Neuro-Fuzzy Shadow Classifier

Benny P.L. Lo, Professor Ara Darzi, Professor Guang-Zhong Yang

Department of Computing, Department of Surgical Oncology and Technology

Imperial College of Science, Technology and Medicine

London SW7 2BZ United Kingdom {benny.lo, a.darzi, gzy}@ic.ac.uk

Royal Society & Wolfson Foundation Medical Image Computing Laboratory

INTRODUCTION

Shadow is one of the main problems in object segmentation for video sequence processing. Due to the difficulty of modelling its statistical behaviour, complete shadow removal remains difficult and can lead to errors in determining both shape and object location. Since shadow normally follows the motion of the object and can introduce significant intensity changes to the background, simple intensity and temporal based filters are not effective in practice. Therefore, techniques for removing shadow are often based on thresholding colour differences, as shadow rarely alters the hue of the background pixels. However, for MPEG and MJPEG image sequences, relying on the hue information alone, one cannot identify shadow accurately as the chrominance information is considerably reduced due to quantisation and compression. Figure 1 shows an image extracted from a MPEG sequence and its corresponding hue distribution. In addition, as the attributes of shadows and objects are often very similar, discrete thresholding cannot reliably distinguish one from the other.

This poster presents a self-adaptive neuro-fuzzy shadow filter that combines different aspects of visual characteristics for shadow removal in MPEG and MJPEG images. The strength of the technique is that it relies on real-time cross-referencing of different filter responses for achieving self-adaptation and learning. The proposed method, therefore, does not require explicit thresholding and is applicable to video sequences acquired in different environmental settings.

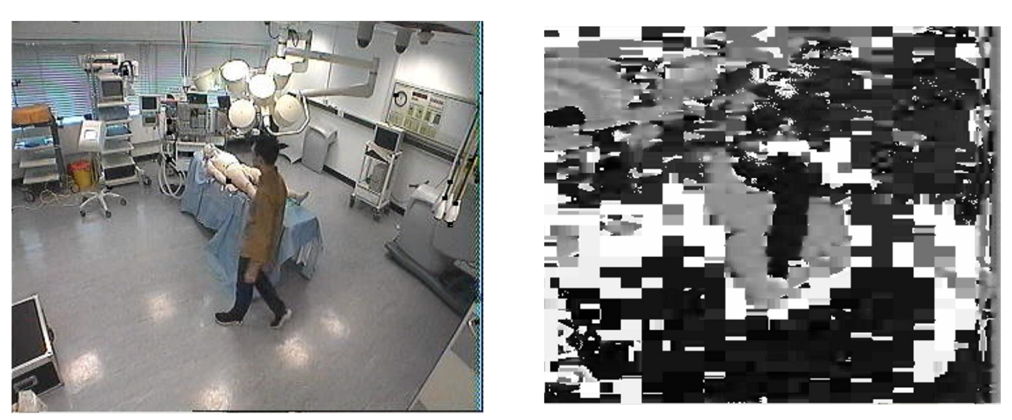

Fig. 1. An image from a MPEG video sequence (left), and the corresponding hue image displayed as a gray scale image (right)

FILTER DESIGN

Figure 2 is a schematic illustration of the proposed filter design. A statistical background removal process based on modelling of the temporal PDF of the in-coming video stream is first applied. The resulting image is then passed onto four different shadow filters (intensity difference, intensity gain, RGB vector angular difference and colour difference) for measuring different visual characteristic. Subsequently, a neuro-fuzzy classifier is then used to combine the outputs of these filters to derive the shadow removed image.

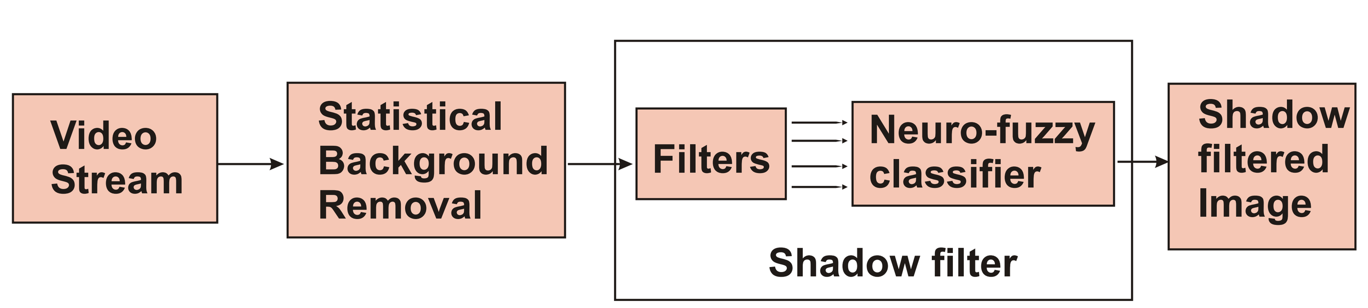

Fig. 2. A schematic illustration of the shadow filter design

Filters

Intensity difference is defined as the absolute difference between the current image $I(x,y)$ and the statistical $B(x,y)$ background image calculated from the peak of the PDF of each pixel, *i.e.*,

$$D(x,y) = \left| I(x,y) - B(x,y) \right| \quad (1)$$

where $D(x,y)$ is the filter output. As dark regions have less significant intensity difference, for darker regions, however, one has to rely on the relative intensity attenuation [1], *i.e.*,

$$G(x,y) = \left| \frac{I(x,y)}{B(x,y)} \right| \quad (2)$$

Based on the property of shadow colour invariance, two colour filters have also been adopted. For compressed images, it has been found that the angle between the RGB vectors provides a good estimation of shadow regions[2].

$$R(x,y) = \frac{\langle I \cdot B \rangle}{\|I\| \|B\|} \quad (3)$$

where I and B are the RGB vectors of the current and the background images. To address the problem of limited colour quantisation steps used in MPEG and MJPEG video sequences, a colour invariant model proposed by Salvador *et al* has been used[3].

$$b_1 = \arctan \frac{R_b}{\max(G_b, B_b)} \qquad c_1 = \arctan \frac{R_i}{\max(G_i, B_i)}$$

$$b_2 = \arctan \frac{G_b}{\max(R_b, B_b)} \qquad c_2 = \arctan \frac{G_i}{\max(R_i, B_i)} \quad (4)$$

$$b_3 = \arctan \frac{B_b}{\max(R_b, G_b)} \qquad c_3 = \arctan \frac{B_i}{\max(R_i, G_i)}$$

$$V(x,y) = (c_1 - b_1)^2 + (c_2 - b_2)^2 + (c_3 - b_3)^2$$

where R_i, G_i, B_i and R_b, G_b, B_b are the RGB components of a given pixel of the current and background images.

Figure 3 illustrates the relative performance of the four shadow filters used. It is evident that none of the filters is ideal for the image concerned. The complementing nature of these filters, however, can be exploited for improved performance in shadow removal.

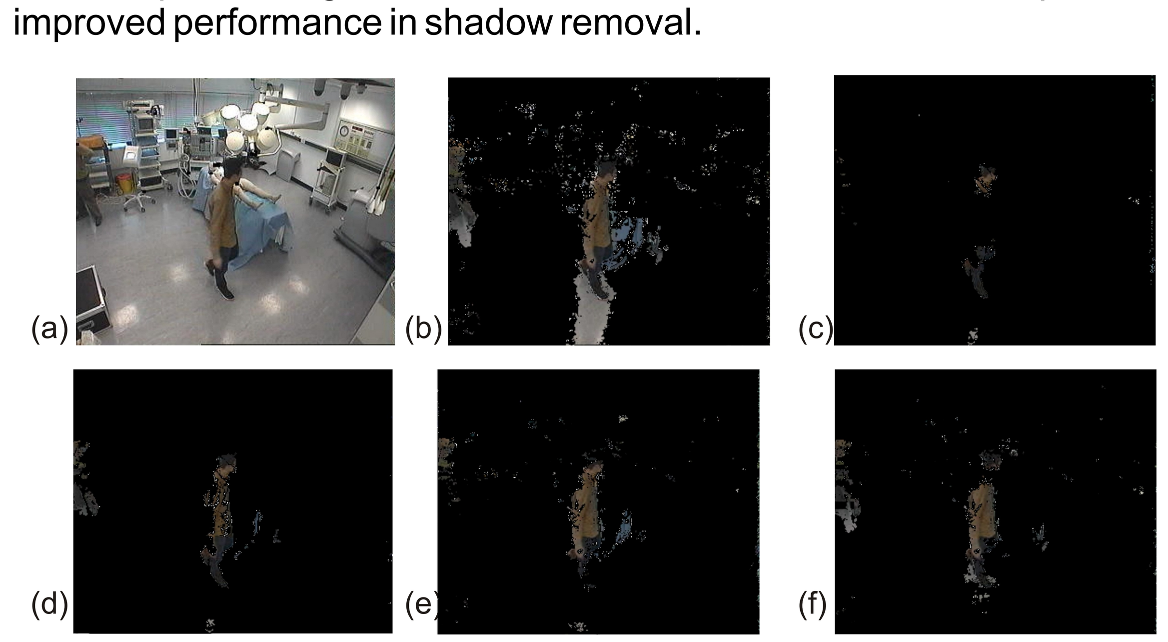

Fig. 3. The relative performance of different shadow filters. (a) The original image, (b) after background removal, (c-f) shadow removal by intensity difference, intensity attenuation, RGB vector angular difference and colour difference, respectively

Neuro-Fuzzy Classifier

In this study, the four filters measure the differences between the current and the background images. As there is no discrete definition on big and small differences between the images, three linguistic meanings are defined to represent the 'low', 'median' and 'high' differences between the images. Accordingly, three fuzzy sets are designed to describe the low, medium and high output levels of each filter. Figure 4 delineates the membership functions of the three fuzzy sets based on the function, where the minimum and maximum range is learnt statistically from the incoming video stream.

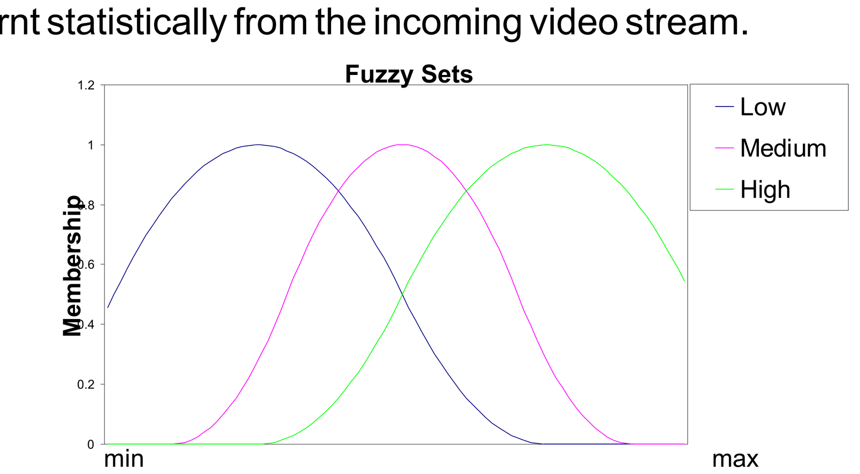

Fig. 4. The definition of the 'low', 'medium' and 'high' membership functions of the fuzzy sets. The 'min' and 'max' are the range for a particular filter output

As such, a total of 12 membership values are obtained from the outputs of the four filters, based on which a multi-layered perceptron (MLP) network with 1 hidden layer and 10 hidden nodes is used to identify shadow pixels. MLP is a neural network that requires supervised learning. For video sequence processing, however, it is difficult in practice to perform such training with example data sets. To address this problem, we have used a contextual based training routine for adapting the shadow filter responses based on the following rules:

1) *If the outputs of the filters area all 'low', thecorresponding pixel is a shadow pixel.*
2) *If the outputs of the filters area all 'high', the corresponding pixel is and object pixel.*
3) *If a shadow pixel is surrounded mainly by object pixels and the outputs of the filters are not 'low', the corresponding pixel should be re-classified as an object pixel instead.*
4) *If an object pixel is surrounded mainly by shadow pixels and the outputs of the filters are not 'high', the corresponding pixel should be re- classified as a shadow pixel instead.*

For rules (3) and (4), the pixels to be tested depends on the chosen neighbourhood. For a eight neighbourhood setting, 'mainly' means that there are at least 5 surrounding pixels that are inconsistent with the classification result of the current pixel. During the processing of the video streams, if any of the rules is violated, the MLP is retrained through back-propagation. Figure 5 shows the testing results for the above four hypotheses. It indicates that pixels, which meet condition 1(green) and 4(yellow), are shadow pixels, and pixels that meet condition 2(blue) and 3 (red), are object pixels.

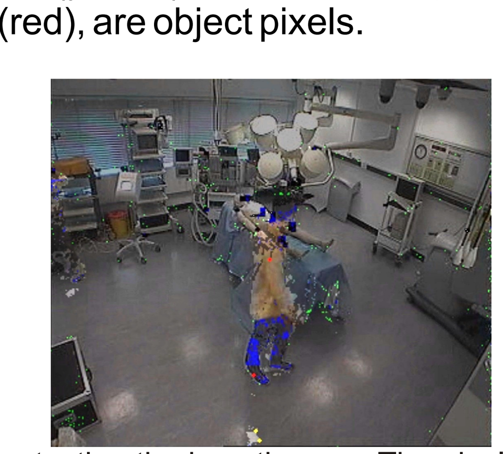

Fig. 5. Experiment results for testing the hypotheses. The pixels that satisfy the first, second, third and forth condition are highlighted with green, blue, red and yellow respectively

RESULT

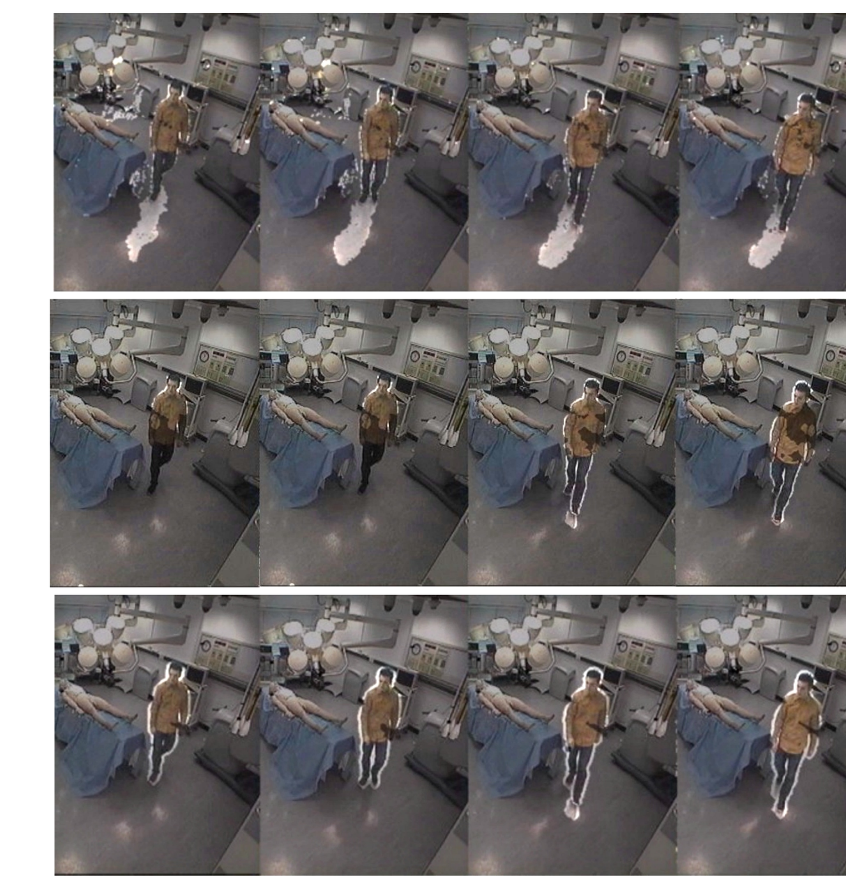

Fig. 6. (Sequence 1) An image sequence showing a person walking inside an operating theatre. Results of object identification based on background removal (top), standard neural network based filter (middle) and neuro-fuzzy filter (bottom) respectively.

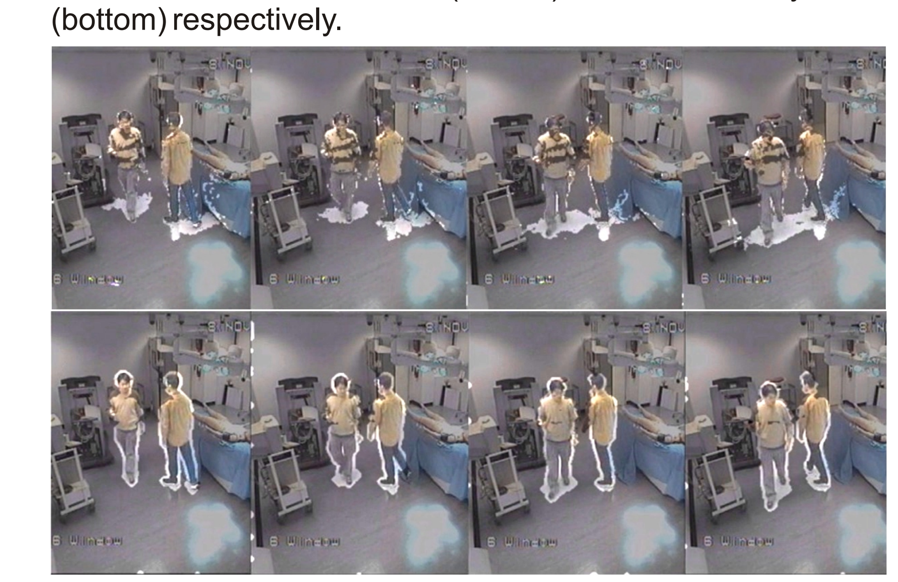

Fig. 7. (Sequence 2) An image sequence showing two people walking towards each other inside an operating theatre. Results of object identification based on background removal (top) and neuro-fuzzy filter (bottom) respectively.

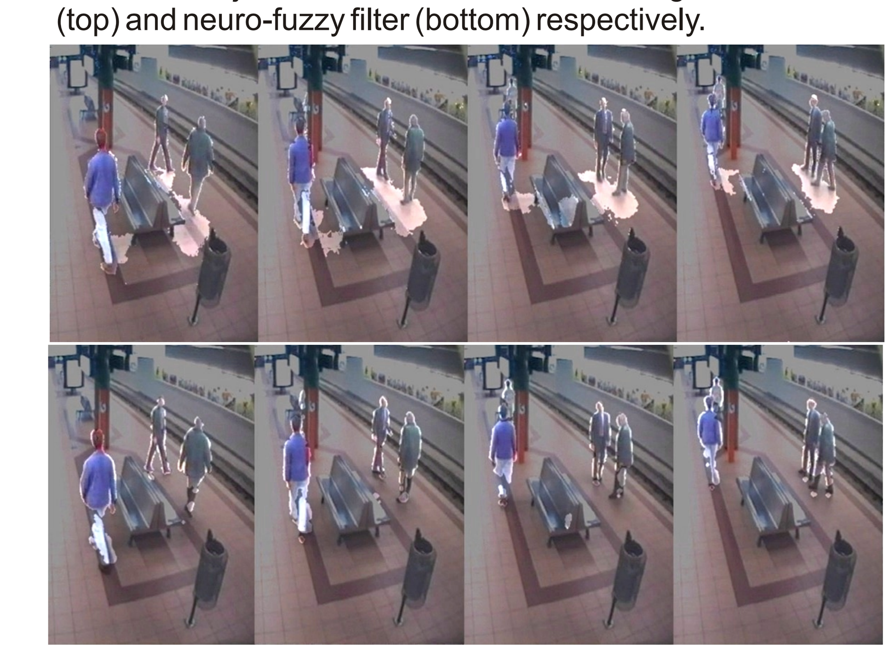

Fig. 8. (Sequence 3) An outdoor image sequence showing three people walking along the platform of a railway station. Results of object identification based on background removal (top) and neuro-fuzzy filter (bottom) respectively.

For quantitative analysis, Figure 9 shows the residual shadow pixels for the above three video sequences. The residual shadow pixels were measured manually before and after the application of the newly proposed shadow filter. The mean and standard deviation of the three video sequences with and without neuro-fuzzy shadow filtering are (269 ± 199, 5870 ± 2999), (501 ± 263, 4545 ± 1920) and (388 ± 274, 3943 ± 1355), respectively. In order to demonstrate the relative merit of different shadow filters, Figure 10(a) illustrates the residual shadow pixels before and after applying these filters for Sequence 1 of Figure 6. Since most shadow filters can also erroneously remove pixels belonging to the moving object, Figure 10(b) measures the amount of distortion introduced by calculating the percentage pixels located within the moving object that have been misclassified. It is evident that the proposed neuro-fuzzy shadow filter provides the best overall performance.

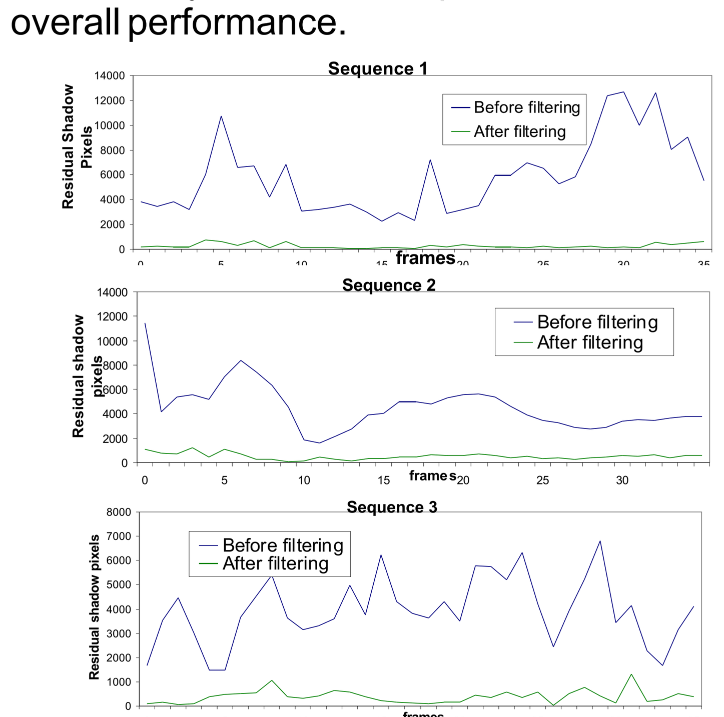

Fig. 9. Residual shadow pixels before and after applying different shadow filters for the three video sequences used in Figure 6, 7 and 8

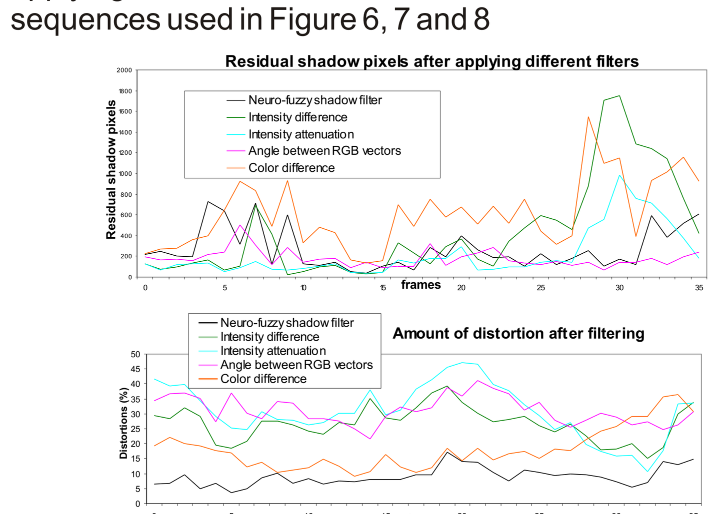

Fig. 10. (a) The relative performance of different shadow filters defined by Equations 1-4, and their combined performance by using the proposed neuro-fuzzy framework. (b) The distortion ratio of the moving object after applying different shadow filters

[1] P.L. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection", *Proceedings of the 6th British Machine Vision Conference*, 347-356, 1995.
[2] W. Lu and Y.P. Tan, "A Color Histogram Based People Tracking System", *IEEE Conf. On ISCAS 2001*, vol. 2, 137-140, 2001.
[3] E. Salvador, A. Cavallaro and T. Ebrahimi, "Shadow Identification and Classification using Invariant Color Models", *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1545-1548, 2001.