MOVING SHADOW DETECTION BY NONLINEAR TONE-MAPPING

Elad Bullkich[1], *Idan Ilan*[1], *Yair Moshe*[1], *Yacov Hel-Or*[2] *and Hagit Hel-Or*[3]

[1] Signal and Image Processing Laboratory (SIPL), Dept. of Electrical Engineering, Technion - Israel Institute of Technology, 32000 Haifa, Israel
[2] Google Inc. Mountain View, U.S.A.
[3] Dept. of Computer Science, University of Haifa, 31905 Haifa, Israel

Email: elad@bullkich.com, ilanidan@gmail.com, yair@ee.technion.ac.il, toky@idc.ac.il, hagit@cs.haifa.ac.il

ABSTRACT

Shadow detection is a well-known problem in image and video processing. Detecting moving shadows is useful in numerous applications such as object detection and tracking. Most works in this area are not suitable for shadow detection using a low-cost outdoor surveillance camera. In this work we suggest a fast shadow detection approach for video surveillance, by comparing each video frame to a continuously updated background image. We differentiate shadow areas from foreground areas by assuming the shadow pixels are associated with background pixels through a nonlinear tone mapping. This assumption is general and applies to various systems and scene conditions. A distance measure between patch images that account for nonlinear tone mapping is calculated by adapting a recently suggested approach for pattern matching termed Matching by Tone Mapping (MTM). We show that the proposed technique is computationally efficient and outperforms state-of-the-art shadow detection techniques in typical surveillance scenarios.

Index Terms— shadow detection, visual surveillance

1. INTRODUCTION

A growing interest in the image and video processing communities is the task of cast shadow detection. In applications such as video surveillance, traffic monitoring and human motion capture, good segmentation and tracking of foreground objects is a core requirement. Unfortunately, moving shadows in these applications may appear as foreground objects, when in fact they are caused by the interaction between light and objects. The inability to distinguish between foreground objects and shadows can cause severe problems such as object merging, false segmentation and identification failure, all of which significantly affect performance. Thus, shadow detection and removal is an important and necessary task.

As an active research area, various techniques for detecting shadows are described in the literature. These techniques can be divided into model-based and property-based techniques. Model-based techniques are based on a model of the scene, the illumination or the acquisition system. An example of this approach is the use of a geometry-based model of the observed surrounding [1]. These studies are usually designed for specific conditions and are difficult to generalize. Property-based techniques try to detect shadows in a more general way by using gradient, color or texture features that discriminate shadows from foreground objects or background. Such features are used to detect distortions in luminance and in chrominance [2], to exploit color differences between background and shadow in different color spaces [3] [4] or to statistically model changes in appearance of a shadowed pixel [5]. These approaches are hindered by requirements which are typically unsuitable for practical systems such as pre-calibration of camera and scene parameters. Furthermore performance of these approaches tend to deteriorate when dealing with shadows in surveillance cameras and outdoor scenes since such shadows are often non uniform and tend to have wide penumbras [6]. Some studies for detecting shadows are based on computing an illumination invariant features that rely on physical properties [7] [8]. Unfortunately, reliable computations of these invariants require high quality images. Furthermore, it was found that shadow invariant methods are sensitive to noise and may fail in outdoor scenes [8]. A comprehensive survey of shadow detection studies is presented in [9].

The work presented in this paper suggests overcoming the drawbacks of previous shadow detection techniques by making a very general assumption on the properties of a shadow. Similar to other approaches, we assume that the structural content (namely, textures and edges) in shadowed regions is preserved from the original non-shadowed scene since shadowing does not alter the background surfaces themselves, only their illumination. However, in contrast to previous studies that use a similar assumption, we assume the shadowing effect can be represented locally by any non-linear tone mapping of the background's gray levels. Unlike linear tone mappings, which can be dealt with a Normalized Cross Correlation (NCC) metric, a nonlinear mapping does not restrict the transformation within the shadowed areas of the image and can thus model any tone mapping of the gray levels (including non-monotonic mappings). We use a non-linear tone mapping invariant metric termed Matching by Tone Mapping (MTM), proposed by Hel-Or et al. [10], to evaluate distances between suspected foreground and background pixels. This distance metric yields a method for distinguishing foreground from shadowed pixels. The MTM metric compensates for the nonlinear mapping existing in the shadowed areas, resulting in small valued distances, while foreground pixels greatly differ from the background producing a large distance measure. The MTM measure is robust thus does not require tweaking of parameters for different scenes and can handle low quality images and complex lighting conditions

The remainder of this paper is organized as follows. In section 2 we present the MTM metric. In section 3 we justify our assumption and present the proposed shadow detection technique. Experimental results are described in section 4. Section 5 concludes our work.

2. MATCHING BY TONE MAPPING

In a recent study [10] Hel-Or et al. proposed the Matching by Tone Mapping (MTM) distance measure between image patches and presented an efficient way for computing it. The MTM distance is invariant to tone mappings existing between image patches, including nonlinear mappings. In this section we briefly describe the MTM distance measure. For further details, see [10].

We wish to evaluate the distance between two image patches under any possible tone mapping. Let $\boldsymbol{p} \in R^m$ and $\boldsymbol{w} \in R^m$ be two patches to be compared and denote by $\mathcal{M}: R \rightarrow R$ a tone mapping function, i.e. $\mathcal{M}(\boldsymbol{p})$ represents the tone mapping applied to each pixel in $\boldsymbol{p}$ independently. The MTM distance is the normed distance between one patch and a tone mapped second patch minimized over all possible tone mappings. It is defined as follows:

$$D(\boldsymbol{p}, \boldsymbol{w}) = \min_{\mathcal{M}} \left\{ \frac{\|\mathcal{M}(\boldsymbol{p}) - \boldsymbol{w}\|^2}{m \cdot \mathrm{var}(\boldsymbol{w})} \right\} \qquad (1)$$

The denominator is a normalization factor inducing the metric to be tone-scale invariant, i.e. $D(\boldsymbol{p}, \boldsymbol{w}) = D(\boldsymbol{p}, \alpha\boldsymbol{w})$ for any scalar α and avoiding a small distance if $\boldsymbol{w}$ is a flat patch. By its definition, MTM allows the measure to be insensitive to the brightness and contrast changes between the patches, while still capturing the structural differences between them. Note that this measure is not symmetric, thus an alternative definition is obtained by exchanging the roles of $\boldsymbol{p}$ and $\boldsymbol{w}$.

Obviously searching over all possible tone mappings is not practically feasible. Hel-Or [10] propose a method for computing the MTM measure very efficiently using a piecewise-constant mapping (PWC) approximation. The range of possible values of $\boldsymbol{p}$ is divided into k bins, and then k binary "slices" of $\boldsymbol{p}$ are produced, each of which representing the pixels belonged to the respective bin. Denoting the jth slice by $\boldsymbol{p}^j$, Eq. 1 can be efficiently minimized using a closed form solution [10]:

$$D(\boldsymbol{p}, \boldsymbol{w}) = \frac{1}{m \cdot \mathrm{var}(\boldsymbol{w})} \left[\|\boldsymbol{w}\|^2 - \sum_j \frac{1}{|\boldsymbol{p}^j|} (\boldsymbol{p}^j \cdot \boldsymbol{w})^2 \right] \qquad (2)$$

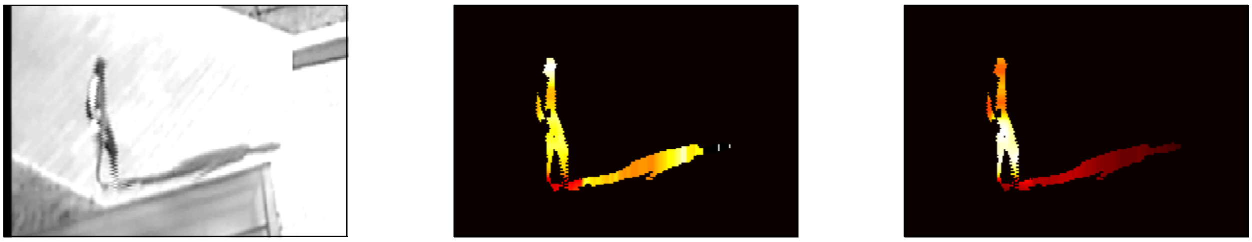
The numerator is now given as the difference between the norm of $\boldsymbol{w}$ and the sum of all k inner products of the $\boldsymbol{w}$ and the slices of $\boldsymbol{p}$.

3. SHADOW DETECTION

The proposed shadow detection technique operates as part of a foreground detection system in video. It detects shadows in the current frame and uses a background image computed by the system using any background subtraction technique (such as temporal median or Mixture of Gaussians) [11]. We assume a nonlinear tone mapping between shadowed pixel and background pixels, namely, the spatial neighborhood of a shadowed pixel in the current frame is a result of a nonlinear tone mapping of the corresponding spatial location in the background frame. An example of the validity of this assumption is depicted in Fig. 1 for one

(a) (b) (c)

Fig. 1: Example of tone mapping due to shadows. (a) Background image of the sequence *Sepm*, (b) Frame #208 of the sequence, (c) The tone mapping applied to shadowed pixels of (b) is nonlinear.



(a) (b) (c)

Fig. 2: (a) A video frame acquired from a low-cost surveillance camera and (b-c) the corresponding distance maps. Bright colors in the distance maps represent high values while dark colors represent low values. (b) NCC results in similar values for foreground and shadow while (c) MTM results in lower values for shadowed pixels compared to foreground pixels.

frame from the video sequence *Sepm* taken from [4].

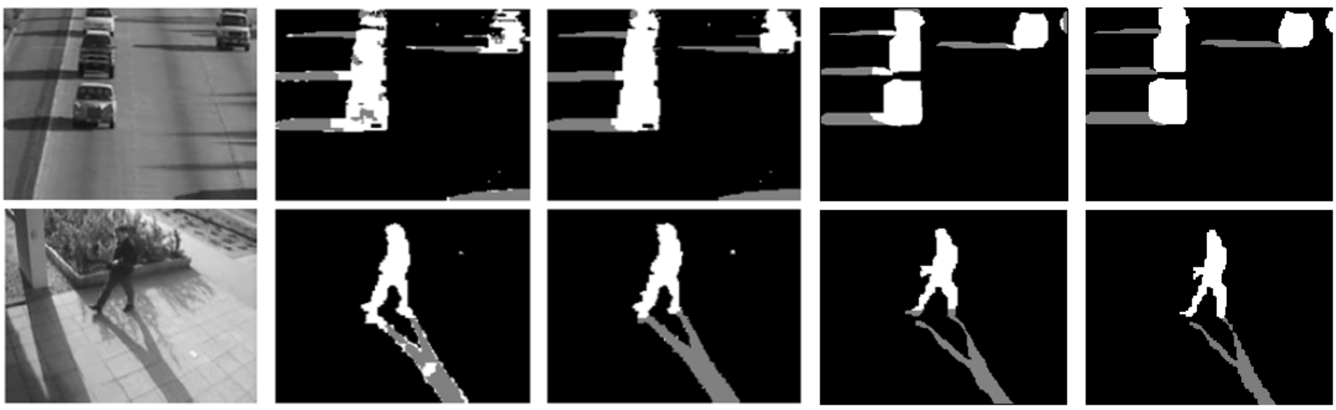
For shadow detection, the MTM measure is applied to a spatial neighborhood of every pixel of the current frame's suspected to be foreground regions. This results in an MTM "distance map" where small values relate to shadowed pixels whereas high values indicate a differently structured content, namely, foreground pixels. As noted above and in [10], the MTM distance is not symmetric and two alternatives are possible. We follow the MTM distance as given in Eq. 1 where suspected foreground patches $\boldsymbol{p}$ are tone mapped towards background patches $\boldsymbol{w}$. The choice of this variation gives better foreground detection results compared with the alternative MTM distance. Fig. 2 depicts the distance map calculated for a video frame acquired from a low-cost surveillance camera. The human figure is about 70×30 pixels and the video quality is low. In such a configuration the assumptions made by most shadow detection algorithms tend to fail. The figure shows that the MTM metric results in substantially lower values for shadowed pixels compared to foreground pixels. This result is compared with the Normalized Cross-Correlation (NCC) metric that fails to do so due to its false assumption of linear mapping between shadowed pixels and their corresponding background pixels.

In order to separate foreground objects from shadows, it is required to threshold the MTM distance map. Values above the threshold are considered as foreground and values below it are considered as shadow. The threshold is determined based on the fact that the distance map contains two classes of pixels – foreground pixels and shadowed pixels. Otsu's thresholding method [12] is used for calculating the optimum threshold for maximizing the intra-class variance.

4. RESULTS

To evaluate performance we compare our work to two shadow detection methods - Constant Ratio (CR) [13] and Statistical Shadow (SS) [4]. We use the *SZTAKI Benchmark Set* [4] [9]. This dataset contains video sequences with moving shadows in different scenarios, all of which have ground truth, and results of the CR and SS methods. Fig. 3 shows two video frames from the dataset and their corresponding shadow detection results according to the three shadow detection methods. It can be seen that the MTM's results are smoother and more accurate with respect to the given ground-truth.

Table 1 compares quantitative shadow detection results of these methods using measures of precision, recall, and *F*-measure. The *F*-measure takes into account the tradeoff



(a) (b) (c) (d) (e)

Fig. 3: Shadow detection results. (a) Frames from the *Highway* (upper) and *Seam* (lower) sequences, their shadow detection results using (b) constant ratio (CR) model of [13], (c) statistical shadow (SS) model of [4] and (d) the proposed method and their (e) ground-truth of [4] [9]. Gray represents the detected shadow and white representes foreground.

	Precision			Recall			*F*-measure		
Sequence	**CR**	**SS**	**MTM-PWC**	**CR**	**SS**	**MTM-PWC**	**CR**	**SS**	**MTM-PWC**
Highway	0.644	0.805	0.925	0.866	0.890	0.914	0.746	0.845	**0.920**
Seam	0.596	0.774	0.971	0.946	0.968	0.947	0.731	0.861	**0.959**
Senoon	0.742	0.833	0.935	0.980	0.963	0.953	0.845	0.894	**0.944**
Sepm	0.621	0.830	0.908	0.972	0.961	0.914	0.756	0.891	**0.911**

Table 1: Quantitative evaluation results: CR [13], SS [4], MTM-PWC (proposed). The proposed method substatially outperforms both CT and SS for all 4 video sequences.

between precision and recall scores by computing the harmonic mean of precision and recall. The results given in Table 1 show that the proposed technique substantially outperforms the CR and SS methods, yielding a significant advantage in terms of the *F*-measure. The proposed technique results in about 21% increase in the *F*-measure compared with the CR method and about 7% increase compared with the SS method on average. The proposed technique has a small number of parameters. The most important parameters are the patch size and the number of bins into which the image is sliced to. All results were obtained using 255 bins, i.e., each image slice represented a single gray level. The optimal patch size was found to be about the same size of the objects in the scene. In our tests the algorithm was found to be highly robust to both patch size and number of bins.

5. CONCLUSIONS

In this paper we present a novel moving shadow detection technique. The technique assumes nonlinear tone mapping between shadows and background and uses the Matching by Tone Mapping (MTM) approach for efficiently comparing patches of video frames with their corresponding background patches. The proposed technique has low computational complexity, is robust, and was shown to substantially outperform state-of-the-art shadow detection techniques in typical surveillance scenarios.

6. ACKNOWLEDGMENTS

We would also like to thank the staff of the Signal & Image Processing Lab (SIPL) for their support, and especially Prof. David Malah, head of SIPL and Mr. Nimrod Peleg, SIPL chief engineer.

7. REFERENCES

[1] A. Yoneyama et al., "Moving Cast Shadow Elimination for Robust Vehicle Extraction based on 2D Joint Vehicle/Shadow Models," in *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, 2003, p. 229.

[2] T. Horprasert et al., "A Statistical Approach for Rreal-time Robust Background Subtraction and Shadow Detection," in *Proc. ICCV Frame-rate Workshop*, 1999.

[3] C. Benedek and T. Sziranyi, "Bayesian Foreground and Shadow Detection in Uncertain Frame Rate Surveillance Videos," *IEEE Trans. Image Processing*, vol. 17, no. 4, pp. 608-621, April 2008.

[4] C. Benedek and T. Sziranyi, "Shadow Detection in Digital Images and Videos," in *Computational Photography: Methods and Applications*, Rastislav Lukac, Ed. Boca Raton, FL, USA: CRC Press, 2011, ch. 11, pp. 283-308.

[5] I. Mikic et al., "Moving Shadow and Object Detection in Traffic Scenes," in *Proc. IEEE Conf. on Pattern Recognition (ICPR)*, 2000, pp. 321-324.

[6] E. Arbel and H. Hel-Or, "Shadow Removal Using Intensity Surfaces and Texture Anchor Points," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, no. 6, pp. 1202-1216, 2011.

[7] G. D. Finlayson et al., "Removing Shadows from Images," *Lecture Notes in Computer Science*, pp. 823-836, 2002.

[8] E. Salvador et al., "Cast Shadow Segmentation Using Invariant Color Features," *Computer Vision and Image Understanding*, vol. 92, no. 2, pp. 238-259, August 2004.

[9] A. Prati et al., "Detecting Moving Shadows: Algorithms and Evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 25, no. 7, pp. 918-923, 2003.

[10] Y. Hel-Or et al., "Fast Template Matching in Non-linear Tone-mapped Images," in *Proc. IEEE Intl. Conf. on Computer Vision (ICCV)*, Nov. 2011. [Online]. http://www.faculty.idc.ac.il/toky

[11] M. Piccardi, "Background Subtraction Techniques: A Review," in *Intl. Conf. on Systems, Man and Cybernetics*, 2004.

[12] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62-66, Jan. 1979.

[13] J. Stander et al., "Detection of Moving Cast Shadows for Object Segmentation," *IEEE Trans. on Multimedia*, vol. 1, no. 1, pp. 65-76, March 1999.