed in the following way:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} a_R & 0 & 0 \\ 0 & a_G & 0 \\ 0 & 0 & a_B \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{8}$$

where the known RGB color space is used instead of the notation $\rho_k$.
This representation of the change of the color values obtained by a variation of the illuminant of the scene corresponds to the Diagonal Color Scaling Model, also known as Von Kries model. This is one of the first models that have been used to solve the so-called *color constancy problem*, where the term "‘color constancy"’ refers to the insensivity of

a vision system to the variation of the color of a scene caused by a change of the illuminant. To some extent, this property is naturally present in the human vision system, while it has to be embedded in a machine vision system. Solving the color constancy problem corresponds to obtaining a representation of the scene which is independent of the illuminant. The Von Kries model allows to implement this property, since it is possible to pass from an image taken under an unknown illuminant into the same image under a different and known illuminant, using the coefficients $a_R$, $a_G$, and $a_B$.

For real images, the spectral responses of a CCD or CMOS camera are not exactly representable as Dirac functions, but in any case, this model can effectively represent the color transition derived from a change in illumination. This is further proved in [18], where it is demonstrated that for the case where surface illumination can be approximated with a 3-dimensional basis and reflectance with a 2-dimensional basis (Maloney's 2-3 restrictions), then, *independently* of the form of the spectral response of the sensor, the color constancy transformation can be expressed by a diagonal matrix. This is a fundamental result, since many scenes are well representable within the Maloney's restrictions, and so the diagonal model fits a great number of real situations.

The diagonal scaling model has been widely used for solving the problem of color constancy in many algorithms where this was a necessary condition to guarantee good results. For example this model is used in [19] in the field of image retrieval, where a change of the illumination in the query image can avoid its retrieval in a given database, where all images are supposed to be obtained under the same illuminant.

In the algorithm proposed in the present paper, this model of color transition will be used to make a prediction of the value of the color components of a pixel in the current frame of a video sequence, obtained from the corresponding values in a reference frame. Thanks to this prediction it will be possible to establish if this pixel has significantly changed by comparing its predicted and actual values. Such predictions correspond to a normalization of the current image with respect to the illuminant of the reference image.

## 4. DESCRIPTION OF THE ALGORITHM

### 4.1. The concept

Detecting a change in an image $I_n$ with respect to a reference image of a video sequence is equivalent to finding a binary image $\mathcal{M}_n$ such that:

$$\mathcal{M}_n(i,j) = \begin{cases} 1 & \text{if } I_n(i,j) \text{ is changed} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $I_n(i,j)$ indicates the pixel of position $(i,j)$ of the image $I_n$. The change in position $(i,j)$ can result from the presence of an object/person in the scene, from an illumination variation, and/or from the presence of noise. The proposed method aims at creating a binary mask that corresponds to the change caused by the object/person only, and is thus robust to illumination variations and noise.

The concept of the algorithm is to find a *prediction* of the values of the pixels in the current frame, based on the values of the pixels in the reference frame and on an estimate of the global color variations of the current frame. Once this prediction is computed for each pixel, then a *distance* between this prediction and the actual values of the pixel is evaluated, enabling the selection of changed pixels from unchanged ones in a successive thresholding phase.

To understand this concept, let us suppose that a possible change of illumination in the current image of a video sequence alters the values of the pixels exactly according to the diagonal scaling model. In this case, the transition of the RGB color components of the pixels in the current frame would be specified by Eq.8, i.e., each pixel affected only by a change of illumination would have the following color components:

$$R_n^{i,j} = a_R^n R_0^{i,j} \quad G_n^{i,j} = a_G^n G_0^{i,j} \quad B_n^{i,j} = a_B^n B_0^{i,j} \tag{10}$$

where $n$ indicates the frame number. On the contrary, the pixels affected by a change due to the presence of a person/object in the scene would have values different from those indicated by Eq.10. By computing a *distance* between the actual values of the pixels in the whole image and their *prediction* found using Eq.10, it is possible to decide upon pixel changes. The parameters $a_R^n$, $a_G^n$, and $a_B^n$, which are not known *a priori*, could be found by considering a pixel which is not changed in the current image, dividing its RGB color components by those of the corresponding pixel in the reference image.

In practice, for real images this method has two problems. First, the diagonal scaling model does not perfectly fit a change of illumination that could occur in real images. In fact, this model does not consider the shadows and highlights that can be present in real images affected by illumination variations, causing a variation of the parameters

$a_R$, $a_G$, and $a_B$ according to the pixel position: higher values for pixels corresponding to highlighted zones, smaller values for pixels in shadowed zones. This shortcome of the model will be partially compensated by the use of an adequate distance metric and a threshold strategy, still maintaining the structure of the change detection algorithm, i.e., the concept of prediction/distance evaluation of the pixels in the current frame.

Second, it is not possible to establish *a priori* which pixels are changed or not, so that it is not possible to find the parameters $a_C^n$ with $C = R, G, B$ on the basis of one pixel only. To solve this problem, the proposed algorithm computes a *global* estimation of the parameters $a_R$, $a_G$, and $a_B$ using the following equations:

$$\hat{a}_R^n = \sum_{i,j \in \mathcal{S}_R} \frac{R_n^{i,j}}{R_0^{i,j}} \quad , \hat{a}_G^n = \sum_{i,j \in \mathcal{S}_G} \frac{G_n^{i,j}}{G_0^{i,j}} \quad , \text{and } \hat{a}_B^n = \sum_{i,j \in \mathcal{S}_B} \frac{B_n^{i,j}}{B_0^{i,j}} \tag{11}$$

where $\mathcal{S}_C = \{(i,j) : C_0^{i,j} \neq 0\}$ with $C = R, G, B$ respectively, the hat on $a_C^n$ indicating the estimate. It can be noted that the estimation corresponds to the mean value of the ratios between the color components, computed all over the images, i.e., for each pixel position. The prediction of the pixel values is done by the following formulae:

$$\tilde{R}_n^{i,j} = \hat{a}_R^n R_0^{i,j} \quad \tilde{G}_n^{i,j} = \hat{a}_G^n G_0^{i,j} \quad \tilde{B}_n^{i,j} = \hat{a}_B^n B_0^{i,j} \tag{12}$$

The definition of the estimate of Eq.11 has two shortcomings. First it is quite computationally expensive, since three divisions per pixel are needed. Second, since the mean is calculated over the whole current ($n$-th) image, it takes into consideration also those pixels that could be affected by a change that is not due to illumination. In order to overcome these two problems, a modified estimation strategy is elaborated. The new parameters are defined as follows:

$$\hat{a}_R^n = \frac{\sum_{i,j \in \mathcal{S}} R_n^{i,j}}{\sum_{i,j \in \mathcal{S}} R_0^{i,j}} \quad , \hat{a}_G^n = \frac{\sum_{i,j \in \mathcal{S}} G_n^{i,j}}{\sum_{i,j \in \mathcal{S}} G_0^{i,j}} \quad , \text{and } \hat{a}_B^n = \frac{\sum_{i,j \in \mathcal{S}} B_n^{i,j}}{\sum_{i,j \in \mathcal{S}} B_0^{i,j}} \tag{13}$$

where $\mathcal{S} = \{(i,j) \mid \text{pixel of position } (i,j) \text{ belongs to the image}\}$. This equation produces estimates that have always a greater value than those found by Eq. 11, but the difference is not so relevant, and practical tests demonstrate that these new parameters perform almost like the previous ones. More important, as can be noted from Eq.13, only three divisions per image have to be done, thus reducing the computational complexity. The second shortcoming mentioned above in connection with Eq.11 will be solved by applying an iterative scheme, as explained in the following part of this section.

### 4.2. Implementation

The change detection scheme proposed in this paper is showed in Fig.1. In the Parameters Estimation block the values $\hat{a}_R^n$, $\hat{a}_G^n$, and $\hat{a}_B^n$ are computed according to Eq.13. For real time applications, where a certain speed of calculation has to be ensured, the estimate of the parameters can be done on subsampled images, resulting in a lower computational cost and still ensuring a good performance of the algorithm.

If the current image is affected neither by a change due to a person/object nor by illumination variations, then the values of these estimates are approximately 1, exhibiting some differences because of noise, anyway producing predicted values that are almost equal to the reference values. On the contrary, if a change of illumination is present in the scene, the estimates indicate the global change of the color components of the image *independently* for each R, G, and B color channel.

In the Prediction block the values of the current image are predicted from those of the reference image according to the estimated parameters of the diagonal scaling model. The prediction is made according to Eq.12. In the Distance block the distance between the predicted values and the actual ones is computed. Many metrics have been tested in order to select one that ensures good performance. The one that has been chosen is the following:

$$d_n^{i,j} = \sqrt{\left(\frac{R_n^{i,j}}{\tilde{R}_n^{i,j}}\right)^2 + \left(\frac{G_n^{i,j}}{\tilde{G}_n^{i,j}}\right)^2 + \left(\frac{B_n^{i,j}}{\tilde{B}_n^{i,j}}\right)^2} \tag{14}$$

where $d_n^{i,j}$ indicates the distance between corresponding pixels of the actual and predicted reference image in position $(i,j)$. The formula of Eq.14 does not actually represent a true distance in the common Euclidean sense, but it allows

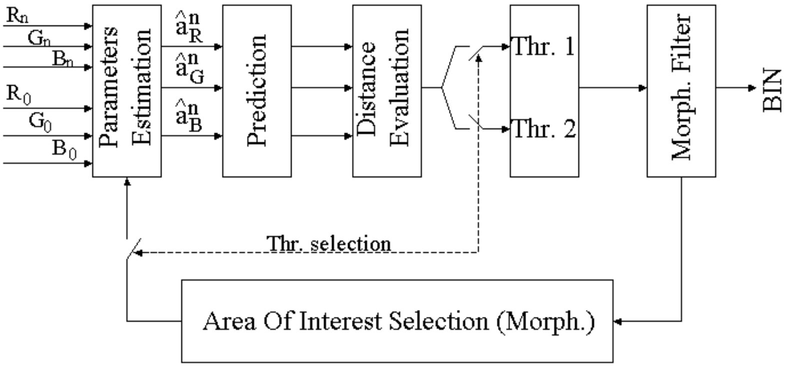

**Figure 1.** Change detection scheme.

obtaining good results in the change detection algorithm. It is worth noticing that the single terms of the sum in Eq.14 can be written as:

$$\frac{C_n^{i,j}}{\tilde{C}_n^{i,j}} = \frac{C_n^{i,j}}{\hat{a}_C^n C_0^{i,j}} = C_n^{i,j}\frac{1/\hat{a}_C^n}{C_0^{i,j}} = \frac{C_n^{i,j}}{(C_0^{i,j})_{\mathrm{norm}}} \tag{15}$$

where as usual $C = R, G, B$. Eq.15 shows that the distance between two images can be seen as the distance between the current image and a normalized image, which is obtained by the reference image imposing a global illumination variation whose characteristics are similar to those of the current frame. In other words, the illuminant of the current image is reported back to that of the reference image, thus solving the color constancy problem mentioned in Section 3.

This normalization can be done using other and more accurate methods proper to the color constancy theory, still maintaining the rest of the scheme. This allows to obtain a better performance by just changing a single block of the scheme, independently of the other ones, ensuring a greater flexibility to the whole change detection scheme.

The Threshold block is deciding whether a pixel of the current image has changed or not. This decision is made by a threshold mechanism applied to the distance between corresponding pixels.

From experimental tests it has been noted that an adaptation of the threshold is required to obtain good results, especially when a strong variation of illumination occurs. These tests have shown that in presence of a positive variation of illumination the threshold has to be more selective. For this reason, the proposed algorithm performs an adaptation of the threshold. This is done using the parameters computed before, with a computational cost of one division per image, which is negligible. The threshold is updated according to the ratio $Z = \hat{a}_R^n / \hat{a}_B^n$ which has been noted to follow the illumination changes within a certain precision, as it is observed from Fig.2.

The horizontal axis indicates the frame index of a test sequence featuring a strong variation of illumination near the 5-th frame where a lamp is switched on in the scene. One notices that the ratio $Z = \hat{a}_R^n / \hat{a}_B^n$ increases after this event, and is then stabilizing while the light is maintained on. This behaviour was verified with various test sequences, as it will be shown in the next section.

This suggests to use the ratio $Z = \hat{a}_R^n / \hat{a}_B^n$ to adapt the threshold to the illumination variation. The binary mask obtained in this block is then:

$$\mathcal{M}_1 = \begin{cases} 1 & \text{if } d_{i,j}^n < \eta_1/Z \text{ or } d_{i,j}^n > \eta_2 \cdot Z \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $\eta_1$ and $\eta_2$ are two parameters that adapt the threshold according to the video sequence characteristics. It can be noticed from Eq.16 that the classification considers either small values of the distance, for which the threshold becomes more selective by a division by $Z$, or large values, for which the threshold becomes more selective by a multiplication by $Z$.

To have an idea of the result obtained in this block, Fig.3(c) reports the change detection mask obtained from Eq.16. The reference and current frame are shown in Fig.3(a) and Fig.3(b) respectively, where it is clearly noted that the change of illumination is very strong. It can be observed from Fig.3(c) that most parts of the change are ignored by

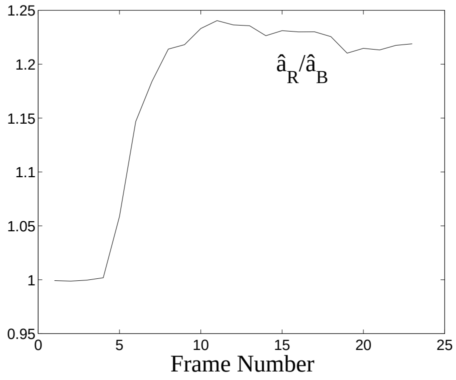

**Figure 2.** Modeling the variation of illumination.

the algorithm, while the change due to the person is taken into consideration.

The Morphological Filter block performs an opening on the binary mask $\mathcal{M}_1$. The aim of this block is twofold, namely to obtain a binary mask which does not contain holes and whose contours are quite regular, and to eliminate part of the noise present in the first mask. This operation is very cheap from a computational point of view, since it involves operating on binary images. Fig.3(d) reports the result of this process.

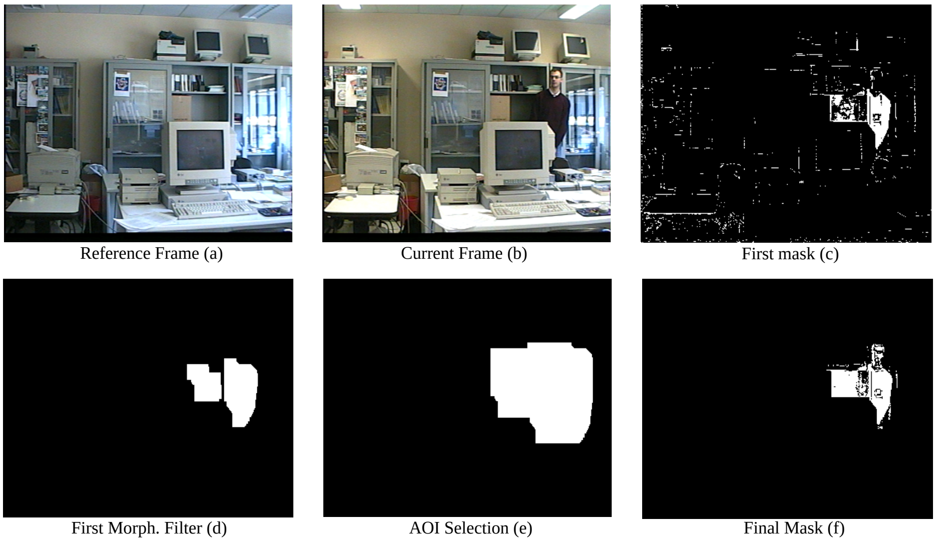

**Figure 3.** Results of the different phases of the algorithm.

As can be noticed, an iteractive process is being carried on. The area corresponding to the last mask in Fig.3(d) is dilated using a second morphological filter, thus obtaining a mask which is called AOI (Area Of Interest) as depicted in Fig.3(e). This mask specifies an area of interest around the changed pixels obtained by the first change detection iteration. The same change detection algorithm described above is then applied to the pixels located within the AOI. The advantage of applying this iterative process lies in the fact that the estimated parameters $\hat{a}^n_C$ become more accurate, since they can be processed in those parts of the image that have not changed due to the person, i.e., Eq.13 is calculated on $\mathcal{S} = \{(i,j) \mid (i,j) \in (AOI)^C\}$, where $(AOI)^C$ is the complementary set of $(AOI)$. Furthermore, within the AOI the threshold can be less selective, thus allowing to obtain a better change detection mask. The final change detection mask is then achieved applying:

$$\mathcal{M}_2 = \begin{cases} 1 & \text{if } d^n_{i,j} < \epsilon_1 / Z \text{ or } d^n_{i,j} > \epsilon_2 \cdot Z \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where $\epsilon_1$ and $\epsilon_2$ have the same function as the parameters $\eta_i$ used in Eq.16. The final result is reported in Fig.3(f).

## 5. EXPERIMENTAL TESTS AND RESULTS

### 5.1. The Test Sequences

The proposed algorithm has been tested on four video sequences. Each of them is representative of a video surveillance situation where the change detection algorithm has to face different problems in order to find a valid detection mask. The sequences used are:

- The Hall-Monitor Sequence;
- The Labo1 Sequence;
- The Labo2 Sequence;
- The Labo3 Sequence.

They have been acquired at a rate of 3 frames per second, and are composed of 30, 24, 24, and 24 frames respectively. The proposed method has been tested on the whole set of data, but for simplicity we show only the results obtained from a single frame for each of the test sequences. In Fig.4 the reference and current frames used to test the change detection algorithm are shown.

The test sequences can be divided into two main groups: sequences that are not affected by a variation of the illumination of the scene, as the sequences Hall-Monitor and Labo3, and sequences that present a strong variation of the illumination, like Labo1 and Labo2 sequences, where the illumination variation is produced by switching on a neon lamp in a laboratory setting. These two types of sequences were used in order to test the algorithm both in a relatively simple case, i.e., when there is no illumination change, and in more difficult situations, which correspond to a variation of illumination. The proposed algorithm should be robust in both cases.

**Figure 4.** Images used for the tests.

The well-known Hall-Monitor sequence was chosen since, although it represents a quite easy situation for a change detection algorithm, it presents some interesting aspects, like the presence of an object which is left in the scene and remains still (the hand-bag of the first man that comes into the scene), the occurrence of *transparencies* caused by the fact that the color of some parts of the subjects walking in the scene is almost the same as the corresponding zones of the background, etc.

The sequence Labo3 was chosen to test the performance of the iterative process which estimates the parameters $a_C^n$ introduced in the previous section. Indeed, this sequence shows in the first frames a man walking across the scene, whose body occupies a big portion of the scene. This characteristic can affect the first estimation, since the parameters $\hat{a}_C^n$ are computed over the whole image. The second iteration should remove this shortcoming, and this sequence is used to verify this assumption.

The two remaining sequences, Labo1 and Labo2, present both a strong illumination variation. This variation has been obtained by turning on the neon lamp of the considered laboratory setting. Simultaneously, a person was walking through the scene. The aim of the change detection algorithm is that of detecting only the person and not the change of illumination. The nature of both sequences renders this task quite difficult, since, as can be seen in Fig.4, the variation of the illumination causes many collateral phenomena, like the formation of shadows, reflections, and diffractions on the objects present in the scene. For this reason these sequences are quite difficult to tackle and represent a relevant task for a change detection algorithm, thus being a good test for assessing the robustness of the proposed method.

### 5.2. Comparison with other Algorithms

In order to test the performance of the proposed algorithm, it is compared to two different change detection algorithms. The first algorithm that has been chosen is the one proposed in [10], which is based on the theory of [7] and has been briefly described in Section 2. This algorithm makes use only of the luminance components of the image. The second algorithm is the one proposed in [16]. This algorithm has been chosen since it presents some similarities with our method, although it is not designed for change detection in presence of strong illumination variations. In fact it uses the color information of the images to handle the color constancy problem. This approach is similar to the one proposed in the present article.

Before presenting the results of the tests, it is necessary to notice one important difference between the two algorithms that have been chosen and the proposed one: the latter is structured iteratively, i.e., to obtain the final change detection mask two scans of the images are needed, while the two former algorithms are single scan-based. This fact has some implications when comparing the three methods, since a fair comparison has to be ensured. For this reason, in the following part of this section, the final result of the change detection algorithm proposed in this paper will be shown not only after two iterations, but also after the first iteration. This allows a fair comparison of the discussed methods.

### 5.3. Results of the Tests

The first tests were devoted to the evaluation of the robustness of the parameter estimation done by the proposed algorithm. In Fig.5 the values of the ratio $\hat{a}_R/\hat{a}_B$ are reported for the frames of the test sequences. It can be noted that this ratio is effective in signaling the presence of an illumination variation. Indeed, in Fig.5(a) this ratio maintains almost the same values, since in the Hall-Monitor sequences there is no illumination change. On the other hand, in Fig.5(b) and Fig.5(c), which correspond to the sequences Labo1 and Labo2, it can be clearly noted that the illumination change present in the scene is well detected by the parameter $\hat{a}_R/\hat{a}_B$. An additional observation is that the first and second iterations give almost the same results, as in these sequences the persons walking through the scene occupy a small part of the scene.

On the contrary, in Fig.5(d) it is clearly observed that the first estimate suffers from the presence of the person in the first frames, but in the second iteration, marked by $(\hat{a}_R/\hat{a}_B)_2$, the estimation is correct, since the ratio remains almost constant, as there is no change of illumination. The estimated parameters $\hat{a}_C^n$ that are used to obtain the

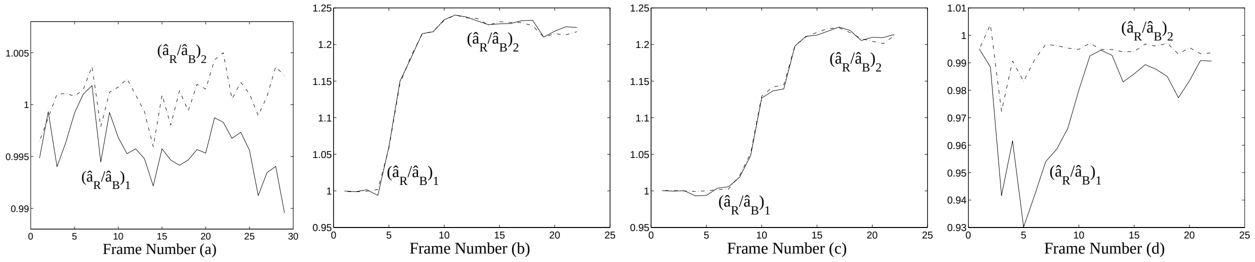

**Figure 5.** Evaluation of parameters estimation; Hall-Monitor (a); Labo1 (b); Labo2 (c); Labo3 (d); continuous line: first scan $(\widehat{a_R}/\widehat{a_B})_1$; dotted line: second scan $(\widehat{a_R}/\widehat{a_B})_2$.

plots of Fig.5 were computed by subsampling the test images by a factor three both horizontally and vertically.

The second tests concern the performance of the whole change detection algorithm. These tests were carried out on the entire video sequences and Fig.6 reports the results obtained for a single frame, where the reference and current frames used are those of Fig.4. The images denoted by $\mathrm{a}_i$ with $i = 1, 2, 3, 4$ present the result of the first scan of the proposed method, while those denoted by $\mathrm{b}_i$ correspond to the final detection mask (second iteration). The images denoted with $\mathrm{c}_i$ and $\mathrm{d}_i$ correspond to the detection masks obtained by the algorithms described in [10] and in [16] respectively.
It can be noticed that for the Hall-Monitor sequence, the three methods produce similar results, while for the sequence Labo1 and Labo2 the proposed algorithm exhibits less noise in the final detection mask, thus offering a more robust solution to the illumination variations. This is obtained mainly by its iterative structure. In any case, the results already obtained in the first scan are comparable with those of the other two methods. The method proposed in [16] presents some difficulties with the sequences affected by an illumination variation, precisely because it is not designed to face strong illumination changes. On the other hand it ensures optimal results in the sequence Labo3, as can be seen in Fig.6($\mathrm{d}_4$).
The algorithm proposed in [10] exhibits a good noise removal ability in the sequences Labo1 and Labo2, whereas it is less effective in sequence Labo3. In Fig.6($\mathrm{c}_4$) it can be noted that many holes appear in the person, and this is mainly because the person present in the scene has a shirt of uniform color: in the zones where the background is also uniform the ratio between the luminance components has a low variance, so that the pixels corresponding to these zones are not detected as changes. This problem could be solved by lowering the threshold, but this would cause an increase of the noise in the image, thus cancelling the expected benefit.

For our method, the threshold parameters used in the experiments were optimized for the sequence Labo1 in order to obtain a mask which most closely corresponds to the actual change of the image. The same parameters were used for the other test sequences, without further optimization; this is an advantage of the technique we propose. In fact, it should be noticed that for comparing the algorithms, it was necessary to perform the parameter optimization separately for each sequence, and the optimal values found differ significantly. The parameter values are provided in Table 1 and Table 2.

| Method | Labo1<br>(Ref. frame index = 10; Cur. frame index = 61) | Labo2<br>(Ref. frame index = 10; Cur. frame index = 61) |
|---|---|---|
| Proposed | $\eta_1 = 1.3$, $\eta_2 = 5$, $\epsilon_1 = 1.5$, $\epsilon_2 = 2$ | $\eta_1 = 1.3$, $\eta_2 = 5$, $\epsilon_1 = 1.5$, $\epsilon_2 = 2$ |
| [10] | $T_{LDD} = 0.2$ | $T_{LDD} = 0.3$ |
| [16] | $\tau_{CD} = 10$, $\tau_{\alpha lo} = -0.5$ | $\tau_{CD} = 10$, $\tau_{\alpha lo} = -0.5$ |

**Table 1.** Parameters used for the sequences where illumination changes were produced.

| Method | Hall-Monitor<br>(Ref. frame index = 1; Cur. frame index = 108) | Labo3<br>(Ref. frame index = 1; Cur. frame index = 16) |
|---|---|---|
| Proposed | $\eta_1 = 1.3$, $\eta_2 = 5$, $\epsilon_1 = 1.5$, $\epsilon_2 = 2$ | $\eta_1 = 1.3$, $\eta_2 = 5$, $\epsilon_1 = 1.5$, $\epsilon_2 = 2$ |
| [10] | $T_{LDD} = 0.03$ | $T_{LDD} = 0.03$ |
| [16] | $\tau_{CD} = 2$, $\tau_{\alpha lo} = -0.2$ | $\tau_{CD} = 10$, $\tau_{\alpha lo} = -0.3$ |

**Table 2.** Parameters used for the sequences where illumination changes were not produced.

It can be noted that, even though the number of parameters to adjust is four in the present algorithm, they ensure good results for sequences that present very different characteristics from one to the other. In any case the research is still going on about reducing the number of these parameters.
The morphological operator that was used in the Morphological Filter block of Fig.1 is based on two consecutive *opening* operations, obtained by an erosion and a dilation done respectively with a $3 \times 3$ and $9 \times 9$ square mask. The AOI is obtained by a dilation operator with a $32 \times 32$ square mask.

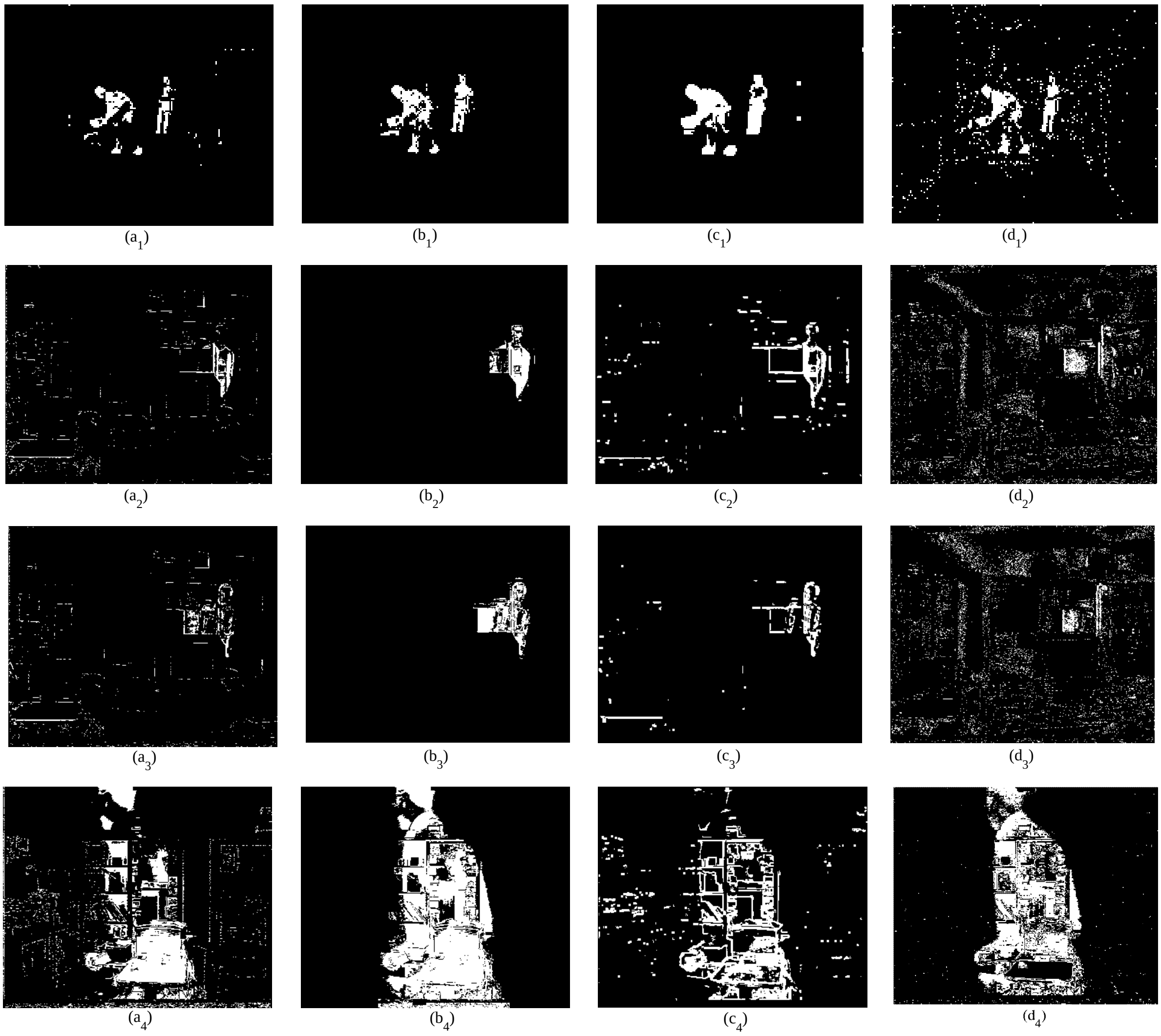

**Figure 6.** Results of the change detection algorithms. (a$_i$) first scan, and (b$_i$) second scan with the proposed method; (c$_i$) method of [10]; (d$_i$) method of [16]; Hall-Monitor: i=1, Labo1: i=2, Labo2: i=3, Labo3: i=4.

## 6. CONCLUSIONS

A new change detection algorithm was proposed whose characteristic is to be robust to strong illumination changes. Based on results known from the color constancy theory, the algorithm compensates for a possible illumination change in an image by reporting back the current illuminant to a reference one, where the detection is carried on. The change detection mask which is obtained is further processed in a second scan, in order to get better results. This scheme was shown to be quite robust to strong illumination changes even in presence of shadows and highlightings affecting the scene, and its computational cost is very low. Further research will be devoted to finding a better model for the change of colors in presence of an illumination change, and to extending the method to various real situations.

## ACKNOWLEDGMENTS

This work was originally part of the first author's Master Thesis performed at the Image Processing Laboratory of the University of Trieste, Italy. The research is presently going on at the Institute of Microtechnology, University of Neuchâtel, Switzerland, in the frame of a project financed by the Swiss Commission for Technology and Innovation (CTI) under Grant C-4374, whose support is acknowledged.

# REFERENCES

1. At disposal at *http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/OWENS/LECT12/node4.htm*
2. B. Galvin, B. McCane, K. Novins, D. Manson, and S. Mills, "Recovering motion Fields: An Evaluation of Eight Optical Flow Algorithms," *Ninth British Machine Vision Conf. (BMVC)*, Southhampton, United Kingdom, 1998.
3. E. Deléchelle and J. Lemoine, "Détection du mouvement fondée sur un modèle d'interaction électrique," *Vision Interface '99*, Trois-Rivières, Canada, May 1999.
4. J. Monteil and A. Beghdadi, "Une méthode rapide et robuste d'estimation du champ de déplacement," *Vision Interface '99*, pp. 520–527, Trois-Rivières, Canada, May 1999.
5. M. J. Black and P. Anandan, "A framework for the robust estimation of optical flow," *Fourth International Conf. on Computer Vision, ICCV-93*, pp. 231–236, Berlin, Germany, May 1993.
6. B.T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, **18**, pp. 311–317.
7. K. Skifstad and R. Jain, "Illumination Independent Change Detection for Real World Image Sequences," *Computer Vision, Graphics, and Image Processing*, **46**, pp. 387–399, 1989.
8. S.-C. Liu, C.-W. Fu, and S. Chang, "Statistical Change Detection with Moments under Time-Varying Illumination," *IEEE Trans. on Image Processing*, Vol. 7, No. 9, pp. 1258–1268, Sept. 1998.
9. E. Durucan, J. Snoeckx and Y. Weilenmann, "Illumination Invariant Background Extraction," *Proc. of the 10th International Conf. on Image Analysis and Processing (ICIAP '99)*, pp. 1136–1139, Venice, Italy, Sept. 27-29, 1999.
10. E. Durucan and T. Ebrahimi, "Robust and illumination invariant change detection based on linear dependence for surveillance application," *European Signal Proc. Conf., EUSIPCO 2000*, Vol. II, pp. 1041–1044, Tampere, Finland, September 5-8, 2000.
11. T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Processing*, **31**, pp. 165–180, 1993.
12. A. Neri, S. Colonnese, G. Russo, P. Talone, "Automatic moving object and background separation," *Signal Processing*, **66**, pp. 219–232, 1998.
13. D. Toth, T. Aach, and V. Metzler, "Bayesian spatio-temporal motion detection under varying illumination," *European Signal Processing Conf., EUSIPCO 2000*, Vol. IV, pp. 2081–2084, Tampere, Finland, Sept. 5-8, 2000.
14. L.Marcenaro, G.Gera and C.Regazzoni, "Adaptive Change Detection Approach for Object Detection in Outdoor Scenes Under Variable Speed Illumination Changes," *European Signal Processing Conference, EUSIPCO 2000*, Vol. II, pp.1025–1028, Tampere, Finland, September 5-8, 2000.
15. G.L. Foresti, "Object Recognition and Tracking for Remote Video Surveillance," *IEEE Trans. on Circuits and System for Video Technology*, Vol. 9, No. 7, pp. 1045–1062, October 1999.
16. T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," *IEEE ICCV'99 FRAME-RATE Workshop*, Kerkyra, Greece, Sept. 1999.
17. S.A. Shafer, "Using color to separate reflection components," *Color Research and Applications*, **10(4)**, pp. 210–218, 1985.
18. G.D. Finlayson, M.S. Drew, and B.V. Funt, "Diagonal Transforms Suffice for Color Constancy," at disposal at *http://www.cs.sfu.ca/~colour/publications/ICCV/iccv_abs.html*
19. D. Koubaroulis, J. Matas, and J. Kittler, "Illumination invariant object recognition using the MNS method," *European Signal Processing Conf., EUSIPCO 2000*, Vol. IV, pp. 2173–2176, Tampere, Finland, Sept. 5-8, 2000.