Neuro-Fuzzy Shadow Filter

Benny P.L. Lo and Guang-Zhong Yang

Department of Computing,
Imperial College of Science, Technology and Medicine,
180 Queen's Gate, London SW7 2BZ, United Kingdom.
{benny.lo, gzy}@ic.ac.uk

Abstract. In video sequence processing, shadow remains a major source of error for object segmentation. Traditional methods of shadow removal are mainly based on colour difference thresholding between the background and current images. The application of colour filters on MPEG or MJPEG images, however, is often erroneous as the chrominance information is significantly reduced due to compression. In addition, as the colour attributes of shadows and objects are often very similar, discrete thresholding cannot always provide reliable results. This paper presents a novel approach for adaptive shadow removal by incorporating four different filters in a neuro-fuzzy framework. The neuro-fuzzy classifier has the ability of real-time self-adaptation and training, and its performance has been quantitatively assessed with both indoor and outdoor video sequences.

Keywords: Grouping and segmentation, neuro-fuzzy classifier, shadow removal

1 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. Due to the prevalence of video based real-time object tracking applications, reliable shadow removal has become an important research topic.

Based on the fact that shadow rarely alters the hue of the background pixels, a number of techniques have been proposed in the literature by using colour difference thresholding[1,2]. Mikić *et al,* for instance, proposed the use of a statistical model based on this observation for processing traffic scenes, where shadowing represents a major obstacle for robust vehicle tracking[3]. In parallel to the use of colour invariance, researchers have also investigated the use of background intensity changes for shadow removal as in certain circumstances, shadow can be identified from the luminance changes of the background[4,5]. Spatial features have also been used for shadow removal[5,6,7]. Salvador *et al* proposed a method of applying edge detection to the luminance and colour features[6]. For separating moving objects from their shadows, Sonada and Ogata proposed the use of core line analysis and its application

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2352, pp. 381-392, 2002.

to the tracking of loci in image sequences[8]. Under the environment of multiple-camera tracking, shadows can also be removed based on multi-view projection geometry[9].

Existing research has shown that the adaptation of the above techniques in a generic sense can be difficult. The use of thresholding depends on a number of environmental factors that are related to lighting and reflection. For MPEG and MJPEG image sequences, this can become even more problematic 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. It can be seen that by relying on the hue information alone, one cannot achieve reliable shadow identification due to the poor marginalisation of regional borders. On the other hand, shadows cannot be accurately removed with intensity or edge based filters either, since the intensity difference between the background and shadows can often be significantly large, and shadows sometime can have stronger edge than the object itself. Furthermore, since the attributes of shadows and objects are often very similar, discrete thresholding cannot reliably distinguish one from the other.

The purpose of this paper is to propose 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).

2 Filter Design

2.1 Shadow Filters

Figure 2 is a schematic illustration of the proposed filter design. A statistical background removal process based on the modelling of the temporal PDF of the in-coming video stream is first applied. This eliminates most of the stationary objects and the resulting image is then passed onto four different filters for shadow removal based on varying visual characteristics of the shadows. As each filter only partially removes the shadow pixels, a neuro-fuzzy classifier is then used to combine the outputs of these filters to derive the final shadow removed image.

For the four shadow filters used, two of them are based on intensity and the other two on colour information. They include

- intensity difference
- intensity gain
- angle between RGB vectors
- colour difference

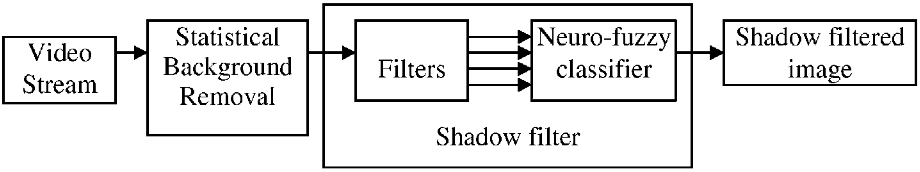


Fig. 2. A schematic illustration of the shadow filter design

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

$$D(x, y) = |I(x, y) - B(x, y)| \tag{1}$$

where $D(x, y)$ is the filter output. The use of intensity difference is biased towards the extraction of shadows in bright regions. For shadows in darker regions, however, one has to rely on the relative intensity attenuation between $I(x, y)$ and $B(x, y)$ [5], *i.e.*,

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

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 [1]. Therefore, the corresponding filter can be defined by the following vector equation

$$R(x, y) = \frac{\langle \vec{I} \bullet \vec{B} \rangle}{\|\vec{I}\| \|\vec{B}\|} \tag{3}$$

where $\vec{I}$ and $\vec{B}$ are the RGB vectors of the current and the background images. It is worth noting that this model operates in the RGB rather than the HSV space, therefore both luminance and chrominance can have an effect on the filter response. 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[6]. This model is based on the follow set of equations

$$c_1 = \arctan\left(\frac{R_i}{\max(G_i, B_i)}\right) \tag{4}$$

$$c_2 = \arctan\left(\frac{G_i}{\max(R_i, B_i)}\right)$$

$$c_3 = \arctan\left(\frac{B_i}{\max(R_i, G_i)}\right)$$

$$b_1 = \arctan\left(\frac{R_b}{\max(G_b, B_b)}\right)$$

$$b_2 = \arctan\left(\frac{G_b}{\max(R_b, B_b)}\right)$$

$$b_3 = \arctan\left(\frac{B_b}{\max(R_b, G_b)}\right)$$

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

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.

2.2 Neuro-Fuzzy Classifier

In practice, the combination of multiple filter responses can be problematic as in video sequences ambient environment can undergo constant changes. Although certain prior information of the filters is known, such as shadow usually induces only a slight colour change to the background, the uncertainties of this information limit its ability and effectiveness in designing the filter. Therefore, it requires a self-adaptive way of adjusting the relative weightings of different filters. One method to adaptively integrate the filters is through using a neural network to learn the responses toward different scenarios and to classify shadows accordingly. In addition, instead of feeding the filter outputs directly to the neural network, fuzzy sets are defined as the network inputs in order to model the filter responses and to include the imprecise prior information for shorten the learning process. To this end, a neuro-fuzzy classifier has been designed to incorporate the outputs of these filters for optimal shadow removal.

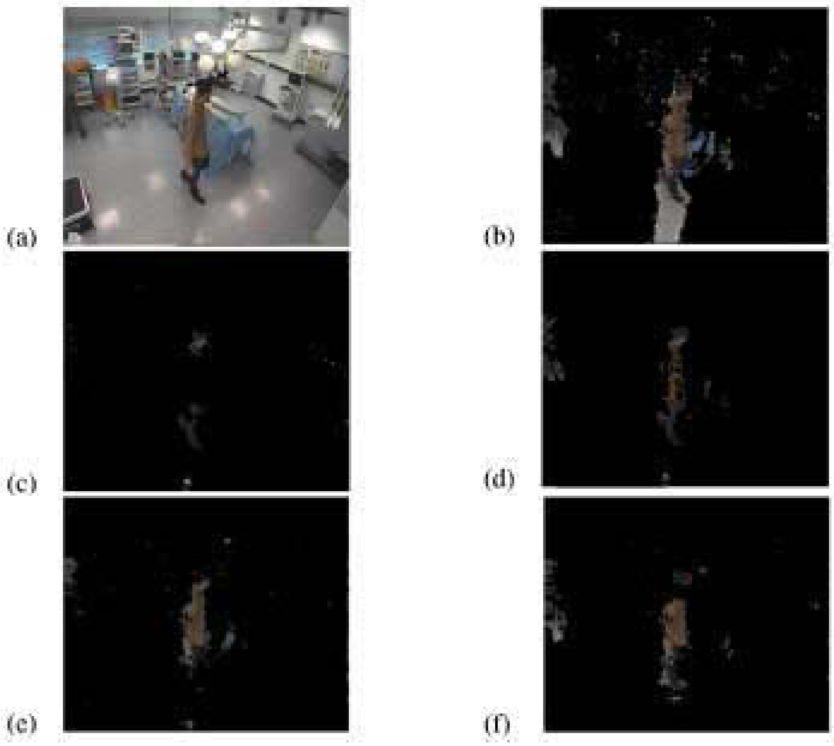


Fig. 3. The relative performance of different shadow filters. (a) The original image, (b) after background removal, (c-f) shadow removal by intensity difference, intensity attenuation, angle difference between RGB vectors, and colour difference, respectively

In this study, the four filters measure the differences between the current and the background image. As there is no discrete definition on big and small differences between the images, three linguistic meanings are defined to represent the 'low', 'medium' 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 π function, where the minimum and maximum range is learnt statistically from the incoming video stream.

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

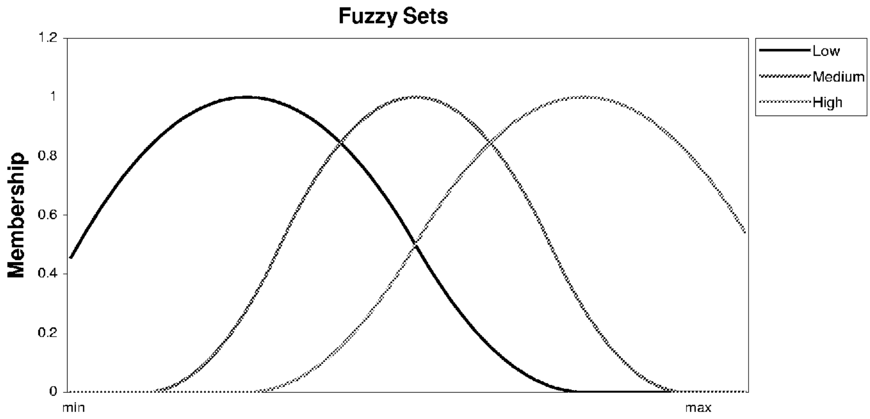


Fig. 4. The definition of the 'low', 'medium' and 'high' membership functions of the fuzzy sets. The "min" and "max" are the range for a particular filter output

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 are all "low", the corresponding pixel is a shadow pixel.*

2) *If the outputs of the filters are all "high", the corresponding pixel is an 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 an 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, the above rules are evaluated in real-time. If any of the rules is violated, the MLP is retrained through back-propagation. Since the number of nodes involved in the network is small, the adaptation of the network is therefore very fast. 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. The above contextual based training process, in fact, serves the purpose of real-time adjustment of the relative weights of the four filters used. Since there is no explicit thresholding involved, the proposed framework is applicable to different video scenes.

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

3 Results

Figure 6 demonstrates a MPEG sequence with a person walking inside an operating theatre. The top row shows the original image for object detection based on background removal. When the person approached the operating table, the overhead theatre light cast a strong shadow to the floor. The corresponding neuro-fuzzy shadow filtered results are listed on the middle row, demonstrating the effectiveness of the proposed technique. In addition, to indicate the improvement of the adaptability after using fuzzy sets to model the filter outputs, the bottom row shows the resulting images from the modified version of the shadow filter, which utilizes the normalized filter outputs as the neural network inputs. It is obvious that without the fuzzy representations, the filter tends to remove the shadows either excessively or insufficiently, because the neural network has not reach convergence. For the purpose of multiple-object tracking, Figure 7 illustrates a MPEG sequence recorded from a different camera showing two people walking towards each other. The detection result based on background removal, as shown on the top row images, mistakenly merged the two objects together due to shadows cast on the ground. After applying the proposed neuro-fuzzy filter, shadows were reliably removed and objects could be individually identified. The applicability of the proposed technique for different scene settings is demonstrated in Figure 8. None of the parameters of the neuro-fuzzy filter was altered for this video sequence, which illustrates three people walking along the platform of a railway station. The shadow in this case was caused by the sunlight and the corresponding shadow removed result is shown on the bottom row of Figure 8.

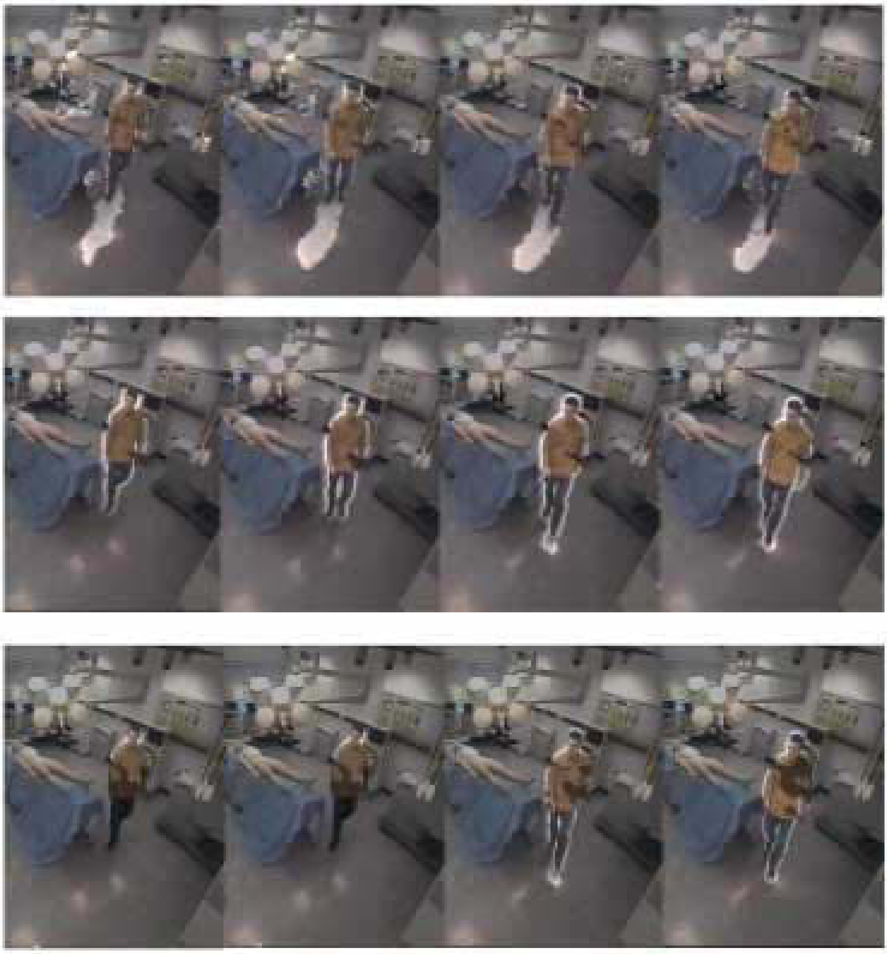


Fig. 6. (**Sequence 1**) An image sequence showing a person walking inside an operating theatre. Results of object identification based on background removal (top), neuro-fuzzy filter (middle), and standard neural network based filter (bottom), respectively

For quantitative analysis, Figure 9 shows the residual shadow pixels for the above three video sequences. An interactive image measurement program was developed for manually classify the moving objects. The residual shadow pixels were measured 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

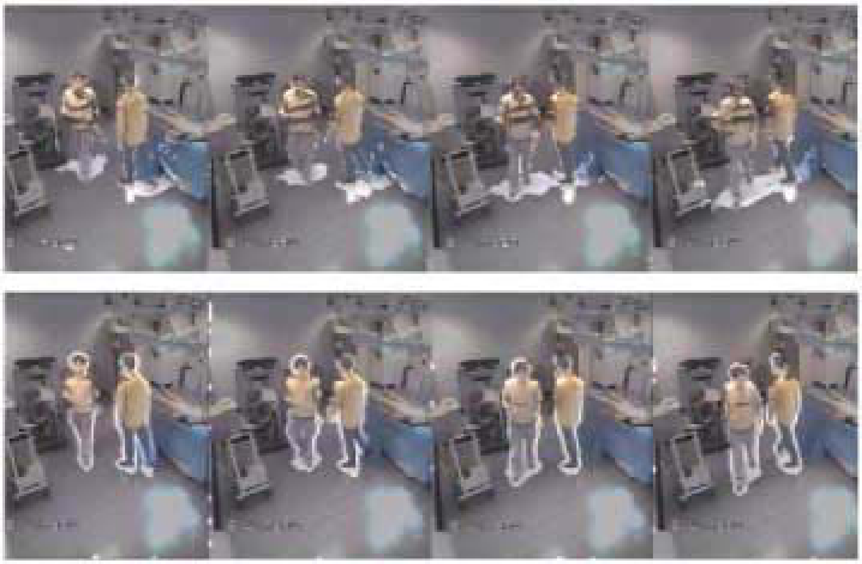


Fig. 7. (**Sequence 2**) An image sequence showing two people walking towards each other inside an operating theatre. Results of object identification based on background removal (top) and neuro-fuzzy filter (bottom), respectively

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

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.

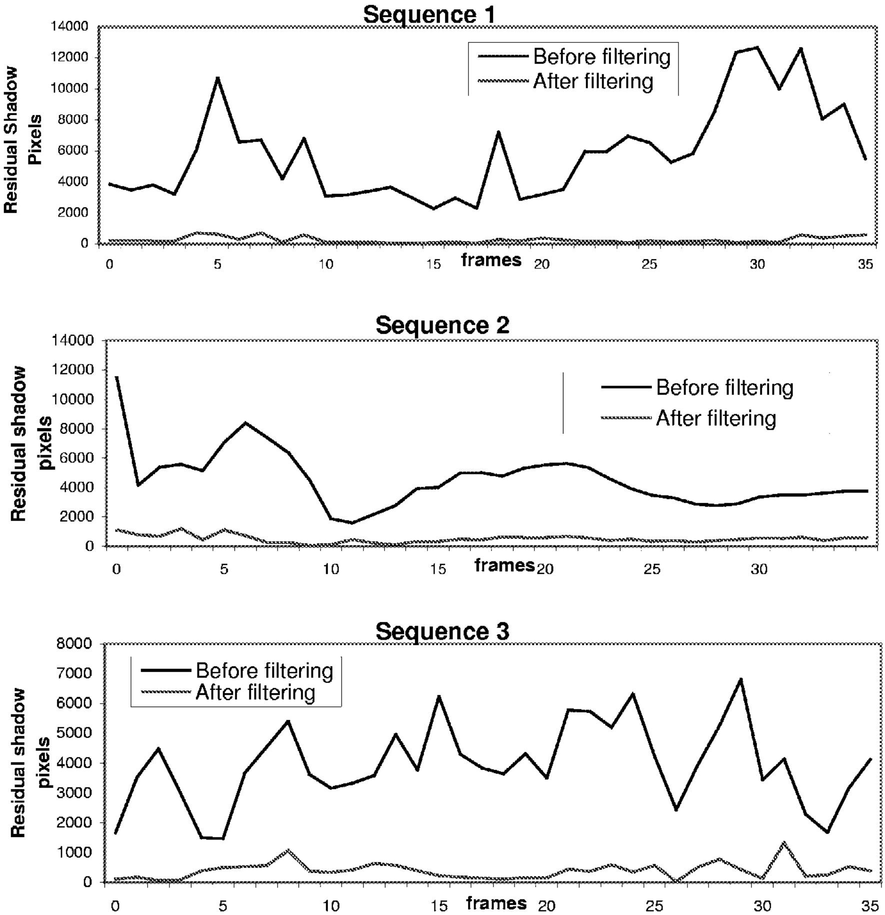
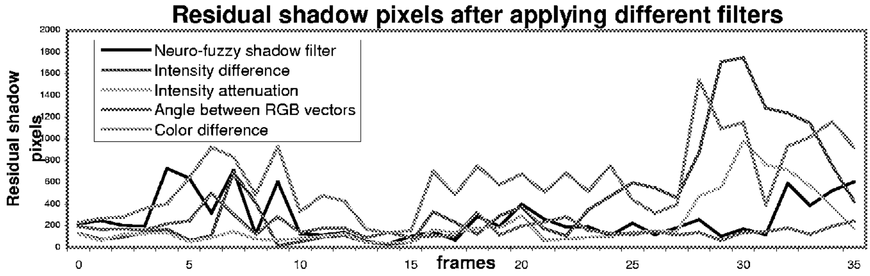


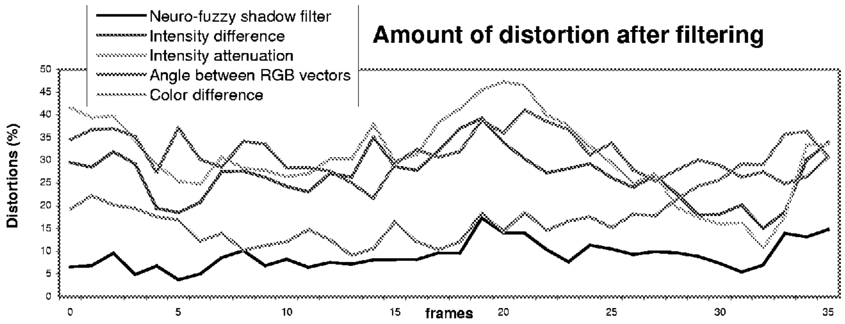
Fig. 9. Residual shadow pixels before and after applying different shadow filters for the three video sequences used in Figures 6,7 and 8

4 Conclusion

This paper presents an effective framework of combining shadow filters based on different visual characteristics. The use of MLP combined with fuzzy classification facilitates the dynamic weighting of different filter responses. Current experimental results have shown that the proposed technique works well for both indoor and outdoor cenes. The major benefit of the proposed framework is that it uses con-



(a)



(b)

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

textual information for real-time self-adaptation and training, and thus avoids the difficulty of creating example data sets for supervised learning. Since the proposed updating algorithm is generic, the method can be readily applied to different video sequences with varying environmental settings. The quantitative analysis demonstrates the strength of relying on the complementing nature of different methods of shadow removal. The proposed technique is simple to implement in practice, and can easily be incorporated into existing real-time video stream processing frameworks. It is also possible to further improve the performance of shadow removal by combining additional filters based on other visual characteristics and by adaptively adjusting the fuzzy representation of each filter to reflect the probability distribution of its responses.

Acknowledgement. The image sequences used for this study were provided by Mr. Lee Edwards of Orbb Ltd., and the EU project ADVISOR (IST1999-11287) and its partners, which includes Thales Research Ltd., Bull, Vigitec, INRIA, Reading University and Kingston University.

References

1. W. Lu and Y.P. Tan, "A Color Histogram Based People Tracking System", *IEEE Conf. On ISCAS 2001*, vol. 2, 137-140, 2001.
2. A. Prati, I. Mikić, C. Grana and M.M. Trivedi, "Shadow Detection Algorithms for Traffic Flow Analysis: a Comparative Study", *Proceedings of 2001 IEEE Intelligent Transportation Systems Conference Processings – Oakland (CA) USA*, 340-345, 2001.
3. I. Mikić, P.C. Cosman, G.T. Kogut and M.M. Trivedi, "Moving Shadow and Object Detection in Traffic Scenes", *IEEE Proceedings of the 15th International Conference on Pattern Recognition*, vol. 1, 321-324, 2000.
4. C, Jiang and M.O. Ward, "Shadow Identification", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1992 CVPR1992, 606-612, 1992.
5. P.L. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection", *Proceedings of the 6th British Machine Vision Conference*, 347-356, 1995.
6. 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.
7. M. Kilger. "A Shadow Handler in a Video-based Real-time Traffic Monitoring System", *Proceedings of the IEEE Workshop on Applications of Computer Vision*, 11-18, 1992.
8. Y. Sonoda and T. Ogata, "Separation of moving objects and their shadows, and application to tracking of loci in the monitoring images", *Proceedings of the 4th International Conference on Signal Processing 1998, ICSP '98*, vol. 2, 1261-1264, 1998.
9. K. Onoguchi, "Shadow Elimination Method for Moving Object Detection", *Proceedings of the 4th International Conference on Pattern Recognition 1998*, vol. 1, 583-587, 1998.