INDIAN INSTITUTE OF TECHNOLOGY, HYDERABAD

Hierarchical Background Subtraction Algorithm For Foreground Background Segmentation

by

Vidhi Rani Kanesh

A thesis submitted in partial fulfillment for the
degree of Master of Technology

in
Computer Science Engineering

under guidance of
Dr. C. Krishna Mohan

June 2016

Declaration

I, Vidhi Rani Kanesh, declare that this thesis titled, "Hierarchical Background Subtraction Algorithm For Foreground Background Segmentation" and the work presented in it are written by my own. I confirm that this work was done wholly or mainly while in candidature for a thesis research. Any part of this thesis has not previously been submitted for any degree or any other qualification at this University or any other institution, this has been clearly stated. I have consulted the published work of others, this is always clearly attributed. I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work. I have acknowledged all main sources of help where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Vidhi

Signature

VIDHI RANI KANESH

Student Name

CS13M1010

Roll Number

Approval

This Thesis entitled "Hierarchical Background Subtraction Algorithm For Foreground Background Segmentation" by Vidhi Rani Kanesh is approved for the degree of Master of Technology from IIT Hyderabad.

L. B. Venu...

Examiner
Dept. of CSE
IITH

R Singh

Examiner
Dept. of CSE
IITH

C. Krishna Mohan

(Dr. C. Krishna Mohan) Adviser
Dept. of CSE
IITH

U. Ramesh

Chairman
Dept. of CSE
IITH

Abstract

Segmentation of moving objects from a video sequence is one of the fundamental and critical task in automated video surveillance, and it gets challenging when the background is non-stationary. Background subtraction is a widely used algorithm for real-time segmentation of moving objects in the presence of the static camera. Various pixel-based or block-based background subtraction approaches are available in the literature. Pixel-based methods generate smooth contours of foreground objects while block-based methods are more robust to the dynamic background. In this dissertation, we propose to combine both block-based and pixel-based background subtraction techniques in a hierarchical manner for better foreground detection. Motion segmentation problems such as dynamic background, illumination variations, and noise are addressed at block-level. Then a fine level processing of the foreground regions is done on the foreground background model obtained at block-level processing to smoothen the foreground objects.

Acknowledgements

I express my deepest gratitude to all the people who have supported me and encouraged me during the course of this project. I am very grateful to my thesis advisor Dr. C. Krishna Mohan for giving me the opportunity of carrying out this research under their guidance. I am very deeply indebted to him for his support, advice and encouragement without which this research could not have proceeded smoothly. I would also like to thank my family members and my friends for their continuous motivation and encouragement.

Contents

List of Figures

Chapter 1

Introduction

Segmentation of moving objects from a video sequence or image sequence is one of the fundamental and critical task in automated video surveillance. Therefore its performance can have a great impact on the performance of applications employing it since they only consider foreground objects of interest. Motion segmentation is employed in various applications such as object detection, object tracking, action recognition, behavioural analysis, personal identification, etc.

The conventional approaches for motion segmentation are temporal differencing, optical flow and background subtraction[2].

In temporal differencing the difference between two or three consecutive frames is calculated to segment foreground and background. These methods are fast and adaptive to dynamic environments with sudden illumination change but does a very poor job of extracting the foreground objects, e.g. there may be holes left in the foreground region(ghosts).

Motion segmentation methods based on optical flow uses the characteristics of flow vectors of objects in motion over time to detect foreground regions in an image sequence. Methods based on optical flow rely on continuous movement of objects and can not detect stationary objects. These methods can not be implemented for real time motion segmentation since they are computationally complex and need special hardware support.

The most common method for real-time segmentation of moving objects is background subtraction. Background subtraction detects the foreground by thresholding the difference between the current video frame and the referenced background.The implementation of these methods is computationally affordable and the accuracy of detection of moving objects is good.

Segmention of the foreground objects of interest from a video sequence is difficult because of some challenging problems such as dynamic background(e.g. swaying trees, moving water, waves, and rain, etc.), illumination changes, noise, shadows, occlusion, etc. To overcome these problems many background subtraction approaches have been proposed, such as a mixture of Gaussians[3–5], codebook[6–10], kernel density estimation, edge histogram[11], normalised vector distances[12], Kalman filtering, etc.

In general, the foreground/background regions are selected in one of the following two ways : (1) pixel-by-pixel basis or (2) block-by-block(region/patch) basis. In proposed method, we perform background subtraction at both block-level and pixel-level. We have discussed various pixel-based and block-based background subtraction methods in section 2.

Either the frames are processed at the pixel level or foreground level, we need efficient texture descriptors. Various kinds of texture descriptors are local binary patterns(LBP) [13–15], block truncation coding, discrete cosine transform(DCT)[16][17], Hadamard transform[18–20], Normalised vector distance[12], Gabor filters[21, 22], etc. In proposed method DCT is used as block descriptor, due to its decorrelation properties[23] and ease of implementation in hardware.

The remaining sections of this dissertation are organized as follows : related work is presented in section 2, section 3 presents the problem definition, section 4 is about the proposed algorithm and section 5 is conclusion.

Chapter 2

Related Work

In this section various background subtraction methods based on pixel-by-pixel and block-by-block processing, proposed by different authors are discussed. Some of the methods based on hybrid level processing are also discussed.

2.1 Pixel-Based Methods

The earlier approaches on background subtraction were pixel-based and considered every pixel to be independent. For pixel-level processing, the simplest techniques are based on applying a median filter on pixels at each location across all the frames. The main disadvantage of this method is the high computational cost of sorting the pixels. Stauffer et al. [3] used a mixture of Gaussians to construct a background model for each pixel, and it achieves good performance using learning procedure and builds a statistical model. However, it cannot detect and remove the shadow, noise and does not consider the spatial correlation of pixels.

The method proposed by Kim et al. [9] created codebooks for each pixel. It does not deal with dynamic backgrounds since it does not consider the spatial correlation of pixels. To overcome this problem, Chih et al. [10] proposed a background subtraction based on single layer codebook model. It has two components background extractor(BE) which is calculated using single layer codebook model, for each frame and spatial information is propagated from neighbouring pixels which in turn handles dynamic background and illumination problem. Then background gradient extractor(BGE) is calculated to ensure completeness of foreground objects.

To handle varying illumination, Song et al. [21] proposed a phase based background extraction. Phase features of pixels are extracted using Gabor filter. Deepak and Sukadev [22] also used Gabor filter to handle illumination problem.

2.2 Block-Based Methods

Zhou et al. [4] proposed a method in which regions are modeled as the mixture of Gaussians. It can handle dynamic backgrounds subject to gradual lighting changes. To improve the convergence of mixture of gaussian distributions, Zhou et al. in [5] introduced the additional term to update equation called as momentum.

Davide et al. [20], used Hadamard transformed matrix as block descriptor. It is used for background initialization in the static background environment.

Reddy et al. [17] proposed a background estimation method on a block-by-block basis. For each block location, a representative set is maintained of unique patches obtained along its temporal line. The background candidates are decided based on some similarity measure. 4-connecting neighbours are used to estimating the background at an empty block location and the blocks are compared in a frequency domain. For decomposition of blocks, 2D DCT is used. This method is used for static background and can handle varying illumination.

In the method proposed by Le et al. [24], the image is pre-segmented into regions. Then within each region, patches are sampled into two bags of patches; which are used for foreground segmentation.The algorithm is conceptually simple but this method suffers from issues such as supervised dimension reduction and density modeling techniques on image patch sets, optimal random patch sampling strategy, and self-tuned optimal image patch size searching.

2.3 Hybrid Methods

A few methods have been proposed which work at both pixel and block level. Various authors like Cai et al. [6], Jing et al. [7, 8], Kim et al. [9] have proposed codebook based background subtraction models; which works at both block and pixel-level. These proposed methods can effectively deal with dynamic backgrounds, shadows, and highlights.

In the method proposed by Omar et al. [25], the image is divided into blocks, then using T frames of the video sequence, blocks of sequence are combined to form a matrix for each block location.After computing the matrices of blocks, within each matrix, the pdf of pixels intensity is computed. The summation of pdfs in each block is calculated. The blocks having highest values are classified as background. These blocks are used to model the background. They have applied structure-texture decomposition on the absolute difference image to minimize the noise in the results of the background subtraction. The structure component which contains just the homogeneous parts of the image is used in the segmentation process. A binary image based on some threshold is used to detect moving objects.

In method proposed by Lee et al. [1], the image is divided into blocks of size 8x8 and block sampling is performed. After the sampling of blocks, pixels of the corresponding blocks are sampled. The MOG-based classification scheme is used to classify a sampled pixel as foreground or background. If the sampled pixel is classified as the foreground pixel, the corresponding block is processed by silhouette detection scheme to refine the foreground model.

Chapter 3

Problem Definition

Pixel based foreground detection methods are more susceptible to illumination changes, noise, and dynamic background. In static background environment, pixel-based methods work well when objects are moving continuously but are likely to fail when the foreground objects are exposed for more time than the background itself. The region based background subtraction methods can handle above-mentioned problems better than pixel-based background subtraction methods. The problem is when motion segmentation is done at a block level, the estimated foreground region is not smooth and suffers from false negative. On contrast the pixel-based background subtraction provides more accurate and precise detection of foreground objects. So in order to alleviate the limitations of a pixel and block based methods and get the benefit of both pixel-wise and block-wise background subtraction methods to perform better foreground detection, we have combined pixel and block-based approaches in a hierarchical manner. In proposed approach first, the foreground background model is generated at a block level. Then it is refined at the pixel level to overcome the low precision problem at the block level.

Chapter 4

Proposed Background Subtraction Algorithm

The steps of the algorithm are as follows

1. The frame is first divided into overlapping blocks of size $N \times N$.
2. A low dimensional descriptor is calculated using DCT.
3. A multivariate Gaussian distribution is calculated to classify the blocks as foreground and background.
4. A cosine similarity measurement is applied on blocks not classified as background and again the blocks are classified as foreground and background.
5. If most of the blocks are classified as foreground, the model is reinitialized.
6. Pixel level detection of foreground objects.

Let the resolution of the image is $W \times H$. The image is divided into overlapping blocks of size $N \times N$. Let the block at location (x, y) be denoted as $b(x, y)$ and pixel at location (i, j) as $p(i, j)$. Since the model contains both the foreground and background blocks, the model is prone to misclassification. To tackle this problem blocks are overlapped.

4.1 Calculation of block description using DCT

Discrete Cosine Transform is used to calculate the low dimensional descriptor of the block. Let the DCT block descriptor of a block $b(x, y)$ is denoted by $d(x, y)$. Discrete cosine transformation transforms an image from it's spatial domain to frequency domain. It can separate the image into high, middle and low-frequency components[16]. Let $b(u,v)_R$, $b(u,v)_G$ and $b(u,v)_B$ are the corresponding R, G, B matrix components of the blocks of size $N \times N$ of source image. Then they are passed through DCT technique to generate the DC coefficients of the blocks of image I and the values are stored in a matrix of size $m = ([N \times N]_R + [N \times N]_G + [N \times N]_B)$. The general equation to calculate 2D DCT of a block at location (u,v) is defined as :

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) cos \left[\frac{(2x+1)u\pi}{2N} \right] cos \left[\frac{(2y+2)v\pi}{2N} \right] \tag{4.1}$$

where $C(u,v)$ is DC coefficient of block $b(u,v)$.

4.2 Classification of blocks

4.2.1 Based on Multivariate Gaussian Distribution

The probability that a block descriptor $d(x,y)$ will be classified as foreground or background can be calculated as :

$$P(d(x,y)) = \frac{1}{(2\pi)^{\frac{m}{2}} \mid \Sigma(x,y) \mid^{\frac{1}{2}}} exp\{ \frac{-1}{2} \left[d(x,y) - \mu(x,y) \right]^T \Sigma(x,y)^{-1} \left[d(x,y) - \mu(x,y) \right] \} \tag{4.2}$$

where $\mu(x,y)$ is the mean vector and $\Sigma(x,y)$ is the covariance matrix for block location (x,y). $\mu(x,y)$ and $\Sigma(x,y)$ are calculated using small training frame sequence which is the first few frames of the sequence. If the probability is greater than some threshold let say G_{th} then the corresponding block is classified as background.

$$P(d(x,y)) \geq G_{th} \tag{4.3}$$

If the block is classified as background, we update the gaussians as

$$\mu(x,y)^{new} = (1-\rho)\mu(x,y)^{old} + \rho d(x,y) \tag{4.4}$$

$$\Sigma(x,y)^{new} = (1-\rho)\Sigma(x,y)^{old} + \rho(d(x,y) - \mu(x,y)^{new})(d(x,y) - \mu(x,y)^{new})^T \quad (4.5)$$

4.2.2 Based on Cosine Similarity Measure

If a block is not classified as background, we calculate the cosine similarity measure of the corresponding block descriptor. Empirical observations show that the angles subtended by block descriptors exposed to illumination changes are almost the same(i.e. The vector distance or cosine distance is not affected by varying illumination[12]).

Cosine similarity is calculated over block descriptor and the mean of corresponding block descriptor as :

$$cosinedist(d(x,y), \mu(x,y)) = 1 - \frac{(d(x,y)^T \times \mu(x,y))}{\|d(x,y)\| \|\mu(x,y)\|} \quad (4.6)$$

if $cosinedist(d(x,y), \mu(x,y)) \leq C$, block is classified as background, where C is threshold and it's value is kept low.

4.3 Reinitialization of Foreground Background Model

Because of quick illumination changes most of the blocks might get classified as foreground even if they belong to the background. To address this problem we perform foreground-background model initialization. We calculate the estimation of a portion of image classified as foreground based on the size of the moving objects and the dynamics of the background . If this estimation increases over some threshold value we perform model reinitialization. So we again repeat steps from 3 to 5. During reinitialization, to calculate Gaussian distribution we keep the covariance matrix while calculating new means since we have small training frame sequence.

4.4 Pixel-level Detection of Foreground Objects

Let the $B_{fg}(i,j)$ be the number of blocks containing pixel $p(i,j)$ which were classified as foreground(Fg) and $B_t(i,j)$ be the total number of blocks containing pixel $p(i,j)$. Now we calculate the likelihood that pixel $p(i,j)$ will belong to foreground as :

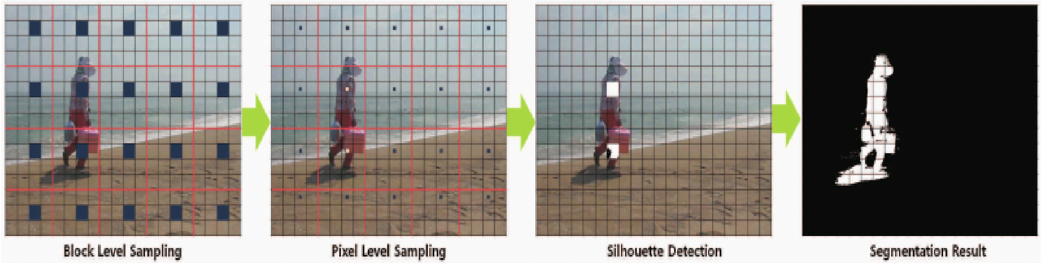

Figure 4.1: Illustration of the proposed algorithm by Lee et al. [1]. Sampling is performed at the block level and at the pixel level. Then, around each sampled foreground pixel, the object silhouette is refined to obtain an accurate segmentation result.

$$P(Fg \mid p(i,j)) = \frac{B_{fg}(i,j)}{B_t(i,j)} \tag{4.7}$$

If $P(Fg \mid p(i,j)) \geq th$, where th is threshold, then pixel $p(i,j)$ is classified as part of foreground.

Chapter 5

Conclusion

We have proposed a background subtraction technique based on both block-level and pixel-level processing for foreground and background segmentation of video frames. At the block level, multivariate Gaussian distribution is used for dynamic backgrounds, while to tackle illumination variations problem cosine similarity measure is used. A low-dimensional block descriptor is used to alleviate the image noise effect. Obtained foreground and background model is processed at pixel level to smoothen the foreground regions.

Bibliography

[1] Jae-Kyun Ahn Dae-Youn Lee and Chang-Su Kim. Fast background subtraction algorithm using two-level sampling and silhouette detection. In *IEEE*, 2009.

[2] Liang Wang Weiming Hu, Tieniu Tan and Steve Maybank. A survey on visual surveillance of object motion and behaviours. In *IEEE Transactions on Systems, Man, and Cybernatics - Part C: Applications and Reviews*, volume 34, August 2004.

[3] W.E.L Grimson Chris Stauffer. Adaptive background mixture models for real-time tracking. In *Proc. IEEE Conf. Computer Vision And Pattern Recognition*, volume 2. IEEE, 1999.

[4] Paul Miller Sriram Varadarajan and Huiyu Zhou. Spatial mixture of gaussians for dynamic background modelling. In *10th IEEE International Conference On Advanced Video And Signal Based Surveillance*. IEEE, 2013.

[5] Paul Miller Sriram Varadarajan, Hongbin Wang and Huiyu Zhou. Regularised region-based mixture of gaussians for dynamic background modelling. In *11th IEEE International Conference On Advanced Video And Signal Based Surveillance (AVSS)*. IEEE, 2014.

[6] Anni Cai Pengxiang Zhao, Yanyun Zhao. Hierarchical codebook background model using haar-like features. In *Proceedings of IC-NIDC*. IEEE, 2012.

[7] Min-Hsiung Shih Yun-Fu Liu Jing-Ming Guo, Chih-Hsien Hsia and Jing-Yu Wu. High-speed multi-layer background subtraction. In *IEEE International Symposium On Intelligent Signal Processing And Communication Systems (ISPACS)*, pages 4–7, November 2012.

[8] Yun-Fu Liu Min-Hsiung Shih Cheng-Hsin Chang Jing-Ming Guo, Chih-Hsien Hsia and Jing-Yu Wu. Fast background subtraction based on a multilayer codebook model for moving object detection. In *IEEE Transactions On Circuits And Systems For Video Technology*, volume 23, October 2013.

[9] D. Harwood K. Kim, T. H. Chalidabhongse and L. Davis. Real-time foreground-background segmentation using codebook model. In *Real-Time Imaging*, volume 11, pages 172–185, June 2005.

[10] Chu-Song Chen Chih-Wei Lin, Wei-Jie Liao and Yi-Ping Hung. A spatiotemporal background extractor using a single-layer codebook model. In *11th IEEE International Conference On Advanced Video And Signal Based Surveillance (AVSS)*, 2014.

[11] M. Mason and Z. Duric. Using histograms to detect and track objects in color video. In *IEEE*, 2001.

[12] Hitoshi Habe Takashi Matsuyama, Toshikazu Wada and Kazuya Tanahashi. Background subtraction under varying illumination. In *Systems and Computers in Japan*, volume 37, 2006.

[13] M. Heikkil A and M. Pietik Ainen. A texture-based method for modeling the background and detecting moving objects. In *IEEE Trans. Pattern Anal. Mach. Intell.*, volume 28, page 657662, 2006.

[14] Shaohui Liu Shengping Zhang, Hongxun Yao. Dynamic background modeling and subtraction using spatio-temporal local binary patterns. In *IEEE ICIP*, 2008.

[15] Shaohui Liu Shengping Zhang, Hongxun Yao. Dynamic background modeling and subtraction using spatio-temporal local binary patterns. In *IEEE*, 2008.

[16] J. K. Mandal Madhumita Sengupta. Self-authentication of color images through discrete cosine transformation (sadct). In *IEEE International Conference On Recent Trends In Information Technology, ICRTIT*, 2011.

[17] C. Sanderson V. Reddy and B. Lovell. An efficient and robust sequential algorithm for background estimation in video surveillance. In *Proc. Of IEEE Intl Conference On Image Processing*, 2009.

[18] A. Vafaei M. J. Nassiri and A. Monadjemi. Texture feature extraction using slant-hadamard transform. In *International Journal Of Applied Science, Engineering And Technology*, 2007.

[19] Nang Aye Aye Htwe Khin Thida Win. Image compression based on modified walsh-hadamard transform (mwht). In *Proceedings Of 3rd Iserd International Conference*, Singapore, May 2015.

[20] Rita Cucchiara Davide Baltieri, Roberto Vezzani. Fast background initialization with recursive hadamard transform. In *Seventh IEEE International Conference On Advanced Video And Signal Based Surveillance*, 2010.

[21] J. Sun G. Xue and L. Song. Background subtraction based on phase and distance transform under sudden illumination change. In *Proc. IEEE International Conference Image Process*, Sept. 2010.

[22] Sukadev Meher Deepak Kumar Panda. Hierarchical background subtraction algorithm using gabor filter. In *IEEE*, 2015.

[23] C. Sanderson and K.K. Paliwal. Polynomial features for robust face authentication. In *IEEE International Conference On Image Processing (ICIP)*, volume 3, page 9971000, 2002.

[24] Gregory Hager Le Lu. Dynamic foreground/background extraction from images and videos using random patches. In *Advances in Neural Information Processing Systems 19*, pages 929–936, 2007.

[25] Soukaina Elidrissi Elkaitouni Omar Elharrouss, Driss Moujahid and Hamid Tairi. An effective foreground detection approach using a block-based background modeling. In *13th International Conference Computer Graphics, Imaging And Visualization*, 2016.