

Wavelet Multi-Scale Transform Based Foreground Segmentation and Shadow Elimination

Ye-Peng Guan^{*,1,2}

¹School of Communication and Information Engineering, Shanghai University; ²Key Laboratory of Advanced Displays and System Application, Ministry of Education, 149 Yanchang Rd., Shanghai 200072, China

Abstract: An algorithm using wavelet multi-scale transform for segmenting foreground moving objects and suppressing shadow is proposed. The optimal selection of threshold is automatically determined which does not require any complex supervised training, manual calibration or hypothesis. The proposed algorithm is efficient enough to segment foreground moving objects with low contrast against the background. The reference image is used to extract foreground no matter the objects enter the field of view before captured or not. The developed method is highly computationally cost-effective since it does not concern with complex computation model, color model or background statistics at a time. By comparisons, it has been shown that the proposed approach is more robust and efficient to detect foreground and suppress shadow during coping with different indoors or outdoors circumstances.

Keywords: Wavelet multi-scale transformation, Foreground segmentation, Shadow suppression, Background subtraction, Threshold.

1. INTRODUCTION

A relevant problem in computer vision is how to detect and track moving objects in video sequences efficiently. The most widely adopted approach for moving object detection is based on background subtraction [1-6]. Moving objects in the scene are detected by the difference between the current frame and the current background model. When the deviation is greater than some critical value, the pixel is considered to be foreground (moving object) [7-9]. A way to model the background is to take the mean, median, or minimum and maximum values of the previous N pixel [2]. It needs to keep track of the pixel value history. To avoid this problem, Heikkila *et al.* [3] used a first order recursive filter to update the background model. However it also causes tailing or ghosting. One of the most important drawbacks of above mentioned techniques is that the same threshold was used for every pixel. A moving object is likely to disappear in the background when entering a darker area in the scene in this way. A popular approach to performing the automatic thresholding of difference images is to assume particular distributions for the difference of image samples and the noise [4-5, 10-11]. These methods use statistical texture properties of the background observed over extended period of time to construct a model of the background, and use this model to decide which pixels in an input image do not fall into the background class. One method to achieve this model is by modeling each pixel as unimodal Gaussian distribution [1]. To overcome some difficulties existing in unimodal background model, parametric background modeling is developed by modeling each background pixel value as a

mixture of Gaussians [12-13]. Since the parametric background model still lacks flexibility when dealing with non-static backgrounds, a highly flexible non-parametric technique is proposed for estimating background probabilities from many recent samples over time using Kernel density estimation [14]. The parametric and non-parametric background subtraction algorithms discussed are based on the probability of observing a certain pixel value. A different approach making use of codebook technique was proposed by Kim *et al.* [15-16]. One of the main drawbacks of codebooks is that the algorithm cannot adapt to changing circumstances when these circumstances were not present in the training phase. Also moving objects that stop moving and should be adopted into the background will get difficulties in this algorithm [9].

In addition, moving object segmentation is affected by the problem of shadows [8, 17-22]. Unfortunately the methods for shadow suppression or identification require manual calibration in terms of the specification of parameters which are related to the environment and the lighting conditions.

In this paper, a method using wavelet multi-scale transform to extract foreground and suppress shadow is proposed, which is applicable for color video images. The proposed method exploits the Hue-Saturation-Value (HSV) color space instead of introducing complex color models. The main contribution of this proposal is to automatically determine the optimal selection of threshold that incorporates the spatial interactions between neighboring pixels for different images. The second contribution is that the developed approach does not require any complex supervised training phase, or make any hypothesis, or need manual calibration for shadow suppression in terms of the specification of parameters related to the environment and the lighting conditions. The experimental results from indoor and outdoor environments have showed the developed approach is efficient

*Address correspondence to this author at the School of Communication and Information Engineering, Shanghai University; Key Laboratory of Advanced Displays and System Application, Ministry of Education, 149 Yanchang Rd., Shanghai 200072, China; Tel: +86 21 56331967; Fax: +86 21 56336908; E-mail: ypguan@shu.edu.cn

and flexible during segmenting moving objects and suppressing shadows in the applications.

The organization of the rest paper is as follows. In the next section, we will discuss wavelet multi-scale transformation. Foreground extraction and shadow suppression are described in section 3 and 4, respectively. Experimental results are given in section 5 and followed by conclusions in section 6.

2. WAVELET MULTI-SCALE TRANSFORMATION

One of the two basic properties, similarity and discontinuity, underlie most current popular segmentation algorithms for images. The similarity category partitions an image into regions that are similar to a set of predefined criteria. The approach for discontinuity is to partition an image based on abrupt changes in intensity. The low-pass and high-pass filters of the wavelet transform (WT) naturally break a signal into similar and discontinuous sub-signals, which effectively combines the two basic properties into a single approach. It is the main motivation using WT to segment foreground.

The 2-D WT of an image $I(x, y)$ at scale 2^j and in orientation k is defined as [23]

$$W_{2^j}^k f(x, y) = I * \psi_{2^j}^k(x, y), k = 1, 2 \quad (1)$$

The two oriented wavelets $\psi_{2^j}^k$ can be constructed by taking the partial derivatives

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x}, \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y} \quad (2)$$

where $\theta(x, y)$ is a separable scaling function which plays the role of a smoothing filter.

The 2-D WT defined by (2) gives the gradient of $I(x, y)$ smoothed by $\theta(x, y)$ at dyadic scales.

$$\nabla_{2^j} I(x, y) \equiv (W_{2^j}^1 I(x, y), W_{2^j}^2 I(x, y)) = \frac{1}{2^{2j}} \nabla I * \theta_{2^j}(x, y) \quad (3)$$

Consider the local maxima of the gradient magnitude at various scales which is given by

$$M_{2^j} I(x, y) \equiv \|\nabla_{2^j} I(x, y)\| = \sqrt{(W_{2^j}^1 I(x, y))^2 + (W_{2^j}^2 I(x, y))^2} \quad (4)$$

A point (x, y) is a multi-scale edge point at scale 2^j if the magnitude of the gradient $M_{2^j} I$ attains a local maximum along the gradient direction $A_{2^j} I$, defined by

$$A_{2^j} I(x, y) \equiv \arctan \left[\frac{W_{2^j}^2 I(x, y)}{W_{2^j}^1 I(x, y)} \right] \quad (5)$$

For each scale, we can collect the edge points together with the corresponding values of the gradient, namely the WT values at that scale. The resulting local gradient maxima set at scale 2^j is

$$P_{2^j}(I) = \left\{ p_{2^j, i} = (x_i, y_i); \nabla_{2^j} I(x_i, y_i) \right\} \quad (6)$$

where $M_{2^j} I(x_i, y_i)$ has local maximum at $p_{2^j, i} = (x_i, y_i)$ along the direction $A_{2^j} I(x_i, y_i)$.

For a J -level 2-D dyadic WT, the set

$$\rho(I) = \left\{ S_{2^j} I(x, y), \left[P_{2^j}(I) \right]_{1 \leq j \leq J} \right\} \quad (7)$$

is called a multi-scale edge representation of the image $I(x, y)$.

Here $S_{2^j} I(x, y)$ is the low-pass approximation of $I(x, y)$ at the coarsest scale 2^j .

The above algorithm is based on a dyadic scale, which reaches large scales quickly. However, when dealing with noisy images, the noise level is sensitive to the change of scales. A dyadic sequence of scales cannot always optimally adapt to the effect of noise [24].

To overcome the drawback, we set a threshold T and consider those points with gradient above the threshold as significant points instead of finding the local maxima. Given the source image $f(x, y)$, the interesting image $E(x, y)$ is given by

$$E = \sqrt{(f \otimes s)^2 + (f \otimes t)^2} \geq T \quad (8)$$

where s and t are two filters of different directions.

3. FOREGROUND EXTRACTION

A popular approach to foreground segmentation is to perform thresholding of difference images. It is necessary to determine the optimal threshold values. Some algorithms proposed require manual calibration in terms of the specification of parameters or some hypotheses.

In this paper, we use wavelet multi-scale transform based background subtraction to extract foreground moving objects. The solution provided in this paper use HSV color space instead of introducing complex color models to detect which pixels have changed due to a moving object, and which pixels have changed due to moving shadow. Since the hue component of HSV might change quite dramatically depending on the strength of the shadow [8], we use value component of HSV to segment foreground objects and adopt saturation component of HSV to suppress shadow.

Suppose the difference between current frame and background model of value component to be ΔV . With a straightforward algorithm, we would use wavelet multi-scale transform to detect the moving objects region based on ΔV . We define a foreground mask P for each pixel (x, y) with the following condition

$$P = \begin{cases} 1 & \text{if } E_{\Delta V} \geq \sigma_{E_{\Delta V}} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where $E_{\Delta V}$ is wavelet multi-scale transform on ΔV , $\sigma_{E_{\Delta V}}$ is a standard deviation of $E_{\Delta V}$.

If the distribution of $E_{\Delta V}$ is assumed to be a Gaussian one, the threshold can be determined by standard deviation of $E_{\Delta V}$ to achieve a desired detection rate. However, the $E_{\Delta V}$



Fig. (1). Detection using value component of HSV for Hall Monitor (frame #47). Original image, $E_{\Delta V}$ image, and extracted possible moving object regions from left to right, respectively.

distribution is not a Gaussian one, but it has a very sharp peak near zero. Thus we use standard deviation of $E_{\Delta V}$ to select threshold. It has proven to be reasonable by experiments on different scenes in real applications also.

Fig. (1) illustrates the moving cast shadows are extracted by the proposed method also. It is desirable to discriminate between moving targets and their detected shadows.

4. SHADOW SUPPRESSION

By shadow suppression, the major problem is how to distinguish between moving cast shadows and moving object points. Stauder *et al.* [25] had done an extensive analysis on the behavior of shadows and has proposed an approach to avoiding the moving shadows. However, the approach is very complex and many computational intensive operations are used to achieve good results. Cucchiara *et al.* [17] presented a non-model based shadow detection approach. They defined a shadow mask S for each point p resulting from motion segmentation based on the following three conditions

$$S = \begin{cases} 1 & \text{if } \alpha \leq I_k^V / B_k^V \leq \beta \wedge |I_k^S - B_k^S| \leq \tau_S \wedge |I_k^H - B_k^H| \leq \tau_H \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where I_k and B_k are the pixel values at point p in the input image (frame k) and the background model (computed at frame k), respectively. The “ \wedge ” symbol represents logical AND.

However, shadow detection algorithm often makes additional assumptions such as small changes in hue and saturation, necessary a prior knowledge of the bands of the changes in the value channel. Moreover choice of parameters τ_S and τ_H is less straightforward and for now is done empirically. Tattersall *et al.* [8] proposed adaptive shadow identification through automatic parameter estimation based on the above method. The shadow mask S is defined as follows

$$S = \begin{cases} 1 & \text{if } \lambda < I_k^V / B_k^V < 1.0 \wedge |I_k^S - B_k^S| < 30 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Different from the method [17], the single variable parameter used is the parameter λ which is automatically calibrated by parameter detection algorithm. However, much additional assumptions must be made also such as number of object regions no more than twice and no less than half the number of shadow regions, average area of shadow regions no significantly larger than the average area of the object regions, number of shadow pixels touching object pixels be small (less than 10%), and so on [8].

In this paper, we propose an efficient method to suppress shadow. We adopt saturation component of HSV to suppress shadow as described in the previous section for foreground extraction. In this way, (9) can be modified as follows for suppressing shadows.

$$P = \begin{cases} 1 & \text{if } E_{\Delta V} \geq \sigma_{E_{\Delta V}} \wedge E_{\Delta S} \geq \sigma_{E_{\Delta S}} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where $\sigma_{E_{\Delta S}}$ is a standard deviation of $E_{\Delta S}$, the “ \wedge ” symbol represents logical AND.

Fig. (2) illustrates the effect of shadow suppression. In the example, we notice that the shadow is suppressed after this stage.

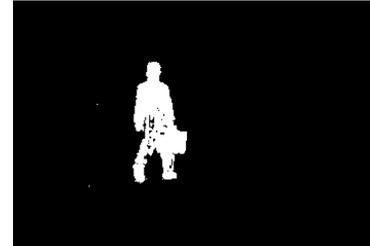


Fig. (2). Detection after shadow suppression for Hall Monitor (frame #47).

Since the detected moving target region may be not complete, some shadow points may be misclassified. To overcome this problem, morphological processing can be applied to the detected foreground objects. This processing allows us to obtain a better performance of clutter background.

5. EXPERIMENTAL RESULTS

To confirm the effectiveness of the above proposed method, we conducted experiments with various indoor and outdoor sequences.

In the experiments, the threshold σ used in (12) is automatically determined. The reference image used in the analysis and described is the first frame of sequence no matter the objects enter the field of view before captured or not. For all sequences, the same set of parameters has been used.

The sequences acquired in two different indoor scenes including *Laboratory* (320×240pixels per frame) from ATON project at <http://cvrr.ucsd.edu:88/aton/>, and *Hall Monitor* (352×240pixels per frame) from the MPEG-4 test set have been used to test the robustness of developed algo-

rithm with different indoor conditions. The corresponding two example results are shown in Fig. (3).



Fig. (3). Detected foregrounds for indoors circumstances. The first column is the first frame of their corresponding sequences, the second column indicates corresponding current frame of the sequence (98#, 90# from top to bottom, respectively), and the third column is the detected foreground.

In Fig. (3), the results obtained with the developed approach are quite good, considering that the contrast between the object and background is very low, shadows projected onto the ground have been suppressed.

In order to test the proposed approach further in outdoor conditions, the sequences acquired in two different outdoor scenes including *Campus* (352×288pixels per frame) from ATON project at <http://cvrr.ucsd.edu:88/aton/> and a pleasure ground scene *Skate* (640×480pixels per frame), where skater stop for a while and skates have been used. The choice of such different scenes is made to emphasize the great reliability and robustness of the propose approach in outdoors circumstances. The two example results are presented in Fig. (4).



Fig. (4). Outdoor detected foregrounds. The first column is the first frame of their corresponding sequences, the second column is the current frame (60#, 115# from top to bottom, respectively), and the third column is the detected foreground.

In Fig. (4), the results show the robustness of the proposed algorithm to cope with different outdoors circumstances. In the second row example of Fig. (4), some points have been extracted as foreground due to the movement of trees.

The above results show qualitative information about the effectiveness of the developed approach. In order to evaluate

the performance of the proposed method quantitatively, false and true positive rate which constructs a ROC (Receiver Operating Characteristic) curve are used to evaluate the developed algorithm.

The sequences are manually segmented to generate the ground-truth. We calculate tpr and fpr to compare the performance of the proposed with that of other algorithms [4, 13], respectively.

We apply the algorithms on the same sequences and the segmented outputs are compared to the hand segmented ground-truth. To establish a fair comparison, we compute the background model (reference image) and other certain parameters from the first 15 frames across the sequence considered. The first 15 frames are considered as the training set and the remaining frames as the testing set for our experimental framework. The calculated parameters remain constant for the whole sequence.

$$fpr = \frac{FP}{FP + TN}, \quad tpr = \frac{TP}{TP + FN} \quad (13)$$

where FP (false positive) indicates pixels falsely marked as foreground, TN (true negative) the number of correctly identified background pixel, TP (true positive) the number of correctly detected foreground pixels, FN (false negative) pixels falsely marked as background.

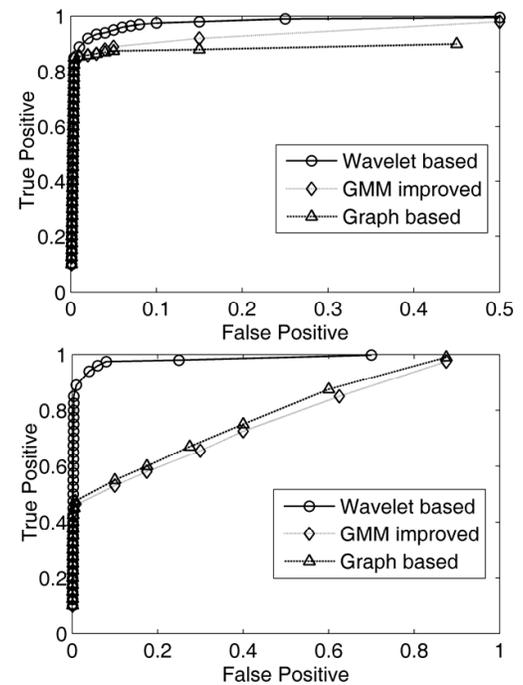


Fig. (5). ROC curves for different approaches including graph cuts based method (Graph based) [4], Gaussian Mixture Model (GMM) [13], and the proposed method (Wavelet based) across the *Laboratory* and *Hall* from top to bottom, respectively.

From the indoors experiments one can notice that the proposed approach is more efficient in detecting foreground moving objects with low contrast against the background than that of others.

From Fig. (6), we further notice that the developed approach is more robust to detect foreground moving objects than that of others in outdoors conditions also. The proposed

approach does not require any complex supervised training phase, and the method for shadow suppression does not require manual calibration in terms of the specification of parameters which are related to the environment and the lighting conditions or makes any hypothesis.

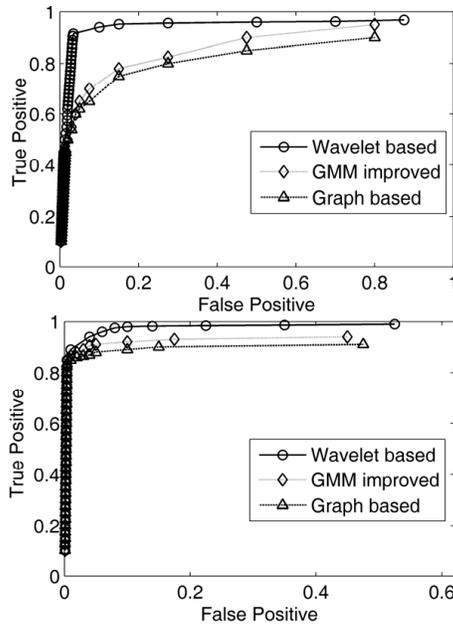


Fig. (6). ROC curves for different approaches including graph cuts based method (Graph based) [4], Gaussian Mixture Model (GMM) [13], and the proposed method (Wavelet based) across the *Campus* and *Skate* from top to bottom, respectively.

For the above three algorithms across the scenes, the average processing times running in MS Visual C++ with a Pentium IV 1.5 GHz 512M RAM computer are given at Table 1.

Table 1. Computational Time Comparison (ms)

Algorithm	Laboratory	Hall	Campus	Skate
Graph	167.8	175.6	187.5	752.
GMM	15.6	16.3	16.8	50.6
Proposed	12.8	13.2	13.7	42.6

By comparison with algorithms [4, 13], it has been shown that the proposed approach is more robust and efficient to detect foreground moving objects and suppressing shadows during coping with different indoors or outdoors circumstances. The developed approach is flexible and computationally cost-effective highly.

6. CONCLUSIONS

This paper has presented an approach to detecting foreground moving objects and suppressing shadows in image sequences. The proposal automatically determines the optimal selection of threshold that incorporates the spatial interactions between neighboring pixels for different images. The proposed algorithm is more flexible in detecting moving objects with low contrast against the background. We consider the reference image used no matter the objects enter the

field of view before captured or not by proposed approach. Besides, the developed algorithm does not require any complex supervised training phase, and the method for shadow suppression does not make any hypothesis or require manual calibration in terms of the specification of parameters which are related to the environment and the lighting conditions. Finally, the method is highly computationally cost-effective since it does not concern with any complex computation model, complex color model or any complex background statistics at a time. By comparison, it has been proven that the proposed approach is more robust and efficient to detect foreground moving objects and suppress shadow during coping with different indoors or outdoors circumstances.

7. ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (60872117).

REFERENCES

- [1] C. Wren, A. Azabayejani, T. Darrell, and A. Pentland, "Pfinder: real-time tracking of the human body", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, pp. 780-785, July 1997.
- [2] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real time surveillance of people and their activities", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 809-830, August 2000.
- [3] J. Heikkila, and O. Silven, "A real-time tracker for visual surveillance applications," in 1st IEEE International. Workshop on Performance Evaluation of Tracking and Surveillance, 2000, pp. 8-13.
- [4] N. Howe, and A. Deschamps, "Better foreground segmentation through graph cuts", Tech. report, Available: <http://arxiv.org/abs/cs.CV/0401017> [Accessed July 16, 2004].
- [5] J. Park, A. Tabb, and A. C. Kak, "Hierarchical data structure for real-time background subtraction," in IEEE International Conference Image Processing, 2006, pp. 1849-1852.
- [6] S. -S. Huang, L. -C. Fu, and P. -Y. Hsiao, "Region-level motion-based background modeling and subtraction using MRFs", *IEEE Trans. Image Process.*, vol. 16, pp. 1446-1456, May 2007.
- [7] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on co-occurrence of image variations," in IEEE International Conference Computer Vision and Pattern Recognition, 2003, pp. 65-72.
- [8] S. Tattersall, and K. Dawson-Howe, "Adaptive shadow identification through automatic parameter estimation in video sequences," in Irish Machine Vision and Image Processing, 2003, pp. 57-64.
- [9] M. Piccardi, "Background subtraction techniques: a review," in IEEE International Conference Systems., Man and Cybernetics, 2004, pp. 3099-3104.
- [10] N. Ohta, "A statistical approach to background subtraction for surveillance systems," in International Conference Computer Vision., 2001, pp. 481-486.
- [11] O. Javed, K. Shafique, and M. Shah, "A hierarchical approach to robust background subtraction using color and gradient information," in IEEE Workshop on Motion and Video Computing, 2002, pp. 22-27.
- [12] C. Stauffer, and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in IEEE International Conference Computer Vision and Pattern Recognition, 1999, pp. 246-252.
- [13] Z. Zivkovic, and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction", *Pattern Recognit. Lett.*, vol. 27, pp. 773-780, May 2006.
- [14] A. Elgammal, D. Harwood, and L. S. Davis, "Nonparametric model for background subtraction," in European Conference Computer Vision, 2000, pp. 751-767.
- [15] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in IEEE International Conference Image Processing, 2004, pp. 3061-3064.
- [16] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model", *J Real Time Imag.*, vol. 11, pp. 172-185, June 2005.

- [17] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1337-1342, October 2003.
- [18] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features", *Comput. Vis. Image Underst.*, vol. 95, pp. 238-259, August 2004.
- [19] T. Thongkamwitoon, S. Aramvith, and T. H. Chalidabhongse, "An adaptive real-time background subtraction and moving shadows detection," in 3rd ACM International Workshop on Video Surveillance and Sensor Networks, 2004, pp. 1459-1462.
- [20] F. Bunyak, I. Ersoy, and S. R. Subramanya, "Shadow detection by combined photometric invariants for improved foreground segmentation," in 7th IEEE Workshop on Application of Computer Vision, 2005, pp.510-515.
- [21] M. Kampel, H. Wildenauer, P. Blauensteiner, and A. Hanbury, "Improved motion segmentation based on shadow detection", *Electron. Lett. Comput. Vis. Image Anal.*, vol. 6, pp. 1-12, Dec 2007.
- [22] T. Zhen, and M. Zhenjiang, "Fast background subtraction and shadow elimination using improved Gaussian mixture model," in IEEE International Workshop on Haptic, Audio and Visual Environments and Games, 2007, pp. 38-41.
- [23] S. Mallat, and W. L. Hwang, "Singularity detection and processing with wavelets", *IEEE Trans. Inf. Theory*, vol. 38, pp. 617-643, March 1992.
- [24] Y. -P. Guan, "Automatic extraction of lip based on multiscale wavelet edge detection", *IET Comput. Vis.*, vol. 2, pp. 23-34, March 2008.
- [25] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation", *IEEE Trans. Multimed.*, vol. 1, pp. 65-76, March 1999.

Received: October 11, 2008

Revised: October 26, 2008

Accepted: October 30, 2008

© Ye-Peng Guan; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.