MOTION DETECTION IN A VIDEO SEQUENCE USING BACKGROUND SUBTRACTION BY 3D STATISTICAL MODELING IN THE VISIBLE SPECTRUM: HOME SURVEILLANCE CAMERA

Ali Ouchar Cherif [1], Mankiti Fati Aristide [2], Mbaiossoum Bery Leouro [3], Abakar Mahamat Ahmat [3]

[1] Institut National Supérieur des Sciences et Technique d'Abéché, Tchad
[2] Ecole Nationale Supérieure Polytechnique de l'Université Marien NGOUABI, Congo Brazzaville
[3] Université de Ndjamena, Tchad

Received 18 June 2023
Accepted 19 July 2023
Published 31 July 2023

Corresponding Author

Ali Ouchar Cherif,
alioucharcherif@yahoo.fr

DOI
10.29121/ijetmr.v10.i7.2023.1351

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Motion recognition is one of the key applications of motion detection, which involves image processing. For this reason, the proposed work is merely an extension of the final year project. It will involve a detailed study of a method for automatic, real-time recognition of motion and identity in a video stream.
Recognizing the suspicious movement of an object or person in an image is a major challenge in the field of camera-based surveillance. This recognition of suspicious object movement involves several phases: detection, tracking, creation of a database or vocabulary against which movements can be compared in order to be recognized and qualified as suspicious or not. In this work, we focus on the phase of detecting a moving object or person in a video sequence. We study a number of existing methods for detecting the movement of objects or people in a video sequence, in order to select the best among them for possible implementation.

Keywords: Motion Detection, Computer Vision, Artificial Intelligence, Video Sequence, Image Processing, Background Subtraction

1. INTRODUCTION

In this article, we present a background subtraction method based on 3D statistical modeling of the image in the visible spectrum. This method takes into account variations in brightness and color due to lighting conditions or shadows, as well as changes in camera orientation or perspective. It is particularly suited to

How to cite this article (APA): Cherif, A. O., Aristide, M. F., Leouro, M. B., and Ahmat, A. M. (2023). Motion Detection in a Video Sequence Using Background Subtraction by 3d Statistical Modeling in the Visible Spectrum: Home Surveillance Camera. *International Journal of Engineering Technologies and Management Research, 10*(7), 35–44. doi: 10.29121/ijetmr.v10.i7.2023.1351

surveillance applications using fixed or moving cameras in outdoor or indoor environments.

There is a wide range of methods available, but this one has been tested on several image sequences from public databases Benabbas Siavosh (2012), Wu and Nevatia (2006). The results show that the method is capable of efficiently detecting moving objects in a variety of scenes, with a low false positive rate and good robustness to changes in lighting and background.

The aim of this research work is to highlight the movement of a person in a video sequence.

To achieve this, we will:

- Review existing methods for detecting moving objects;
- Justify the choice of moving object detection method in terms of the socio-economic context and other existing methods in the literature;
- Implement the chosen detection method with a real case of a moving person.
- Discuss the results with other results from previous work.

Finally, in the conclusion, we outline the prospects for integrating our work into the recognition of suspicious movements in a video sequence.

2. METHODS AND MODELING

The selected method will be used in the context of a surveillance camera system.

2.1. STATE OF THE ART

It is essential to justify the choice of this method after a comparative study of the advantages and disadvantages of other existing methods or approaches. There are various algorithms in the literature which have been designed to model everything static, eliminate shadows, and reward background evolution. The performance of these algorithms in estimating a more robust background varies from one to another.

Table 1

Table 1 Comparative Table of Methods Discussed Medjahed (2012)

Type		Advantages	Disadvantages
Background subtraction	**SAP (visible 2D)**	- Low complexity algorithm - Simple classification - Clear results	- Initialization/static scene - Non-rejected shadows
	SAP (visible 3D)	- Shade resistance - Depth information	-Initialization/static scene - Complexity and calculations - Multiple cameras
Differences of consecutive images		- Flexibility of use - Low complexity of basic algorithm - Flexible initialization	- Compulsory movement - Shadows not rejected - Incomplete detection
Optical flow		- Precise movement information - Tracking/prediction possible	- Complexity and calculations - Non-rejected shadows - Compulsory movement

		- Difficult to interpret
Color	- Easy to implement - Recognition of color to be detected	- Color database 2653 - Loss of detection due to color change.
Histogram	- Particle identification - Easy selection of reference particles - More information, fewer errors	- Recognition mismatch - Lack of balance - Difficulty of implementation

Of all the methods studied and compared in the table above, few meet all the requirements and needs of the objective we have set.

Hardware constraints preclude the use of three-dimensional information or infrared imaging equipment. Thus, thanks to their low complexity and reasonable processing time, two different approaches were finally selected: background subtraction (SAP) by statistical modeling (3D visibility) and motion detection by difference of consecutive images.

The latter method, while having a definite advantage in terms of initialization, suffers from certain limitations on the classification side, ultimately favoring the selection of the 3D statistical modeling background subtraction method in the visible spectrum, whose advantages are Kovac and Solina (2003) , Gobeau (2007)

- Detection of moving objects in a complex, dynamic scene, taking into account variations in lighting, shadows and background changes;
- 3D statistical modeling of the background, based on a robust estimation of the parameters of the multivariate Gaussian distribution for each pixel. This modeling makes it possible to adapt to background changes and reduce false positives;
- Operation in the visible spectrum, without the need for infrared or thermal sensors, reducing system costs and complexity.

2.2. STUDY MODEL: BACKGROUND SUBTRACTION USING STATISTICAL MODELING

Background subtraction simplifies further processing by locating regions of interest in the image. Based on a model of the environment and an observation, the aim is to detect what has changed. In the context of this work, regions of interest are those areas of the image where there is a high probability that a person or object is present.

The algorithm used for background subtraction by statistical modelling comprises three important steps: initialization, motion extraction (foreground) and model update.

2.2.1. INITIALIZATION

The first step is to model the background from the first N frames (N ≈ 30) of a video sequence. An intensity average is then calculated from these images for each pixel and for each channel (R, G, B). The intensity average of a given pixel is then reduced to the following equation:

$$\boldsymbol{\mu_c(x,y) = \frac{1}{N}\sum_{i=0}^{N} I_{i,c}(x,y)} \tag{1}$$

Where I_i is the ith initialization image, ***N*** the number of images used and c the channel channel selected.

The next step is to calculate the standard deviation σ for each pixel (and for each each channel) to be used as a detection threshold. This operation usually involves storing the first ***N*** images. However, a modified equation to circumvent this constraint incrementally and thus reduce memory of memory space. To achieve this, two accumulators are used: ***S(x, y)*** to store the sum of the pixel intensities, and ***S(x, y)*** to store the sum of the pixel intensities. store the sum of pixel intensities and ***SC(x, y)*** to store the sum of squares. Standard deviations can then be calculated using equation (2) Cucchiara et al. (2001):

$$\boldsymbol{\sigma_c(x,y) = \sqrt{\left(\frac{SC_c(x,y)}{N}\right) - \left(\frac{S_c(x,y)}{N}\right)^2}} \quad (2)$$

Moreover, it's interesting to note that **S(x, y)** can be reused for averaging, which avoids additional, superfluous operations.

2.2.2. FOREGROUND EXTRACTION

In order to extract motion in an image, the background pattern must first be subtracted from it. Each pixel, whose difference in absolute value exceeds the value **α × σ**, is then classified as a moving pixel. In the previous expression, the variable **α** represents a certain fraction of the standard deviation **σ**. In practice, this parameter lies in the interval **[2.0,4.0]** and depends on the desired level of exclusion. A binary motion mask can then be generated for each channel using the equation below [14]:

$$\boldsymbol{m_c(x,y) = \begin{cases} 1 \ si \ |I_c(x,y) - \mu_c(x,y)| > \alpha.\sigma_c(x,y) \\ \quad 0 \ \ \text{autrement} \end{cases}} \quad (3)$$

Where $\mathbf{m_c(x, y)}$ represents the motion mask for channel c and $\mathbf{I_c(x, y)}$ the input image to be analyzed. to be analyzed.

The equation $\mathbf{m_c(x, y)}$ represents the calculation of the motion mask for a single channel.

To use this algorithm with the 3 **(RGB)** channels of the images used, the individual must first be generated independently and then combined using a logical OR operator. As a result, if motion is detected for a pixel in a single a pixel in a single channel, this will be sufficient to modify its state. The following equation represents this combination, producing a single-channel motion mask Cucchiara et al. (2001):

$$\boldsymbol{M(x,y) = m_r(x,y) \cup m_g(x,y) \cup m_b(x,y)} \quad (4)$$

Once this operation has been completed, certain mathematical morphology operations [14] must be applied to eliminate noise and false detections. To achieve this, two (2) erosions and two (2) dilations are applied respectively in this order to the motion mask. Finally, the input image is combined with the mask to produce a three (3) channel (foreground) image containing only those pixels representing motion. This operation can be summarized in equation (5):

$$\boldsymbol{F(x,y) = M(x,y).I(x,y)} \quad (5)$$

Where **F(x, y)** represents the foreground image and ***I(x, y)*** the input image. The two images are combined by pixel-by-pixel multiplication for each channel.

2.2.3. UPDATING THE MODEL

During the acquisition period, certain regions of the scene may undergo lighting changes, making it essential to update the statistical background model. Thus, a gradual change in brightness (e.g., sunrise) will be incorporated into the model and will not be considered as motion. To achieve this, foreground extraction is performed with the current image, generating a motion mask M. The background model is then updated from the complement of M, i.e., using all pixels that are labeled as part of the background. Abrupt changes in the image are therefore not added to the model. Equation (6) illustrates this updating process Cucchiara et al. (2001):

$$\boldsymbol{\mu'}_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{y}) = (\mathbf{1} - \boldsymbol{\eta}).\, \boldsymbol{\mu}_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{\eta}.\, \boldsymbol{I}_{\boldsymbol{c}}(\boldsymbol{x}, \boldsymbol{y}).\, \overline{\boldsymbol{M}}(\boldsymbol{x}, \boldsymbol{y}) \tag{6}$$

Where ***μ'(x, y)*** represents an updated average background pixel and ***η*** the learning rate. The expression ***I_c(x, y).$\bar{M}$(x, y)*** represents the static pixels of the current image, i.e. those for which no change is associated Sigal and Athitsos (2000).

In order not to radically alter the background model, only a fraction ***η*** of the temporary image I***c(x, y).$\bar{M}$(x, y)*** is used. In practice, this learning rate can take values in the range [0.05, 0.25]. The higher the value of this parameter, the faster the changes will be integrated. This is equivalent to quickly forgetting the model built during the initialization phase. We recommend using relatively low values (e.g., 0.05).

Finally, the standard deviation is not adjusted or updated while the algorithm is running (i.e., once initialization has been performed), in order to reduce the amount of computation required. However, some further experiments should be carried out to verify the usefulness and impact of this update on the results Sigal and Athitsos (2000).

3. THE TOOLS WE USED

We used a number of tools and techniques to prepare our Code::Block IDE platform, so that we could carry out manipulations on image sequences.

3.1. MINGW (MINIMALIST GNU FOR WINDOWS)

This is an adaptation of the GNU development and compilation software (GCC-GNU Compiler Collection) for the Win32 platform. Unlike other applications, programs generated with MinGW don't need an intermediate compatibility layer (in the form of a DLL (Dynamic Link Library)) Gobeau (2007).

3.2. CMAKESFILES

It is mainly used to facilitate compilation and link editing, since in this process the final result depends on previous operations. The language system used in Cmakefiles is declarative programming. Unlike imperative programming, this means that the order in which instructions are executed is irrelevant Gobeau (2007).

3.3. CODE::BLOCKS

This is our free, cross-platform IDE (Integrated Development Environment). It is written in C++ using the wxWidgets library. For the moment, Code::Blocks is C and C++ oriented, but it can also be supported by other languages.

3.4. OPENCV (OPEN-SOURCE COMPUTER VISION)

OpenCV (Open Computer Vision) is an open-source library of computer vision algorithms algorithms, and is available for the C, C++, and Python languages.

With this library, you can load, display, and modify images, work with histograms and apply basic transformations (thresholding, segmentation, morphology...).

OpenCV is made up of 5 different libraries: **CV, CVAUX, CXCORE, HIGHGUI** and **ML**. **HIGHGUI**, for example, enables file manipulation and the display of a graphical interface, while ML enables data classification. There are many applications for this library, such as computers unlocked by facial recognition, or object tracking on video. Using OpenCV, it is possible to control a machine by eye movements and blinks for disabled people, and so many other possibilities Gobeau (2007) et Benabbas Siavosh (2012).

4. IMPLEMENTATIONS OF THE BACKGROUND SUBTRACTION DETECTION METHOD

We have chosen and studied the best method among those detailed in the state of the art to implement, lead a discussion, and underline the importance of this chosen method.

Detecting people in video sequences involves determining, for an image or sequence of images, whether people are present, and if so, their position. It can be used to identify moving objects in the background of a video sequence. It is a necessary segmentation technique, often widely used in computer vision. Most vision algorithms regard this step as a necessary pre-processing step, the aim of which is to reduce the search space and improve the computational cost performance of an application (possibly recognition or tracking or detection) at a higher level of abstraction. More specifically, this step is needed for object tracking, human action recognition, video surveillance, fall detection and so on. There is a range of object detection methods based on the background subtraction technique. Object detection using this technique involves the subtraction of two images, the current image and the image representing the static part(s) of the scene Viola et al. (2005). One of the major problems with this technique is how to automatically obtain a background of the static scene that is as robust as possible to changes in lighting, shadows, and noise present in the video sequence.

4.1. STUDIED TECHNIQUE

The reference image can be the average of the N previous images in the sequence (publicly available). To calculate this image, we create an array of buf images into which we copy these N images. This array has a fixed size, contains N monochrome images and the i^{th} buf[i] image can be manipulated like any other image Zhao et al. (2006).

To read or write these images, we declare an integer last indexing time t: buf[last] is thus the current image. Once the first N images of the buf array have been stored (buf[i] for i ranging from 0 to N-1), the N+Ist image of the sequence is stored at the beginning of the array, overwriting the old value of buf[0], the next image at t=N+2 overwrites buf[1] etc......The image of the sequence is therefore stored at position last +1 modulo N of the buf buffer. The sum of the N cvAcc images is calculated first, which requires a 32-bit coded image to store this intermediate result. This image is then divided by the integer N, and converted into an 8-bit coded image Zhao et al. (2006).

4.2. IMPLEMENTATION DIAGRAM

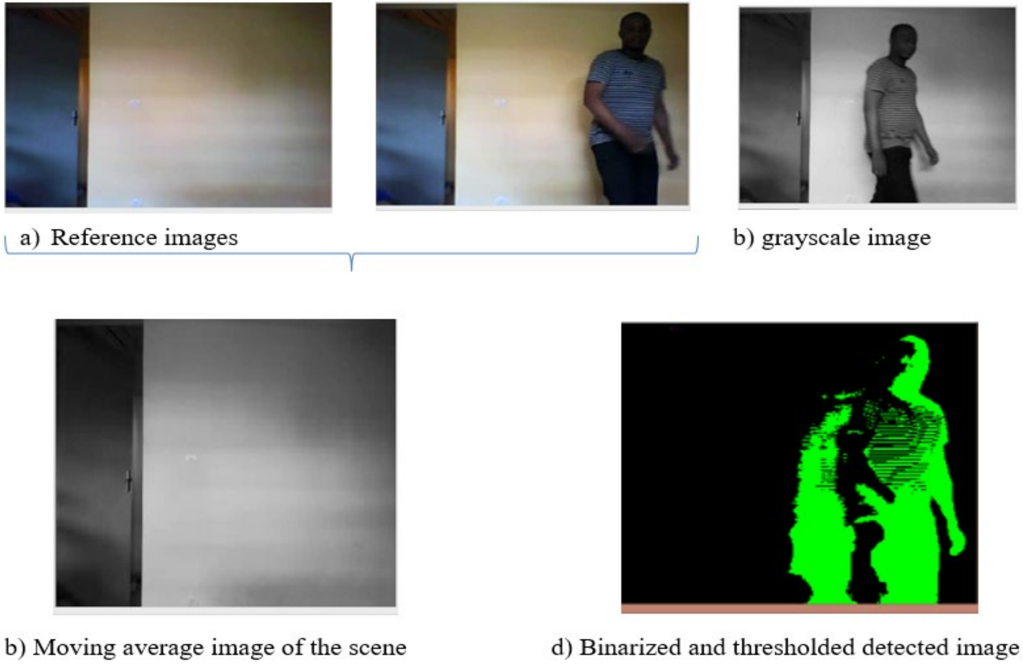
Modeling the background of a scene using the method proposed by [15] has become an increasingly popular reference in the field of moving object detection. This method is robust and effective in a large number of use cases, including dynamic backgrounds (such as foliage, fountains, seashores, flags...) and slight changes in illumination Kim et al. (2005).

The first observation of this algorithm is that false detections are generally located in dark areas of the image. Dark (and therefore less luminous) colors are inherently more difficult to differentiate, leading to greater uncertainty as to their final classification (background or shape). We therefore assume that brightness should be an important factor when comparing color ratios between two pixels Sigal and Athitsos (2000).

5. RESULTS AND DISCUSSION

ver time, we also observe that the color of a given pixel is distributed on a straight line aligned with the origin (0,0,0). These observations motivate the creation of a new color model enabling separate evaluation of color distortion and pixel brightness.

Figure 1



a) Reference images

b) grayscale image

b) Moving average image of the scene

d) Binarized and thresholded detected image

Figure 1 Motion Detection by Mobile Means

5.1. DISUCSSION

The images obtained from the results of the moving average detection method give us five images, including:

Image a) is made up of two images considered as reference images at two different times: at time t, when no moving person is detected, and at time t+1, when a moving person is detected in the scene. The difference between these instants will enable us to average the image acquired later between the image at time t and that at time t+1.

Image b) is the grayscale-transformed image. The purpose of grayscale image transformation is to convert color images into binary images where luminance and chrominance are zero.

Image c) shows the average of the difference between the reference image at time t and that at time t+1. Our objective here is to see nothing on this image, whereas in reality we have detection activities. By averaging the images at time t and time t+1, we call up a specific function in the OpenCv library to cancel out the set of moving pixels, so we'll have an image with no mouth; called the image of the image average.

Image d) is obtained after detection of the image mean, as shown in image c), the result is binarized and thresholded. We can see the person after detection, binarization and thresholding, but with a spot that does not allow us to specify the person as in the other cases.

On the other hand, the presence of a shadow can lead to false detection in certain cases: the presence of a shadow is considered as a second person. For good vision, we need to eliminate the presence of the shadow using the manual thresholding method, while adjusting the desired thresholding level as a function of the cursor.

5.2. ELIMINATING SHADOWS

Shadows can sometimes be generated by the effect of light sources on moving objects, as shown in the images from our various experiments. These special cases can produce inconveniences, for example when matching parts between two regions, as shown in **images c)** and **d)** of figure 1. However, shadows are not too disturbing in a face recognition context, as they can be easily eliminated by a masking operation.

On the other hand, whole-body methods of individual detection will suffer enormously from these side effects. Some techniques, however, can solve most or all of this problem. These include the Gomez and Morales (2002) method discussed above in Mikiæ et al. (2000) and that of Benezeth et al. (2008) which uses the color information obtained from HSV to eliminate shadows. In principle, a shaded background should have the same color at lower brightness.

Apart from the undeniable advantages of eliminating shadows, it should not be forgotten that false detections can occur during this additional phase. However, in the current project, it's better to remove unnecessary areas than to see false detection, which will multiply the number of false alarms.

6. CONCLUSIONS AND RECOMMENDATIONS

In this work, we focus on one of the major problems of motion analysis in an image sequence: the detection of moving objects in simple scenes. This method can efficiently detect motion in complex scenes, taking into account variations in color, brightness and orientation. It can be used for surveillance applications requiring robust and accurate motion detection.

After reviewing the literature, we had to justify the choice of a background subtraction detection model, the one selected. An implementation of this model was carried out on various video sequences. In the case of fast motion, the background subtraction method is recommended. It should be noted that, for greater efficiency, two or more methods could be used in conjunction with image segmentation methods, which have not been used in the present work.

Remember that this work is the first phase of a project to recognize suspicious moving objects in an image. This goal opens up the following perspectives:

- Search for a single method for the detection of moving objects for both types of motion (slow and fast).
- Applying detection methods to sequences with dynamic backgrounds;
- Study object motion tracking in video sequences;
- Study object recognition after detection and tracking.
- Applying computer science to image matching, which is a key stage in many computer vision and photogrammetry techniques, whether for image registration, position estimation, shape recognition, image indexing or three-dimensional recognition, in terrestrial, aerial or satellite imagery.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Benabbas Siavosh (2012). Département d'Informatique, & Université de Toronto, Efficace. Sum-Based Hiérarchique Lissage Sous \ ell_1-Norm.

Benezeth, Y., Jodoin, P. M., Emile, B., Laurent, H., & Rosenberger, C. (2008). Review and Evaluation of Commonly Implemented Background Subtraction Algorithms. In Proceedings of the International Conference on Pattern Recognition, 1–4. https://doi.org/10.1109/ICPR.2008.4760998

Cucchiara, R., Grana, C., & Piccardi, M. (2001). Andrea Pratti et Stefano Sirotti: Improving Shadow Suppression in Moving Object Detection with HSV Color Information. Dans Intelligent Transportation Systems. IEEE Publications, 334–339.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition, 886–893. https://doi.org/10.1109/CVPR.2005.177

Gobeau, J. F. (2007). Détecteurs de Mouvement à Infrarouge Passif (Dé- Tecteurs IRP). Colloque Capteurs.

Gomez, G., & Morales, E. (2002). Automatic Feature Reconstruction and a Simple Rule induction Algorithm for Skin Detection. Dans ICML Workshop On Machine Learning in Computer Vision, 31–38.

H. Nicolas et C. Labit (1993). Motion and Illumination Variation Estimation Using a Hierarchy of Models: Application to Image Sequence Coding, Publication Interne., 742, IRISA.

Kim, K., Chalidabhongse, T. H., Harwood, D., & Davis, L. (2005). Real-time Foreground-Background Segmentation Using Codebook Model. Real-Time Ima- ging, 11, 172185. https://doi.org/10.1016/j.rti.2004.12.004

Kovac, J., & Solina, P. F. (2003). Human Skin Color Clustering for Face Detection. EUROCON 2003 Computer as a tool, 2, 144–148. https://doi.org/10.1109/EURCON.2003.1248169

Medjahed, F. (2012). Détection et Suivi d'objets en Mouvement Dans une Séquence d'images, Mémoire Magister, Université Mohamed Boudiaf faculté de génie Electrique.

Mikiæ, I., Cosman, P. C., & Greg, T. (2000). Kogut et Mohan M. Trivedi: Moving Shadow and Object Detection in Traffic Scenes. Dans International Conference on Pattern Recognition (ICPR), 321–324. \

Sigal, L., & Athitsos, S. V. (2000). Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination. Dans IEEE Conference on Computer Vision and Pattern Recognition, 2, 152–159.

Viola, P., Jones, M. J., & Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. In International Journal of Computer Vision, 63(2), 153–161. https://doi.org/10.1007/s11263-005-6644-8

Wu, B., & Nevatia, R. (2006). Tracking of Multiple, Partially Occluded Humans Based on Static Body Part Detection. In IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition, 951–958.

Zhao, Q., Kang, J., Tao, H., & Hua, W. (2006). Part Based Human Tracking in A Multiple Cues Fusion Framework. In Proceedings of the International Conference on Pattern Recognition, 450–455.