

FOREGROUND DETECTION IN CAMOUFLAGED SCENES

Shuai Li^{*}, Dinei Florencio[†], Yaqin Zhao[‡], Chris Cook^{*}, Wanqing Li^{*}

University of Wollongong^{*}, Microsoft Research[†], Nanjing Forestry University[‡]

ABSTRACT

Foreground detection has been widely studied for decades due to its importance in many practical applications. Most of the existing methods assume foreground and background show visually distinct characteristics and thus the foreground can be detected once a good background model is obtained. However, there are many situations where this is not the case. Of particular interest in video surveillance is the camouflage case. For example, an active attacker camouflages by intentionally wearing clothes that are visually similar to the background. In such cases, even given a decent background model, it is not trivial to detect foreground objects. This paper proposes a texture guided weighted voting (TGWV) method which can efficiently detect foreground objects in camouflaged scenes. The proposed method employs the stationary wavelet transform to decompose the image into frequency bands. We show that the small and hardly noticeable differences between foreground and background in the image domain can be effectively captured in certain wavelet frequency bands. To make the final foreground decision, a weighted voting scheme is developed based on intensity and texture of all the wavelet bands with weights carefully designed. Experimental results demonstrate that the proposed method achieves superior performance compared to the current state-of-the-art results.

Index Terms— Foreground detection, background subtraction, camouflaged scenes, wavelet transform

1. INTRODUCTION

Detection of moving foreground objects is a crucial step in many vision-based systems. Most foreground detection algorithms are based on background subtraction. More specifically, they compute a background model and make foreground decisions by comparing the current image against that background model. A classic example (and still one of the most popular) of a background subtraction method is the Gaussian Mixture Model (GMM) [1] and the improved adaptive Gaussian Mixture Model [2, 3] known as “MOG” and “MOG2”, respectively, which uses a few Gaussian distributions to model the intensity of each pixel. A pixel-based adaptive method which, instead of using the distribution of pixels, uses a history of recently observed pixels for background

modeling was proposed in [4], noted as “SubSENSE”. While all these methods provide useful classification for many situations, they all use only the intensity of the current pixel, and will likely fail when the pixels of the background and the foreground object share similar color.

Another class of algorithms [5, 6, 7, 8, 9, 10], instead of looking simply at the pixel intensity, tries to detect foreground objects based on features that intend to capture the texture around each pixel, i.e., the relationship between the current pixel and a small predefined neighborhood. These features include “Haar”, “Gradient” and “Local Binary Pattern (LBP) [11]”. However, these kinds of features can only exploit a certain type of texture in a small predefined region and cannot adaptively change the scale. When a pixel is located in a region which has relative poor texture at the chosen resolution, the method will likely fail. There have been a few methods [12, 13, 14] in the literature using wavelet transform for background subtraction. However, the existing methods only use the wavelet transform as an additional feature without considering the characteristics of wavelet transform, and thus do not take full advantage of the wavelet transform.

A comprehensive review of recent background subtraction and foreground detection methods can be found in [15]. Existing methods were developed to deal with general foreground objects that show distinct intensity or texture changes. However, there are situations where foreground objects may share similar intensity and texture as the background, such as camouflaged scenes, especially those with poor texture. For example, a person may wear clothes that share similar color to that of the background wall as shown in Fig. 1. These cases pose great challenges to foreground detection but are extremely important for surveillance applications, but they have not yet received much attention in the literature.

This paper focuses on the foreground detection for camouflaged scenes, addressing the above-mentioned problem where foreground objects are visually similar to the background and lack distinct intensity or texture differences. The contributions of this paper are summarized as follows.

- We investigate the differences between foreground objects and visually similar background in the image and wavelet domain, and show that the small differences in the image domain may be able to be detected in certain wavelet frequency bands.
- We propose a weighed voting method to combine all the decisions made on the differences in terms of intensity and texture of the bands, respectively.

Shuai Li is partially supported by the Global Challenge Project, Assistive Systems for the Ageing, of University of Wollongong.



Fig. 1. Example of a camouflaged scene and its differences in terms of intensity and LBP in the image domain between the current image and the background image.

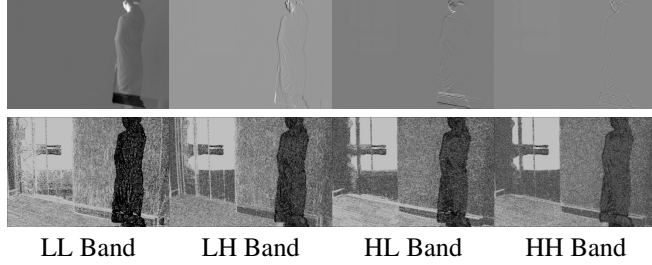


Fig. 2. Differences in terms of intensity (wavelet coefficients) and texture (LBP) between the current image and the background image in each wavelet band of the 3-rd level.

- Three weights are carefully designed, namely noise-induced weight, texture-guided weight and translation weight, to deal with noise, different characteristics and scales of different frequency bands, respectively.
- A new dataset for camouflaged scenes captured in real life is collected with labeled groundtruth. Experiments on both the existing dataset (artificially generated) and the new dataset demonstrate significantly better performance using the proposed method over the state-of-the-art methods.

2. FOREGROUND DETECTION IN THE WAVELET DOMAIN

In foreground detection, much effort has been spent on obtaining an appropriate background model but detecting the foreground from the background model is still largely based on the differences between the current image and the background image. Although this may work for scenes where foreground object and background show distinct changes in intensity and texture, it cannot work appropriately for camouflaged scenes where foreground objects are visually similar to background. The first image in Fig. 1 gives an example of a person wearing clothes that share similar color as the wall. The corresponding differences in the image domain in terms of intensity and texture (LBP) between the foreground and background (obtained by MOG2) are shown in Fig. 1 where the values in the image are properly scaled for display. It can be seen that the differences between the foreground person and background in the camouflaged scene are really small and consequently directly detecting the foreground in the image domain may fail.

2.1. Intensity and Texture Analysis in Wavelet Domain

Wavelet transform decomposes images into different frequency bands, allowing for multiresolution analysis. Specially, each stage of a wavelet transform decomposes an image into four frequency bands, typically denoted as LL, LH, HL, and HH. The LL band contains the low frequency approximation of the image, while the other three bands contain the details (high frequency) of the image in the horizontal, vertical and diagonal directions, respectively. This paper employs a stationary wavelet transform [16, 17], i.e., a non-decimated redundant wavelet transform. The Haar wavelet basis is used, as it provides a simple and effective basis. Considering that we only make use of the wavelet bands as features instead of reconstructing the original image, all wavelet bands including the low frequency wavelet bands at all levels are used in the proposed method.

Different levels of wavelet decomposition show different frequency bands and each frequency band extracts different characteristics of an image. Small differences in the image domain may be highlighted in one or a few wavelet bands when large differences in certain frequencies exist between foreground objects and the background. Fig. 2 shows the intensity (wavelet coefficient) differences and the texture (LBP) differences in the wavelet domain, respectively, in each band of level 3 (other levels are very similar and thus not shown here). Compared to the difference in the image domain shown in Fig. 1, it can be seen that the differences become apparent in the wavelet domain, which makes the detection of the foreground from the visually similar background possible.

2.2. Texture Guided Weighted Voting (TGWV)

The framework of the TGWV based foreground decision method developed in this paper is shown in Fig. 3. First a background model is constructed and the background image for the current time is formed. As this paper mainly focuses on foreground detection for camouflaged scenes, for simplicity existing background modeling processes such as the GMM in MOG2 [3] and the SuBSENSE in [6] are used. We have evaluated in the Experimental Results Section that our proposed foreground detection method can be efficiently combined with any background modeling method while achieving superior performance. Then the stationary wavelet transform is employed to decompose the background image and the current image into wavelet bands of M levels (M is empirically set as 7 in the current experiments). The uniform LBP histogram is then extracted for each wavelet band to represent their textures, referred as LBP hereafter. The differences between the current image and the background image in the wavelet domain are obtained as the absolute differences of the coefficients and the LBP differences measured by the histogram intersection kernel [18]. Then foreground decisions are made separately for the coefficients and LBP in each wavelet band as in GMM by formulating a Gaussian distribution on the differences. Finally the decision for each pixel is obtained by a weighted voting scheme which will be explained in the following.

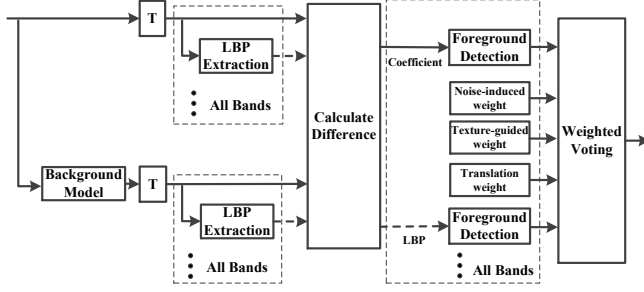


Fig. 3. Framework of the foreground detection method developed in this paper.

2.2.1. Noise-induced Weight

It is known that noise exists in images, even in uncompressed ones, coming from the acquisition process. Usually white Gaussian noise, which has zero mean and shows the same energy at different frequency bands, is assumed. The decisions made on each wavelet band may be affected by the noise and thus the confidence level of each decision could be compromised. However, energies of different wavelet bands are quite different and thus the effects of the noise on the decisions are different. Since that the decision on each wavelet band is based on the value of the coefficient, the confidence ω_{ni} of the decision by considering the effect of the noise can be determined as

$$\omega_{ni} = \frac{\sigma_{si} - \sigma_n}{\sigma_{si}} \quad (1)$$

where σ_{si} and σ_n are the standard deviation of the i -th wavelet band and noise, respectively.

2.2.2. Texture-guided Weight

Different frequency bands show different responses to different textures. For example, high frequency wavelet bands only respond to the changes across pixels, which makes them not discriminative for flat regions. Therefore, to adaptively adjust the confidence of the low frequency and high frequency wavelet bands in each level on the final decision for different regions with different textures, a texture guided weight w_t is used by taking into consideration the texture (measured by LBP) of the region that the current pixel located in.

$$\omega_{tLi} = 1 - f((n_i(BG) + n_i(C))/K)$$

$$\omega_{tWi} = \begin{cases} 1 - f((n_i(BG) + n_i(C))/K), & \text{if } i \in \{LH, HL, HH\}; \\ 1 + \sum_{k \in \{LH, HL, HH\}} 2 \cdot f((n_k(BG) + n_k(C))/K) \\ \quad + f((n_i(BG) + n_i(C))/K), & \text{if } i = LL \end{cases} \quad (2)$$

where ω_{tLi} and ω_{tWi} represent the weights for the decisions made on LBP and coefficients, respectively. $n_i(BG)$ and $n_i(C)$ represent the number of the last two patterns in the uniform LBP (the pattern with no changes and the pattern with more than two changes indicating how flat the region is) in the background image and the current image, respectively, and K is the total number of all the patterns. $f(x)$ is a non-decreasing function that maps x to a value between 0 and 1.

It can be seen that when the current region is very flat, the foreground detection of the wavelet bands in this level will be determined mostly based on the low frequency wavelet band information.

2.2.3. Translation Weight

As each wavelet band can be considered as a representation of the image in a certain frequency, final foreground decisions of pixels can be obtained by effectively combining all the decisions together. Considering that each wavelet coefficient is obtained based on a few pixels in a block, a translation process is needed to transfer the decisions in the wavelet domain to the image domain. It is clear that if all the pixels related to a coefficient belong to the same object, the decision made on this coefficient can well represent the decision on these pixels in this frequency. On the contrary, if pixels related to a coefficient belong to different objects, the decision made on this coefficient may not be correct for all these pixels. In order to characterize this relationship, the correlation among pixels is first modeled using a first-order autoregressive process as

$$\rho(i+1) = \alpha \cdot \rho(i) + \epsilon(i+1) \quad (3)$$

where α , $0 < \alpha < 1$, is the autoregressive coefficient and ϵ is a white noise process with zero mean. When the distance between two pixels increases, their correlation reduces. Therefore, as the wavelet decomposition level increases and the number of pixels related to one coefficient increases, the correlation between the central and the related pixels gets smaller as their distance apart get greater. Consequently, the confidence of the decision made on the coefficient being correct to all its related pixels gets smaller. A translation weight ω_{ci} is used for each wavelet coefficient to transfer the decision to its related pixels, which is determined as the average correlation of all the pixels related to the coefficient.

$$\omega_{ci} = \frac{1}{N_p} \sum_{i=1}^{N_p} \rho(i) \quad (4)$$

where N_p is the total number of the pixels related to one coefficient.

2.2.4. Weighted Voting based Foreground Detection

By combining the noise induced weight in (1), the texture guided weight in (2) and the translation weight in (4), we can obtain the final decision for each pixel using a weighted voting strategy as

$$V = \sum_{i=1}^N \omega_{ni} \cdot \omega_{ci} \cdot (\omega_{tWi} \cdot V_{Wi} + \omega_{tLi} \cdot V_{Li})$$

$$F = \begin{cases} 1, & \text{if } V > T_V; \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where V_{Wi} and V_{Li} represent the foreground decisions made based on the wavelet coefficient and LBP in the i -th wavelet band, respectively, and N represents the total number of wavelet bands. F represents the final foreground decision for a pixel, and T_V is the minimum number of votes needed to determine the current pixel as foreground.

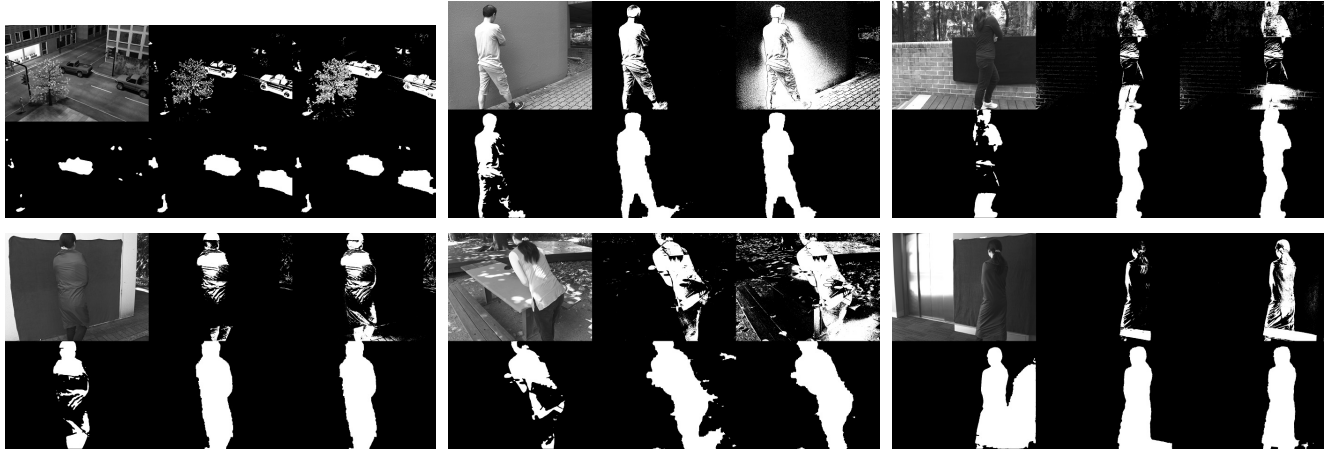


Fig. 4. Detection result examples. For each figure, the images in the first row from left to right are the current image, the results obtained by “MOG2 [3] without shadow”, “MOG2 [3] with shadow”, respectively. The images in the second row from left to right are the results obtained by “SubSENSE[6]”, “Proposed + MOG2” and “Proposed + SuBSENSE”, respectively.

3. EXPERIMENTS

3.1. Datasets

As reported in [19], a camouflaged video has been used to test the performance of the existing algorithms for detecting camouflaged objects, which is adopted in this paper for evaluation. However, this dataset was artificially generated by computer graphics and it can hardly simulate the texture on the real foreground object. Therefore, we captured another dataset containing 5 grayscale video sequences of resolution 1536*1152 in real scenes. The foreground person wears clothes in a similar color as that of the background. Examples of the scenes with the foreground person in the collected dataset together with the camouflaged video [19] are shown in Fig. 4 (first image of each figure).

3.2. Results

The proposed foreground detection method is compared with the popular MOG2 [3] and one most recent work SubSENSE [6] which has reported the state-of-the-art performance on background subtraction. The performance is measured using Recall, Precision, F-Measure and PSNR obtained by the BMC Wizard software [20]. Due to the page limitation, only the average experimental results of all the videos including the camouflaged video in [19] are shown as in Table 1. As expressed in the previous Section, the proposed method focuses on detecting the foreground from the visually similar background and the background modeling process is implemented using the GMM model in MOG2 [3] or SubSENSE [6] which are denoted by “Proposed + MOG2” and “Proposed + SuBSENSE”, respectively. From the results, it can be seen that the proposed method can achieve significant improvement. Especially in terms of Recall, the proposed method can perform much better and can detect most of the foreground objects. It can be also seen that regardless of the background modeling methods used, the proposed method consistently achieves superior performance. Some example detection results are also shown in Fig. 4. It can be clearly seen that the results obtained

Table 1. Average Performance of All Tested Videos

Method	Recall	Precision	F-Measure	PSNR
MOG2[3] with shadow	0.859	0.731	0.788	30.162
MOG2[3] without shadow	0.786	0.879	0.826	36.792
SuBSENSE[6]	0.847	0.911	0.873	38.994
Proposed + MOG2	0.984	0.876	0.926	42.904
Proposed + SuBSENSE	0.979	0.899	0.937	45.645

by the proposed method are better than the existing methods in detecting the foreground objects that are visually similar to the background.

The computational complexity of the proposed method is higher than MOG2 [3] and SubSENSE [6] due to the multiple level wavelet decomposition. However, the algorithm is highly parallel and can be easily implemented on GPU [21].

4. CONCLUSION

This paper presents a texture guided weighted voting scheme in the wavelet domain to addresses the foreground detection problem in camouflaged scenes. It is first shown that small and hardly noticeable differences between foreground objects and visually similar background can be well captured by a few wavelet bands. Then foreground decisions are made for each wavelet band based on wavelet coefficients and local binary patterns obtained on the coefficients, respectively. To effectively combine all the decisions for the final detection on the pixels, the property of the wavelet transform is considered and a texture guided weight is adaptively determined for each decision. Experimental results have verified the efficacy of the proposed foreground detection method.

5. REFERENCES

- [1] Pakorn KaewTraKulPong and Richard Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Video-based surveillance systems*, pp. 135–144. Springer, 2002.
- [2] Zoran Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition*. IEEE, 2004, vol. 2, pp. 28–31.
- [3] Zoran Zivkovic and Ferdinand van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters*, vol. 27, no. 7, pp. 773–780, 2006.
- [4] Martin Hofmann, Philipp Tiefenbacher, and Gerhard Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2012, pp. 38–43.
- [5] Marko Heikkilä and Matti Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 657–662, 2006.
- [6] Pierre-Luc St-Charles, Guillaume-Alexandre Bilodeau, and Robert Bergevin, "Flexible background subtraction with self-balanced local sensitivity," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2014, pp. 414–419.
- [7] Bohyung Han and Larry S Davis, "Density-based multi-feature background subtraction with support vector machine," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 1017–1023, 2012.
- [8] Xu Jian, Ding Xiao-qing, Wang Sheng-jin, and Wu You-shou, "Background subtraction based on a combination of texture, color and intensity," in *9th International Conference on Signal Processing*. IEEE, 2008, pp. 1400–1405.
- [9] Srenivas Varadarajan, Lina J Karam, and Dinei Florencio, "Background recovery from video sequences using motion parameters," in *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 989–992.
- [10] Csaba Benedek and Tamás Szirányi, "Bayesian foreground and shadow detection in uncertain frame rate surveillance videos," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 608–621, 2008.
- [11] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [12] J-C Huang and W-S Hsieh, "Wavelet-based moving object segmentation," *Electronics Letters*, vol. 39, no. 19, pp. 1380–1382, 2003.
- [13] J-C Huang, T-S Su, L-J Wang, and W-S Hsieh, "Double-change-detection method for wavelet-based moving-object segmentation," *Electronics Letters*, vol. 40, no. 13, pp. 798–799, 2004.
- [14] Alok Kumar Singh Kushwaha and Rajeev Srivastava, "Framework for dynamic background modeling and shadow suppression for moving object segmentation in complex wavelet domain," *Journal of Electronic Imaging*, vol. 24, no. 5, pp. 051005–051005, 2015.
- [15] Andrews Sobral and Antoine Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.
- [16] Guy P Nason and Bernard W Silverman, "The stationary wavelet transform and some statistical applications," *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG*-, pp. 281–281, 1995.
- [17] Jean-Christophe Pesquet, Hamid Krim, and Hervé Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1964–1970, 1996.
- [18] Michael J Swain and Dana H Ballard, "Color indexing," *International journal of computer vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [19] Sebastian Brutzer, Benjamin Höferlin, and Gunther Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 1937–1944.
- [20] "Bmc wizard software," http://bmc.iut-auvergne.com/?page_id=63.
- [21] Joaquín Franco, Gregorio Bernabé, Juan Fernández, and Manuel E Acacio, "A parallel implementation of the 2d wavelet transform using cuda," in *17th Euromicro International Conference on Parallel, Distributed and Network-based Processing*. IEEE, 2009, pp. 111–118.