ONLINE SUNFLICKER REMOVAL USING DYNAMIC TEXTURE PREDICTION

A. S. M. Shihavuddin, Nuno Gracias and Rafael Garcia
Computer Vision and Robotics Group, Universitat de Girona, Girona, Spain

Keywords: Sunflicker, Dynamic Texture, Illumination Field, Shallow Underwater Images.

Abstract: An underwater vision system operating in shallow water faces unique challenges, which often degrade the quality of the acquired data. One of these challenges is the sunflicker effect, created from refracted sunlight casting fast moving patterns on the seafloor. Surprisingly few previous works exist to address this topic. The best performing available method mitigates the sunflickering effect using offline motion compensated filtering. In the present work, we propose an online sunflicker removal method targeted at producing better registration accuracy. The illumination field of the sunflicker effect is considered as a dynamic texture, since it produces repetitive dynamic patterns. With that assumption, the dynamic model of the sunflicker is learned from the registered illumination fields of the previous frames and is used for predicting that of the next coming frame. Such prediction allows for removing the sunflicker patterns from the new frame and successfully register it against previous frames. Comparative results are presented using challenging test sequences which illustrate the better performance of the approach against the closest related method in the literature.

1 INTRODUCTION

In the field of underwater image sensing, new and complex challenges need to be addressed like light scattering, sunflicker effects, color shifts, shape distortion, visibility degradation, blurring effects and many others. The research work presented in this paper deals with the presence of strong light fluctuations due to refraction, commonly found in shallow underwater imaging. Refracted sunlight generates dynamic patterns, which degrade the image quality and the information content of the acquired data. The sunflickering effect creates a difficult challenge to the scientists to understand and to interpret the benthos. Development of an online method to completely or partially eliminate this effect is a prerequisite to ensure optimal performance of underwater imaging algorithms.

2 RELATED WORK

Motivated by the work by on effective shadow removal method (Weiss, 2001) in land scenes and the work on shadow elimination (Matsushita et al., 2002) on video surveillance, Schechner and Karpel developed an approach (Schechner and Karpel, 2004a) to solve the sunflicker removal problem based on the observation that the spatial intensity gradients of the caustics tend to be sparse. Under that assumption, doing the temporal median over the gradients of a small number of images, a gradient field was obtained where the illumination effect was greatly reduced. The flicker free image was obtained by integrating the median gradient field. This sunflicker removal method (Schechner and Karpel, 2004a) did not take into consideration the camera motion for which registration inaccuracies are likely to appear. Using the method by Sarel and Irani (Sarel and Irani, 2004; Ukrainitz and Irani, 2006) two transparent overlapped video can be separated using the information exchange theory, if one of the video mentioned contains a repetitive dynamic sequence. This method (Sarel and Irani, 2004) claims to work for the repetitive dynamic sequence with variation in each repeated cycle, though the variation amount in the repetitive dynamic sequence is not quantitatively defined. On the other hand, a large number of frames would be required to grab a complete cycle of a repetitive dynamic sequence, making it impractical for moving cameras. An interesting approach to this problem can be to use polarization information (Schechner and Karpel, 2004b), to improve visibility underwater and in deep space. The refracted sunlight underwater has unique polarization characteristics. The exploitation

S. M. Shihavuddin A., Gracias N. and Garcia R..
ONLINE SUNFLICKER REMOVAL USING DYNAMIC TEXTURE PREDICTION.
DOI: 10.5220/0003811901610167
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 161-167
ISBN: 978-989-8565-03-7

of such characteristics is possible, but requires special cameras with polarized filters or imaging sensors, which are not commonly deployed. Another way of looking in to the same problem can be to recover the low rank matrix (Candès et al., 2009) which presents the illumination field. But this method only works when there is no camera motion which in our case is impractical.

The work by Gracias et al. (Gracias et al., 2008) can be considered as the state of art among the methods that have addressed the removal of sunflicker in shallow underwater images. This motion compensated filtering approach accounted for the basic requirement that the camera should be allowed to move during acquisition, thus being more adequate to real optical survey work. The method is based on the assumption that video sequences allow several observations of the same area of the seafloor, over time. If the assumption is valid, it is possible to compute the image difference between a given reference frame and the temporal median of a registered set of neighboring images. The most important observation is that this difference will have two components with separable spectral content. One is related to the illumination field (which has lower spatial frequencies) and the other to the registration inaccuracies (mainly having higher frequencies). The illumination field can be approximately recovered by using low–pass filtering. The main limitation is that the median image for each frame is obtained from both past and future frames, thus making it non causal.

3 APPROACH

The presented method uses dynamic texture modeling and synthesizing (Doretto et al., 2003), using principal component analysis, to predict the sunflicker pattern of the next frame from the previous few frames. The prediction is then used to coarsely remove the illumination field from the current working frame. The presented approach attains higher registration performance even under heavy illumination fluctuation. Also this method is strictly causal (*i.e.* does not rely on future observations), fulfilling the important condition of online operation required for visual-based robot navigation.

3.1 Dynamic Texture Modeling and Synthesizing

The Open Loop Linear Dynamic system (OLLDS) model (Doretto et al., 2003) is used to learn the dynamic model of sunflicker illumination pattern and to synthesize it for unknown cases. Once modeling is done, the OLLDS can potentially be used for extrapolating synthetic sequences of any duration at negligible computational cost. The underlying assumption in this approach is that, the individual images are realizations of the output of a dynamical system driven by an independent and identically distributed (IID) process.

For learning the model parameters, the OLLDS model uses one of two criteria: total likelihood or prediction error. Under the hypothesis of second-order stationarity, a closed-form sub-optimal solution of the learning problem can be obtained as follows (Doretto et al., 2003).

1. A linear dynamic texture is modeled as an auto-regressive moving average process (ARMA) with unknown input distribution, in the form,

$$\begin{array}{lll} x(t+1) = Ax(t) + Bv(t) & v(t) \approx N(0,Q) & x(0) = x_o \\ y(t) = Cx(t) + w(t) & w(t) \approx N(0,R) & \end{array}$$

where y(t) belongs to R_n is the observation vector; x(t) belonging to R_r, is the hidden state vector (r is much smaller than n); A is the system matrix; C is the output matrix and v(t), w(t) are Gaussian white noises. Here y(t) presents the noisy output, in this case the image sequences.

2. Taking the SVD of y(t), C can be found.

$$Y_1^\tau = U\Sigma V^T$$
$$C(\tau) = U \qquad X(\tau) = \Sigma V^T$$

3. A can be determined uniquely by,

$$A(\tau) = \Sigma V^T D_1 V (V^T D_2 V)^{-1} \Sigma^{-1}$$

Where

$$D_1 = \begin{bmatrix} 0 & 0 \\ I_{\tau-1} & 0 \end{bmatrix} \qquad D_2 = \begin{bmatrix} I_{\tau-1} & 0 \\ 0 & 0 \end{bmatrix}$$

Also, Q and B can be found by

$$Q(\tau) = \frac{1}{1-\tau} \sum_{i=1}^{\tau-1} v(i) v^T(i)$$
$$BB^T = Q$$

3.2 Motion Compensated Filtering

Let us consider a set of registered images. We refer to a given image by the discrete parameter i which indexes the images temporally. The radiance L of a given pixel with coordinates (x, y) can be modelled as

$$L_i(x,y) = E_i(x,y) \cdot R_i(x,y)$$

where E_i is the irradiance of the sunlight over the 3D scene at the location defined by pixel (x, y) at time i, after absorption in the water. R(x, y) is the bidirectional reflectance distribution function. For underwater natural scenes, where diffuse reflectance models are applicable, R is assumed independent of both light and view directions.

Converting the expression for L_i to a logarithmic scale allows the use of linear filtering over the illumination and reflectance.

$$l_i(x,y) = e_i(x,y) + r_i(x,y)$$

For approximately constant water depth, and for realistic finite cases, the median converges significantly faster to average value of l_i than the sample mean.

$$I_{med}(x,y) = med_{[i0,i1]} I_i(x,y) \approx e + r_{med}(x,y)$$

Here r_{med}(x, y) stands for an approximation to the median of reflectance. The difference d_i(x, y) of a given image l_i(x, y) with the median radiance l_{med}(x, y) is used to recover the approximate background image.

$$d_{l_i}(x,y) = l_i(x,y) - I_{med}(x,y)$$

$$\approx (e_i(x,y) - e) + (r_i(x,y) - r_{med}(x,y))$$

This difference d_{l_i}(x,y) has two main components. The first component relates to the instant fluctuation of the illumination field and has lower spatial frequencies. This component will have positive values in the over-illuminated areas where there is convergence of the refracted sunlight and will have negative values in the areas where the sunlight is diverted away. The second component relates to inaccuracies in the image registration and has higher spatial frequencies. After applying a low pass filter on the difference image, the low frequency components which resembles the illumination field is kept in the output. This approximated illumination field is later used to correct the main input image and recover a flicker–free image.

4 PROPOSED ALGORITHM

In the online sunflicker removal method proposed in this work, the approximate illumination field of the current frame is found from the dynamic texture prediction step. The homography is mainly calculated between the temporary recovered current image (using the approximate illumination of the current frame) and the last recovered image. Finally, the new image can be recovered by using a median image calculated from the current and previous few images warped with respect to the current image using the calculated homography. The median image and the current images can be used to find the difference image for the current frame. The difference images after being filtered through a low pass filter only holds the low frequency components which represents the correct illumination field. Using this correct illumination field the working image is recovered from the sunflicker effect. The steps are demonstrated as in fig. 1

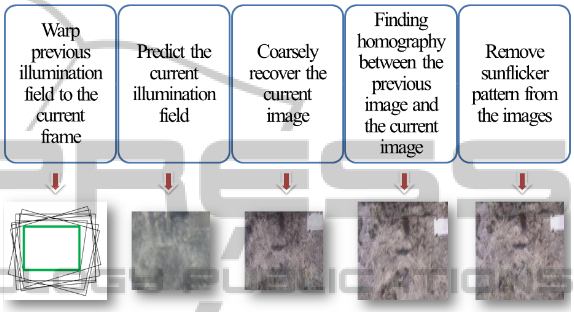


Figure 1: Step by step flow diagram of the proposed online method for sunflicker removal. From left to right, (1)warping previous illumination field to the current frame, (2)predicting the current illumination field, (3)coarsely recovering the current image, (4)finding homography between the current and the previous frame and (5)removing sunflicker pattern from the image using the calculated homography.

The proposed approach considers the following mentioned assumption to be valid

- Illumination field is a dynamic texture
- Camera movement in the video sequence is smooth
- Bottom of the sea is approximately flat

The main algorithm contains several steps which are presented next, using the following nomenclature:

- $I_{0,k}$ - Original input image obtained at time instant k
- $I_{R,k}$ - Recovered image obtained from $I_{0,k}$ after sunflicker removal
- H_k - low pass filtered version of the difference image at time k, which is used as estimate of the illumination field at time k
- $I_{M,k}$ - Median image obtained from N frames after being warped into the frame of image $I_{0,k}$
- N - the number of images present in the learning sequence.
- $M_{k,k-1}$ - Homography relating the image frames at time k and $k-1$.

The algorithm comprises the following steps:

1. Apply the motion compensated filtering method (Gracias et al., 2008) to register and recover

first few images from the sun flickering effects. In our implemented system, we used first the 25 frames for this step.

2. Get the new image in the sequence $I_{0,k}$ assuming that the previous images $I_{0,k-1}, \ldots, I_{0,k-N}$ have been recovered after sunflicker removal ($I_{R,k-1}, \ldots, I_{R,k-N}$). Advance time k (i.e. all previous data structures that had index k now have index $k-1$).

3. Predict the flicker pattern by

- Warping all the filtered version of the difference images, $H_{k-1}, \ldots, H_{k-N}$ with respect to the current frame $I_{0,k}$ to be recovered , assuming that $M_{k,k-1} \approx M_{k-1,k-2}$. All other previous homographies were obtained from actual image matches and thus known previously.
- Learning the sunflicker pattern from the registered filtered difference images. After registration, the motion of the camera is being compensated. In the learning phase, all the difference images of the previous frames H ($H_{k-1}, \ldots, H_{k-N}$) are converted into a column matrix. Using the array of all the registered H ($H_{k-1}, \ldots, H_{k-N}$) a large matrix called W_{t-1} is created having P rows and N columns. P is the number of pixels per frame and N is the total number of frames in the learning sequence.
- Predicting the $\hat{H}_k$ using the learned model. For learning, open loop linear dynamic model (Doretto et al., 2003) is being used. In this step the last frame H_{k-1} of the learned sequence is considered as the first frame for the synthesize part.

4. Create the correction factor $\hat{C}_k$ using the predicted low pass filtered version of the difference image $\hat{H}_k$ and the approximate difference image $\hat{I}_{d,k}$. To find the approximate difference image, the previously recovered image $I_{R,K-1}$ is warped into the current frame $I_{0,k}$ position using the last homography. Using the warped portion of the previous recovered image and the rest from current original image, a approximate median image is created, which is fused into the system to find the approximate difference image, $\hat{I}_{d,k}$ of the current frame. Using this approximated difference image $\hat{I}_{d,k}$ and the predicted $\hat{H}_k$, the correction factor $\hat{C}_k$ is found.

5. Apply predicted correction factor to $I_{0,k}$. Using the correction factor $\hat{C}_k$ found in the last step, the current image is approximately recovered from sunflicker effect. This recovered image is denoted by $\hat{I}_{R,k}$.

6. Perform image registration between $\hat{I}_{R,k}$ and $I_{R,k-1}$. From this obtain the real $M_{k,k-1}$.

7. Update $I_{M,k}$. Using the motion compensated filtering method (Gracias et al., 2008) create a median image for the current frame using the last few original frames. In this case, use the $\hat{I}_{R,k}$ to do the registration during finding the current median image, $I_{M,k}$.

8. Obtain final $I_{R,k}$. Using $I_{M,k}$ find the real difference image for the current frame, $I_{d,k}$ and the correct sunflicker pattern H_k and later the correct recover image $I_{R,k}$ removing the sunlight properly.

9. Go to step 2 and do the same for the next frames

The image registration is performed using the classic approach of robust model based estimation using Harris corner detection (Harris and Stephens, 1988) and normalized cross–correlation (Zhao et al., 2006). This method proved more resilient to strong illumination differences when compared with the result doing the same with SIFT (Lowe, 2004). Furthermore, the operation is considerably faster since the search areas are small.

We assume the knowledge of the gamma values for each color channel. For unknown gamma values one can apply blind gamma estimation. An efficient method is described in (Farid, 2001; Farid and Popescu, 2001), which exploits the fact that gamma correction introduces specific higher-order correlations in the frequency domain. Having the gamma values we transform the intensities to linear scale. After the deflickering the final output images are transformed into the sRGB space with the prescribed gamma value.

The steps above are applied over each color channel independently. Strong caustics lead to overexposure and intensity clipping in one or more of the color channels, resulting in chromaticity changes in the original images. These clippings typically affect different regions of the images over time, given the non-stationary nature of the caustics. The median is not affected by such transient clippings, whereas the average is. The low pass filtering is performed using a fourth order Butterworth filter (Kovesi, 2009), with a manually adjusted cutoff frequency.

Due to the camera motion, the stack of warped difference images described in step 3, may not cover the entire area of the current frame. If one considers the whole area of the current frame, this creates a condition of missing data in the W matrix for PCA. To circumvent this condition, only the maximum sized area of the current frame which is present in each warped frame at current frame location, is considered.

5 SELECTED RESULTS

The performance evaluation of the proposed system was done comparing with the previously available offline method (Gracias et al., 2008) by evaluating both on several test datasets of shallow water video sequences having distinct refracted sunlight conditions. The main evaluation criterion is the number of inliers found per time–consecutive image pair in each registration step. This criterion was found to be a good indicator the image de–flickering performance. The better the sunflicker removal is achieved, the larger the number of found inliers per time–consecutive image pair, considering all the other influencing factors constant.

5.1 Grounding Sequence

The first test sequence was obtained from a ship grounding survey and contains illumination patterns of relatively low spatial frequency. Fig. 2 shows an example of how in each step the image is coarsely recovered, registered and finally cleaned. The top left and bottom left parts of the image 2 show the original input image and real illumination field accordingly. Using this initial estimate, the warped illumination field is learnt and predicted for the current frame; the new image is temporarily cleaned with the prediction of illumination field and registered with respect to the last recovered image. In the bottom middle of Fig. 2, the predicted illumination field is shown and the upper middle part of the Fig. 2 represents the intermediate condition of the illumination field. The prediction illumination fields approximates the real illumination field with some spatial displacements and intensity variation. On the top right of Fig 2 shows the final recovered image and the bottom right of Fig 2 shows the final median image. Some examples of the finally recovered images for difference video sequences are given in Figs. 3(b), 3(d) and 3(f).

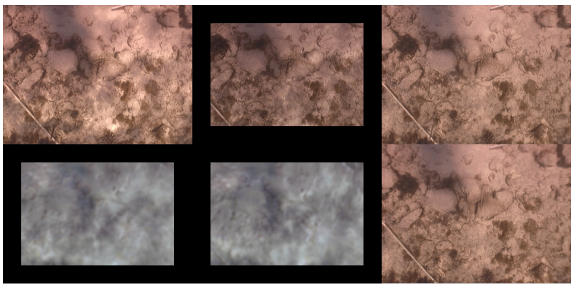
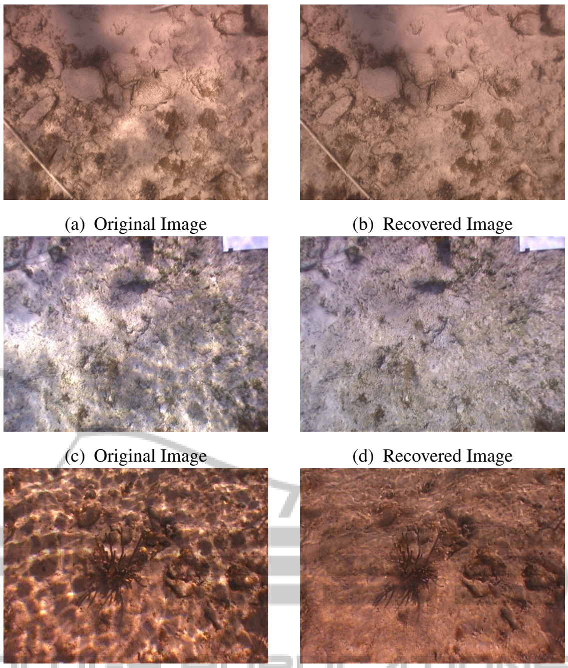


Figure 2: Step by step results for the *grounding* sequence. The top left is the original input image and bottom left is the original illumination field in the input image. In the top middle, the image is of the temporary recovered image, $\hat{I}_{R,k}$ and the bottom middle is the predicted illumination field. This was done using the predicted illumination field as shown in the bottom middle. On the top right is the finally recovered image, $\hat{I}_{R,k}$ and on the bottom right is the final median image.

In Fig. 4 the proposed online method and the existing offline method (Gracias et al., 2008) are being compared in terms of the number of inliers during the registration step. From the graph, it can be seen that the new method is outperforming the old method in



(a) Original Image (b) Recovered Image

(c) Original Image (d) Recovered Image

(e) Original Image (f) Recovered Image

Figure 3: Illustration of the online sunflicker removal performance.

almost all the frames. Quantitatively, matching performance is improved by 46% based on the inliers detected in every pair of frames. Without prediction, the registration or the number of found inliers are varying very rapidly because of complete dependency on the camera motion. In the case of prediction, it includes the displacement errors creating a generalized solution which provides approximately constant outcomes.

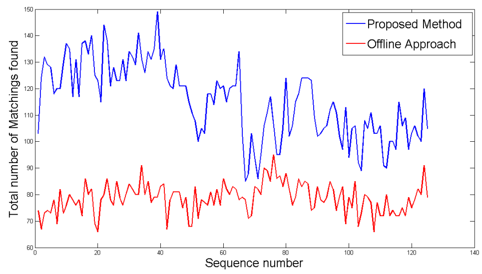


Figure 4: Comparison between proposed method and the offline method for grounding sequence.

5.2 Rock Sequence

The Rock sequence is more challenging than the grounding sequence, captured in shallow waters having a very rocky bottom. Both the methods are being tested for sunflicker removal keeping other parameters unchanged. The proposed online method is also performing significantly better than the offline method in terms of found inliers per registration steps as shown in Fig. 5. Overall gain in the matching performance in the proposed method is about 67% compared with the offline method.

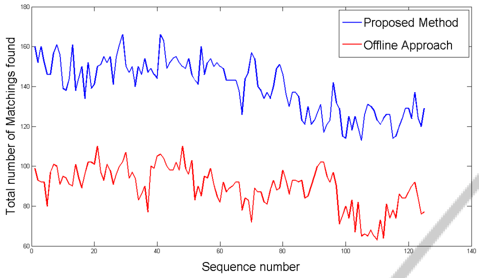


Figure 5: Comparison between proposed method and the offline method for Rock sequence.

5.3 Andros Sequence

This is the most challenging sequence of the three, captured in very shallow waters of less than 2 meters depth under intense sunlight. In this case, the illumination patterns have simultaneously very high spatial and temporal frequencies which leads to the method of (Gracias et al., 2008) to fail. In such situation it is very hard to get a good registration performance. In the Fig. 6, the graph shows that the proposed method performs 15% better. This performance is achieved because of the steady motion of the camera in this particular sequence. It proves the robustness of the online method achieving better results even in such a difficult video sequence. In case of large registration errors, the prediction model performance degrades rapidly.

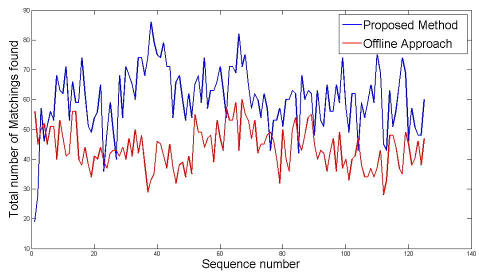


Figure 6: Comparison between proposed method and the offline method for Andros sequence.

6 CONCLUSIONS AND FUTURE DIRECTION

This work addresses the specific problem of sunflicker in shallow underwater images, by presenting a method which is suited for de–flickering on the fly. In the current online method, only the previous few frames with the current frame are used to create median image. In this case, the homography is calculated by registering the sunflicker removed version (using prediction from dynamic texture model learned from the last few frames) of current image with the last flicker–free image. This results in higher image registration accuracy than the offline method where the registration is carried out over the original images affected by the illumination patterns. The better registration then reflects in better median images estimation and ultimately in better sunflickering correction.

This research was motivated by the fact that the flicker–free images will help creating more accurate mosaics of the sea floor. Currently the method is implemented in Matlab, and the code has not been optimized for speed. It takes 6.39 seconds per frame, on average, when executed on an Intel core 2 Duo 2 GHz processor. However, being an on-line approach, the method has the potential to be implemented for real–time operation.

An extension to the current work is to relate the illumination frequency with the number of frames required to perform sunflicker removal optimally. It can be a way to know beforehand the minimum frame rate of the camera required to remove the sunflicker effect properly. Also, for instrumented imaging platforms, the camera motion can be estimated using a motion sensor, such as a rate gyro or an accelerometer. This estimate can be used during the initialization phase of the method, or whenever the image registration is not possible.

Another extension addresses highly slanted camera configuration, where the illumination patterns present distinct spatio-temporal frequencies for different regions of the image. In the described method, the cutoff frequency of the the low-pass filter is constant. As future work, this cutoff frequency will be made adaptive to the spatial location and estimated from image data.

ACKNOWLEDGEMENTS

This work was partially funded by the Spanish MCINN under grant CTM2010-15216, and by the EU Project FP7-ICT-2009-248497. ASM Shihavuddin and Nuno Gracias were supported by the MCINN under the FI and Ramon y Cajal programs.

REFERENCES

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2009). Robust principal component analysis? *CoRR*, abs/0912.3599.

Doretto, G., Chiuso, A., Wu, Y. N., and Soatto, S. (2003). Dynamic textures. *International Journal of Computer Vision*, 51:91–109.

Farid, H. (2001). Blind inverse gamma correction. *Image Processing, IEEE Transactions on*, 10(10):1428 –1433.

Farid, H. and Popescu, A. C. (2001). Blind removal of image non-linearities. *Computer Vision, IEEE International Conference on*, 1:76.

Gracias, N., Negahdaripour, S., Neumann, L., Prados, R., and Garcia, R. (2008). A motion compensated filtering approach to remove sunlight flicker in shallow water images. In *OCEANS 2008*.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Fourth Alvey Vision Conference, Manchester, UK*, pages 147 –151.

Kovesi, P. D. (2009). Matlab and octave functions for computer vision and image processing. *School of Computer Science Software Engineering*.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.

Matsushita, Y., Nishino, K., Ikeuchi, K., and Sakauchi, M. (2002). Shadow elimination for robust video surveillance. *Motion and Video Computing, IEEE Workshop on*, page 15.

Sarel, B. and Irani, M. (2004). Separating transparent layers through layer information exchange. In Pajdla, T. and Matas, J., editors, *Computer Vision - ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 328–341. Springer Berlin / Heidelberg.

Schechner, Y. and Karpel, N. (2004a). Attenuating natural flicker patterns. In *IEEE TECHNO-OCEAN, 04*, volume 3, pages 1262 –1268.

Schechner, Y. and Karpel, N. (2004b). Recovering scenes by polarization analysis. In *OCEANS '04. MTTS/IEEE TECHNO-OCEAN '04*, volume 3, pages 1255 –1261 Vol.3.

Ukrainitz, Y. and Irani, M. (2006). Aligning sequences and actions by maximizing space-time correlations. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Computer Vision ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 538–550. Springer Berlin / Heidelberg.

Weiss, Y. (2001). Deriving intrinsic images from image sequences. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 68 –75.

Zhao, F., Huang, Q., and Gao, W. (2006). Image matching by normalized cross-correlation. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2.