Gait Silhouette Extraction from Videos Containing Illumination Variates

Amalina Ibrahim, Wan-Noorshahida Mohd-Isa and Chiung-Ching Ho

Abstract We present a heuristic method to automatically adjust pixel intensity per frame from video by analyzing its colour type and level of brightness before initiating silhouette extraction phase. As this is performed at the pre-processing phase, our proposed method aims to show that it is an improvement or solution for videos containing inconsistency of illumination compared to normal background subtraction. We are introducing two modules; a prior processing module and an illumination modeling module. The prior processing module consists of resizing and smoothing operations on related frame in order to accommodate the subsequent module. The illumination modeling module manipulates pixel values in each frame to improve silhouette extraction for a video containing inconsistency of illumination. This proposed method is tested on 1072 videos including videos from an external KTH database.

Keywords Pixel value · Background subtraction · Gait recognition · Illumination · Brightness · Contrast

1 Introduction

Varying illumination poses a challenge to videos shot in an outdoor scene. Such videos are elements from a typical camera setting in an outdoor surveillance system. In such setting, the normal background subtraction may not be that robust when performing segmentation and extraction of foreground objects from its background.

A. Ibrahim (✉) · W.-N. Mohd-Isa (✉) · C.-C. Ho (✉)
Faculty of Computing & Informatics, Multimedia University,
63100 Cyberjaya, Malaysia
e-mail: amalina_ibrahim@hotmail.com

W.-N. Mohd-Isa
e-mail: noorsha@mmu.edu.my

C.-C. Ho
e-mail: ccho@mmu.edu.my

This is because the background illumination keeps changing with the change in light intensity; thus requiring an update of the background information at some time intervals. This has led to the motivation for developing an adaptive algorithm that can perform robust segmentation and extraction of foreground objects for outdoor videos. In this paper, the foreground object is a person where his/her collection of silhouette that describes his/her gait is what we are interested in.

In machine vision (MV) gait recognition research, such varying illumination has become the kryptonite of gait video data acquisition. This type of illumination is said to be uncooperative and against not just foreground but also background segmentation. Thus, it threatens object (person) extraction process as the object and background are unable to be distinguished. Another approach is to model the background (or background modeling) that looks at using pixel-level light intensity information such as Mixture of Gaussians (MOG). This method has been shown to robustly model complicated backgrounds, especially those with small repetitive movements such as swaying tree, ocean waves and pouring rain [1]. There are now MOG2 and MOG3 variants. These MOGs background model has become very popular because of its efficiency in modeling multi-intensity background distribution, its ability to adapt to changes of background information, and the potential for real-time implementation [1]. In addition, it selects the appropriate number of Gaussian distribution for each pixel and provides better adaptability to varying scenes due to illumination changes [2]. In this paper, we are proposing an algorithm that uses an automated process of pixel value manipulation technique founded on original frame brightness. This approach is followed by applying MOGs as background modeling medium. Our proposed approaches will be performed during preprocessing phase in the aim to extract accurate gait silhouettes from videos with illumination variates.

2 Related Work

Wang, Shin, and Klette have proposed a shadow removal algorithm in the context of silhouette detection for moving people in gray level video sequences. The coloured video is converted to a shadow of gray sequence of frames during preprocessing in order to overcome illumination varying problem [3]. While Cheng et al. work uses binarization method to work well under various kinds of illumination conditions whereby a dynamically adjusted cumulative histogram is computed to find the most suitable threshold for every pixel in the target image [4]. Due to the extraction of binarized silhouettes, in essence, only edge information at the boundary is captured. Any edges or gradients inside the silhouette are completely discarded [5].

Exploiting gradient histograms for person identification based on gait is another approach to improve silhouette extraction accuracy [5, 6]. This method also captures gradients within the person's silhouette. This work has improved the perspective correction technique to normalize oblique-view walking sequences to

side-view plane. The silhouettes from oblique-view walking sequences are vertical and horizontal adjusted to fit the side-view [7]. Unfortunately, the work mentioned here are conducted under controlled environment or laboratory environment and uses surveillance videos that are of grayscale type. Thus, such approaches may affect person segmentation and feature extraction process when applied to videos under varying lighting environment. Tan et al. [8] might have found a solution for this problem. Regrettably, the work focusses on face recognition and may not work properly for gait recognition [8].

3 Methodology

Our algorithms are implemented in OpenCV and we adopt its built-in library background_segm and built-in MOG2 foreground mask. We are proposing two processes; a Prior Processing Module and an Illumination Modeling Module for silhouette segmentation as shown in Fig. 1.

We start with a Prior Processing Module that consists of frame resizing process to reduce storage usage since the later process of classification normally will require a huge storage space. We are using bilinear interpolation because of its simplicity. After that, the resized frame undergoes smoothing process for minor noise reduction. Resizing is always initiated first because extracted features are affected by the size of the image (in pixels).

After the frame is being resized. Then the frame type is identified, whether it is a coloured or grayscale frame. After that, the level of brightness of frame is identified. Three types of brightness level are proposed in this work: bright, normal and dark. Then the resized frame will go through the silhouette extraction, feature extraction and classification. However the results of the final two processes that include performance measure are not covered in this paper.

3.1 Illumination Manipulation

In the Illumination Modeling Module, we manipulate pixel values in each frame to increase accuracy of extracting potential features. In other words, it is an adjustment of brightness and contrast on sequence of frames. This technique is proven effective in enhancing details. This technique has been extensively applied [9–11]. The process involves identifying the values of ksize_, α and β that represent variables
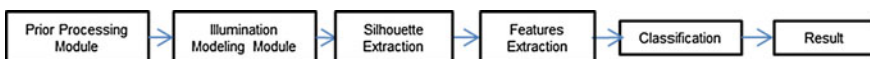

Fig. 1 Process flow of our proposed method

for the operation of smoothing, contrast and brightness, respectively to produce the best silhouette extraction.

The value of ksize_ can be found using the Normalized Box Filter as in Eq. 1. The value of ksize_ will be manipulated based on frame brightness or pixels value intensity. The Normalized Box Filter is a kernel to common type of filter that is found by a weighted sum of input pixel values ($f(i+k, j+1)$) as shown in Eq. 2.

$$h(k,l) = \frac{1}{\text{ksize\_width} * \text{ksize\_height}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ & & & & & \end{bmatrix} \tag{1}$$

$$g(i,j) = \sum_{k,l} f(i+k, j+l) h(k,l) \tag{2}$$

with $g(i,j)$ as an output pixel value, i and j indicates that the pixel is located in the ith row and jth column and $h(k,l)$ are the coefficients of the filter.

The values of α and β are found by two commonly used point processes of multiplication and addition with a constant as shown in Eq. 3 below,

$$g(i,j) = \alpha f(i,j) + \beta \tag{3}$$

where, brightness is represented by β and contrast by α. The parameters α and β are often called the gain and bias parameters.

The values of ksize_, α and β are determined based on categorizing frame brightness from a data driven process. This is done by counting and organizing the pixel values from MOG into a set of predefined bins. Equation 4 shows our bin size = 15 from our gathered data:

$$\begin{aligned} \text{range} &= \text{bin}_1 \cup \text{bin}_2 \cup \ldots \cup \text{bin}_{n=15} \\ [0, 255] &= [0, 15] \cup [16, 31] \cup \ldots \cup [240, 255] \end{aligned} \tag{4}$$

As illustrated in Fig. 2a, bin_1 contains total amount of pixels value in range of 0–15, bin_2 is from 16 to 31 and continue until bin_{15} counts pixel that has value from 240 to 255. It has been found that dark type of frames, region A are from bin_1 until bin_7 that covers range of pixel value from 0 to 127. Normal brightness frames, region B are from bin_8 until bin_{10}, range from 128 to 175 while bright frames, region C are from pixel value 176–255, bin_{11} until bin_{15}. Figure 2b displays the same comparison but for shades of gray type of frame. However the boundaries are different from coloured frames whereby dark frames are from bin_1 until bin_9, normal frames are from bin_{10} to bin_{12} and remaining bins are for bright frames.

From our experiment, we found that the best value of the related variables based on brightness on coloured frames are as in Table 1. From Table 1, the values are null for the normal coloured frames because the extracted silhouettes from our
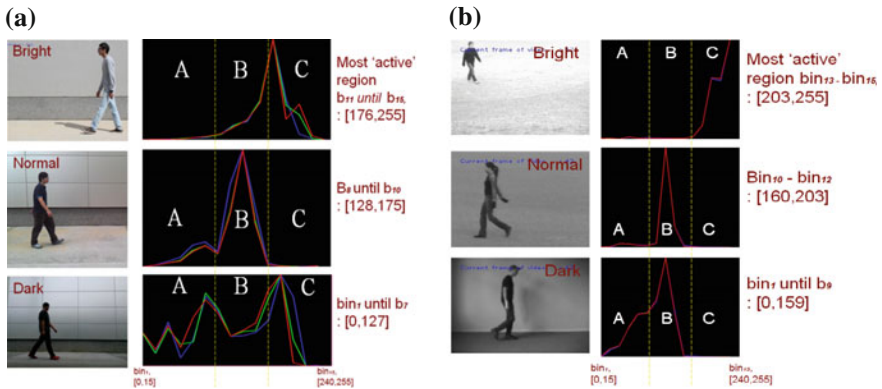
Fig. 2 Graphical comparison of number of pixel versus bin for 3 types of frame brightness. **a** Coloured. **b** Grayscale

Table 1 Identified values of manipulator based on brightness of coloured frame for illumination modeling module

Type of frame brightness	Coloured				Grayscale			
	Bin_n	ksize	α	β	Bin_n	ksize	α	β
Bright	11–15	3	2	−150	13–15	1.5	2	−300
Normal	8–10	Null	Null	Null	10–12	1.5	3	−200
Dark	1–7	1.5	3	−50	1–9	Null	Null	Null

proposed method are no different from the common background subtraction. Similarly, for the dark type of grayscale frame, the variables are null because the extracted silhouettes from our proposed method are no different from the common background subtraction.

4　Results

The proposed method is tested on coloured sequences taken in daylight outdoor environment with 343 different people as our subjects. There are 972 videos recorded on different days and venues. For these videos, each subject walked a fixed straight path, approximately two and a half gait cycle long and back-and-forth at their natural pace. Also, our method is tested on KTH Database of 100 videos. This is to test our method on data collected from surveillance camera. The KTH videos are of grayscale type. The following are the results from several selected frames from different types of brightness and colour on one gait cycle. For detailed results, readers can visit the following website: http://amalinaibrahim5.wix.com/mmue130146.
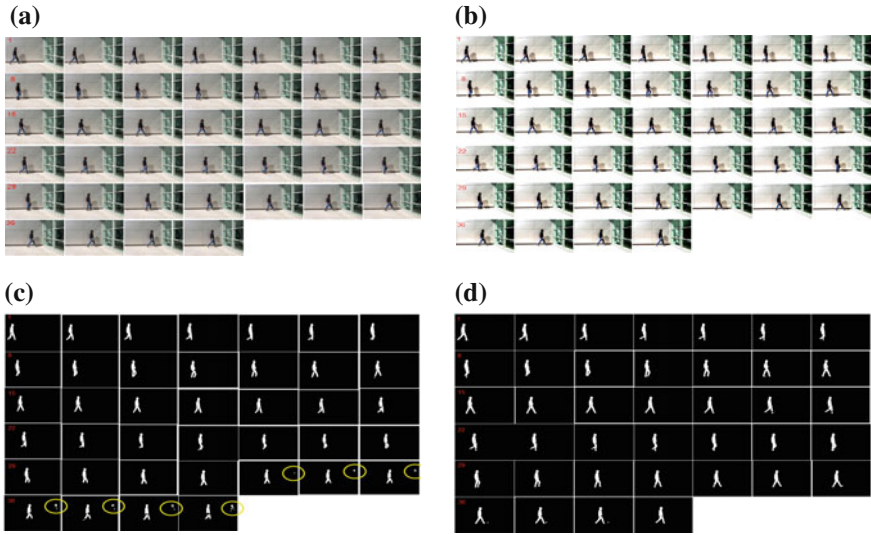
Fig. 3 Silhouette extraction from coloured normal frames on our own collected data with glass reflection on the *right side* of frames. **a** Original coloured normal frames. **b** Illumination adjustment. **c** Silhouettes of background subtraction. **d** Silhouettes of our method
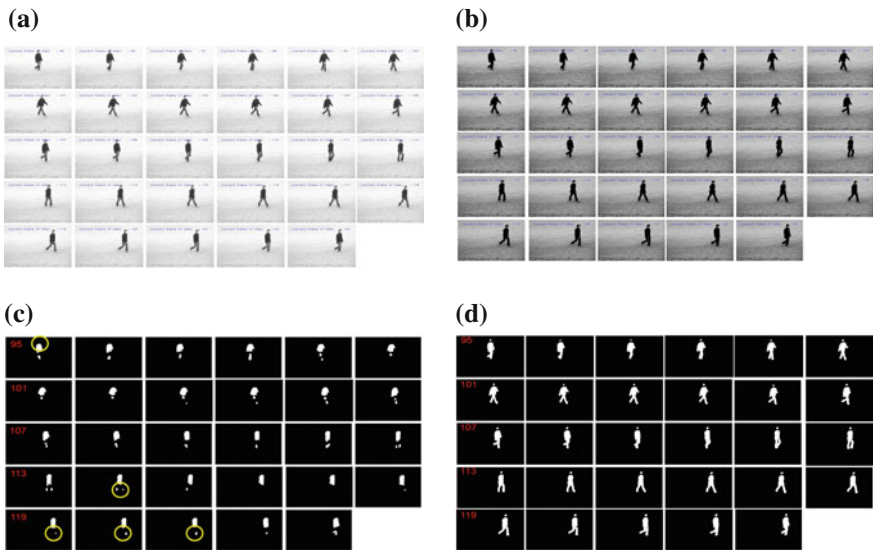

Fig. 4 Silhouette extraction from grayscale bright frames on KTH Database. **a** Original grayscale bright frames. **b** Illumination adjustment. **c** Silhouettes of background subtraction. **d** Silhouettes of our method
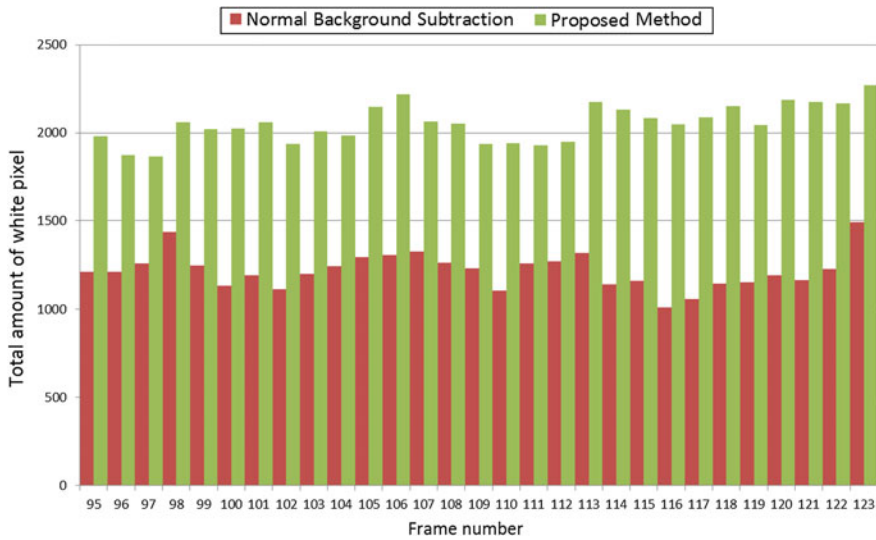
Fig. 5 White pixel counts of Fig. 4 between our method to background subtraction

Figure 3 shows that the proposed method increases the accuracy of silhouette extraction by isolating between actual figures and noise which in this example, a reflection on the right side of each frame. The silhouettes captured (Fig. 3d) are more accurate when compared to standard background subtraction Fig. 3c.

Figure 4 illustrates that the proposed method improves the silhouette extraction which in this example represents a bright type of frame from grayscale KTH video. The silhouettes of our method (Fig. 4d) capture more relevant areas than normal background subtraction where there are missing body and legs (Fig. 4c). Figure 5 illustrates the difference in the number of white pixels (area of subject covered) for these frames where it has shown improvements by using our proposed method of illumination adjustment in our background model.

The average read-write processing time per 35 coloured frames (an average of 1 gait cycle) of our method was 1:49.8 s. This is not far from the standard background subtraction processing time of 1:46.2 s.

5 Conclusion and Future Work

The primary goal of this work is to improve silhouette extraction on videos with illumination inconsistency that may cause problem for person recognition by gait. We presented a heuristic approach that achieves its accuracy by exploiting the intensity of the pixel values to produce better silhouette than background subtraction. Based on a database of 343 subjects, we have found significant improvement in extracting silhouette images as almost every test shows

improvement of at least 50 % in detecting details of subject compare to normal background subtraction method. Future work would be to test the robustness of the captured silhouettes in a person recognition task.

References

1. Wang H, Suter D (2005) A re-evaluation of mixture of Gaussian background modeling. In: IEEE International conference acoustics, speech, and signal processing, 2005. Proceedings. (ICASSP 2005), vol 2
2. Zivkovic Z, van der Heijden F (2006) Efficient adaptive density estimation per image pixel for the task of background subtraction. Elsevier B.V.
3. Wang Z, Shin B-S, Klette R (2006) Accurate human silhouette extraction in video data by shadow evaluation. IJCTE 22:545–557
4. Cheng H-Y, Wu Q-Z, Fan K-C, Jeng B-S (2006) Binarization method based on pixel-level dynamic thresholds for change detection in image sequences. J Inf Sci Eng 22(3):545–557
5. Hofmann M, Rigoll G (2013) Exploiting gradient histograms for gait-based person identification. IEEE international conference on image processing (ICIP), pp 4171–4175
6. Xu M, Ellis T (2001) Illumination-invariant motion detection using colour mixture models. In: Proceedings of British machine vision conference, Sept 2001
7. Ng H, Tan W-H, Abdullah J (2013) Multi-view gait based human identification system with covariate analysis. Int Arab J Inf Technol 10(5)
8. Tan X, Triggs W (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process Inst Electr Electron Eng (IEEE) 19 (6):1635–1650
9. Menotti D, Najman L, Facon J, de Araujo AA (2007) Multi-histogram equalization methods for contrast enhancement and brightness preserving. IEEE Trans Consum Electron 53 (3):1186–1194
10. Ibrahim H, Kong NSP (2007) Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans Consum Electron 53(4):1752–1758
11. Agaian SS, Silver B, Panetta KA (2007) Transform coefficient histogram-based image enhancement algorithms using contrast entropy. IEEE Trans Image Process 16(3):741–758