Unveiling Contrast in Darkness

Yuen Peng Loh   Chee Seng Chan

Centre of Image & Signal Processing, Fac. Comp. Sci. & Info. Tech.,
University of Malaya, Malaysia

`lohyuenpeng@siswa.um.edu.my   cs.chan@um.edu.my`

Abstract

At nighttime, reduced visibility could cause foreground and background images to appear to blend together. However, ambient light is always present in the natural environment, and as a consequence, it creates some contrast in darkness. In this paper, we formulate a visual analytic method that automatically unveils the contrast of dark images (i.e. nighttime), revealing the "hidden" contents. We utilize the traits of image representations obtained from computer vision techniques through a learning based inversion algorithm, eliminating the reliance on night vision camera and at the same time minimizing the need for human intervention (i.e. manual fine-tuning the gamma correction using Adobe Photoshop software). Experiments using the new Malaya Pedestrian in the Dark (MyPD) dataset that we have collected from the website Flickr, and in comparison with conventional methods such as image integral and gamma correction, it shows the efficacy of the proposed method. Additionally, we show the potential of this framework in applications that could benefit public safety.

1. Introduction

Biologically, human vision drops as the day approaches night, making it difficult for a regular person to locate, detect or recognize objects or people in the dark of the night. For instance, can you tell which image in Fig. 1 contains person(s) and their location? Even when these images are enlarged, it is still very difficult to notice the content of the images, nor to find or locate the person(s). As a consequence, the risk of accidents, especially while moving or driving, will increase. Moreover, based on the statistics provided by the Philadelphia Police Department and London Home Office, the lack of visibility lets crime to take place inconspicuously where most murders occur at nighttime from 8 p.m. to 12 a.m [7, 8, 16].

Inspired by Allen et al. [1], who stated that "*Humans do not see light but contrast*", it is eventually the lack of contrast between objects and the environment that hinders
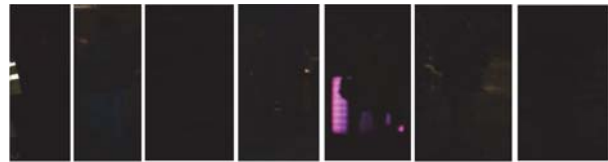

Figure 1. Example of RGB images captured at nighttime. Can you guess which image contains person(s), as well as their location. The answer is revealed in Fig. 7.

our ability to distinguish them as they appear to blend into each other (even if a high resolution camera is used), as noticed in Fig. 1.

However, it does not mean that the captured image is without contrast because of the existence of ambient light[1] naturally. Ambient light reflects small amount of visual information and as a result creates contrast within dark images and videos. We intend to unveil this contrast to make the blended information more apparent. This is not new as photo editing software (e.g. Adobe Photoshop) had made use of this to enhance poor quality images (i.e. poor lighting, exposure error).

Though widely used and gives good results, they require manual processing of images and videos which are both tedious and impractical for real-time employment. On the other hand, existing approaches that deal with applications at nighttime commonly adopt specified cameras such as infrared and thermal imager [12, 17]. However, the support provided by the cameras also becomes a constraint because the specified hardware must be available for their systems to work, otherwise they become ineffective.

This paper proposes a visual analytic approach to unveil the contrast in darkness. This is made possible by utilizing the traits of image representations obtained from computer vision techniques through a learning based inversion algorithm, eliminating the reliance on camera hardware and at the same time minimizing the need for human intervention. We also propose potential applications that would benefit

---

[1]In photography, ambient light refers to available light in the surroundings that is not artificially provided by the photographer

from our proposed framework, and introduce a new night-time dataset, Malaya Pedestrian in the Dark (MyPD)[2], because to our knowledge, such dataset is not available publicly. It is a challenging dataset collected from the image sharing website, Flickr, consisting of 223 images with and without people, captured at twilight or nighttime.

2. Proposed Framework

Ambient light is always available in our surroundings even in the darkness of the night. To put it into perspective, we are still able to see inside a dark room (without switching on indoor lighting) because of moonlight or streetlights penetrating the room through the windows. Unless the room is fully sealed where absolutely no light can penetrate.

Similarly, images captured in the dark (which we call dark images) that appear to look uniform to us actually contain details due to the ambient light. However, the relatively low intensity that leads to low contrast renders the details not easily noticeable by simple human observation. Unlike in real environments where our eyes can adapt to low light, the intensities captured within an image is fixed. Hence, to be able to see the contents captured in the dark image, we need to adjust the contrast of the image, and we intend to develop an algorithm to do so, making use of the small differences provided by the ambient light's intensities.

The term contrast is a commonly used term in computer vision referring to image gradients or difference in intensity that enables us to see edges and textures. To this end, our strategy is to derive a formula that expands (maximizes) the gradient of dark image. As a result of that, we obtain an image with the information revealed. Formally, given a dark image, I, we define the gradients from ambient light as $\nabla(I)$ and the resultant image, I_R can be defined as follows:

$$I_R = \underset{\nabla(I) \in \mathbb{R}^D}{\arg\max}(\nabla(I)) \tag{1}$$

At the same time, we note that feature descriptors can maximize $\nabla(I)$ of an image as follow:

$$\phi(I) = \max_{\nabla(I) \in \mathbb{R}^D}(\nabla(I)) \tag{2}$$

where $\phi(I)$ is the feature representation. By inverting Eqn. 2, the representations can be formulated as an image that has maximized gradients,

$$\phi^{-1}(I) = \underset{\nabla(I) \in \mathbb{R}^D}{\arg\max}(\nabla(I)) \tag{3}$$

It is apparent that Eqn. 3 is equivalent to Eqn. 1, thus, we can represent the resultant image, I_R by the inverse of gradient based features of I, $\phi^{-1}(I)$.
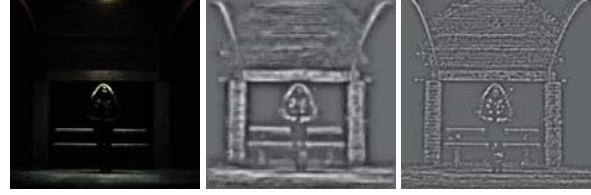
(a) Original image   (b) $\phi(I)$ = HOG   (c) $\phi(I)$ = SIFT

Figure 2. Example of different visualizations, I_R

In the literature, there are great deals of notable gradient based features, so it is not necessary to design new ones. Instead, we selected them based on our objective, that is gradient maximization. Our choices are, but not limited to, SIFT [11] and HOG [3] as these representations are computed with a series of gradient magnitude calculation, orientation binning, normalization and thresholding that ultimately bring about optimized gradients representations.

However, inversion of these gradient based features through analytical approach is not viable due to the representations' computations that are not easily reversible. Hence, one of the best strategies is by approximating the image from the nearest feature descriptor, as follows:

$$I_R = \underset{I_R \in \mathbb{R}^D}{\arg\min} \|\phi(I_R) - \phi(I)\|_2^2 \tag{4}$$

The inversion formulation in Eqn. 4 is also employed in [4, 14, 15] that has gained interest in the computer vision community lately. While their purpose is to study features, we approach it to enable us to unveil contrast. We engage the paired dictionary learning method [14] to solve Eqn. 4, and example of the resultant images are shown in Fig. 2. The resultant images from both gradient based features are distinctively different due to their dissimilar calculation algorithms.

It is evident that the images produced by using both SIFT and HOG as the gradient based features (denoted as $I_{R,SIFT}$ and $I_{R,HOG}$ respectively for the rest of the document) are similar contrast and brightness adjusted grayscale images (SIFT and HOG does not encode color information). Both accentuates the edges and objects' outlines in the dark image, albeit at varying degrees of detail, revealing details of objects or people that were unobservable previously in the original image (e.g. the bench behind the person in Fig. 2). As a result, such emphasized information makes detection or recognition of objects such as people, characters etc. in the dark possible.

3. Experiments

In our experiments, we evaluated the proposed method in comparison with other approaches, namely image integral and gamma correction.

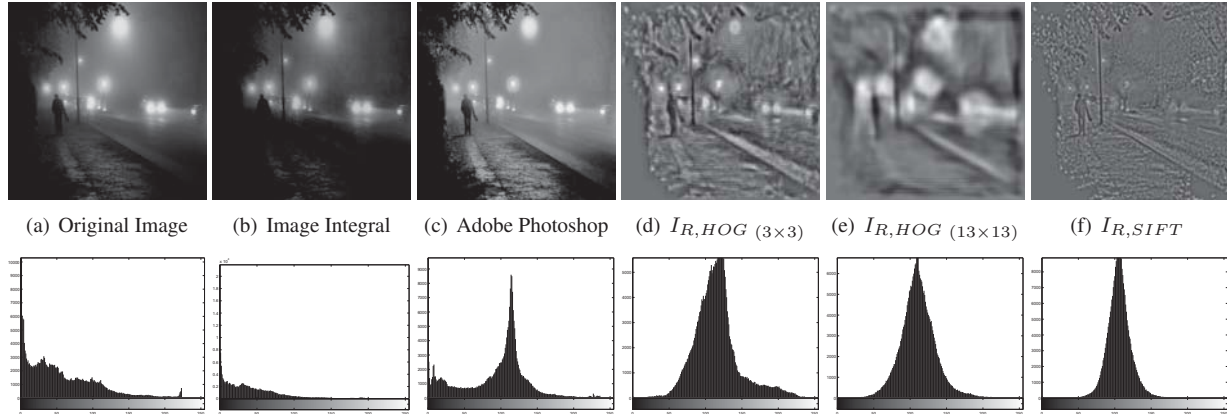|					
(a) Original Image	(b) Image Integral	(c) Adobe Photoshop	(d) $I_{R,HOG\ (3\times3)}$	(e) $I_{R,HOG\ (13\times13)}$	(f) $I_{R,SIFT}$

Figure 3. Example of image with different visualization and their corresponding intensity histograms (x-axis: intensity values from 0 to 255; y-axis: frequency)

3.1. Dataset

The experiments and evaluation of our framework was performed using the new MyPD dataset we collected from Flickr, an image sharing website, consisting of 223 dark images. They are a mix of images captured at nighttime or twilight with inconsistent lighting that restricts our observation of details. There are 81 images without people and 142 images that contain varying numbers of people from one person to groups of people at different sizes and backgrounds including crowds. We additionally annotated the people in the images so that the dataset could be used for other tasks such as people detection. The annotation was done using bounding box that covers the entire person, mostly in upright position, amounting to a total of 373 persons.

3.2. Experimental Settings

Among the four methods we applied to unveil contrast, the image integral is the most basic approach where all the dataset images were subtracted with the mean across the whole dataset. On the other hand, we had chosen gamma correction approach (i.e. using Adobe Photoshop software) as our baseline where it was done by manually[3] adjusting specific parameters (i.e. contrast, brightness, saturation, sharpness) of each image until the contents were clearly seen. As for our approach, in SIFT we used regular grid to densely sample the features, also referred to as Dense SIFT (DSIFT) [13], to enable the rebuilding of the full image structure of $I_{R,SIFT}$.

We train paired dictionaries of SIFT and HOG at patch size of 13×13 from Caltech-256 [10] as it consists of plain object-only images that gives simple and distinct patch basis for better visualization quality. It could be seen from our experiment that a 13×13 $I_{R,HOG}$ (denoted as $I_{R,HOG}$ (13×13)) did not show much comprehensible in-

---

³We employed 5 volunteers for this task, and selected the best qualitative result

Gamma Correction	Integral	$I_{R,HOG}$ (3×3)	$I_{R,HOG}$ (13×13)	$I_{R,SIFT}$ (13×13)
14.38	24.40	8.94	7.55	9.92

Table 1. Mean PSNR (in dB) comparison between approaches.

formation, therefore, we additionally train another dictionary for $I_{R,HOG}$ using patch sizes of 3×3 ($I_{R,HOG}$ (3×3)).

3.3. Performance Comparisons

We evaluated the four different approaches qualitatively and quantitatively using peak signal-to-noise ratio (PSNR). PSNR is commonly used to measure the quality of reconstruction from image compression. Here, it is used to measure and compare the resultant quality of the adjustments made by the different methods. Table 1 shows the mean PSNR of the set of images we tested on and as expected, image integral has the highest PSNR among all approaches. This is because the process only involves subtracting the dataset mean. However, through visual observation in Fig. 3, it is obvious that image integral has difficulty revealing the details of the images. In fact, almost nothing can be noticed from the image integral. Conversely, gamma correction performs much better visually despite a low PSNR. In the mean time, the PSNR of $I_{R,SIFT}$ is better in comparison with $I_{R,HOG}$ in both patch sizes, and has a closer PSNR value to the gamma correction. This suggests that the contrast adjustment by our approach with SIFT as the gradient based features results is a closer match to the baseline.

A comparison of intensity histograms in Fig. 3 shows a similar trend between the gamma corrected and our proposed results while, the image integral skews the intensities towards lower values. This similarity is in agreement with the PSNR results that show our approach is similar to the baseline. This is further supported by visual observation in the respective images whereby details in Fig. 3 can be seen, particularly in $I_{R,HOG}$ (3×3) and $I_{R,SIFT}$, similar
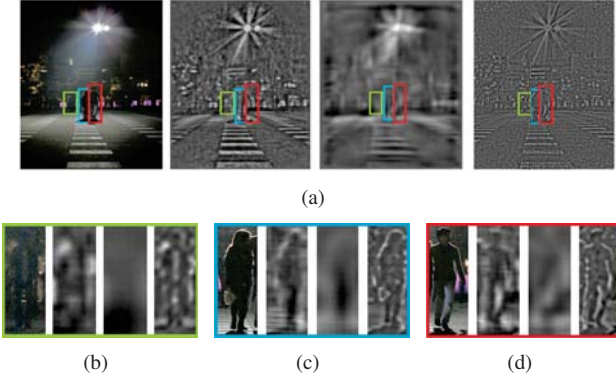
(a)

(b)　　　　(c)　　　　(d)

Figure 4. Image with people of varying scales: (a) From left to right: Baseline (Gamma), $I_{R,HOG}$ (3×3), $I_{R,HOG}$ (13×13), $I_{R,SIFT}$ (13×13); Crop images of people from the colored bounding boxes in (a): (b) Green, (c) Blue, (d) Red

Figure 5. Example of the detections of people with varying scales, number of people and background complexity with $I_{R,SIFT}$. All these images are in their original size.

to the gamma corrected image. Furthermore, we establish that $I_{R,SIFT}$ retains more details than $I_{R,HOG}$. This is supported with an example in Fig. 3 where it is easier to see fine details in $I_{R,SIFT}$ such as the well defined figure of the person, street lights, and leaves of the tree, as opposed to the slightly distorted images of $I_{R,HOG}$ (3×3) and $I_{R,HOG}$ (13×13). We can associate this finding to the differences of the chosen SIFT and HOG as the gradient based features.

On further observations as depicted in Fig. 4, we deduced that using $I_{R,SIFT}$ as the gradient based features is more robust to scale changes than $I_{R,HOG}$. For example, Fig. 4(a) illustrates $I_{R,SIFT}$ and $I_{R,HOG}$ images containing 3 people with different scales, in comparison with the gamma corrected approach as the baseline. The person in the green bounding box, enlarged in Fig. 4(b), is at a smaller scale compared to the other two in blue and red boxes, enlarged in Fig. 4(c) and Fig. 4(d), respectively. As seen, it is hard to visualize any people in $I_{R,HOG}$ (13×13). Downsizing the patch (i.e 3×3) to produce another $I_{R,HOG}$, two of the people are now visible (as shown in Fig. 4(c) and Fig. 4(d)), but the other (smaller) person (i.e in Fig. 4(b)) is still missing. In the case of $I_{R,SIFT}$, it clearly visualizes all the people in one single size (13×13). Based on this analysis, it seems to us that the robustness of $I_{R,SIFT}$ provides a more suitable visualization platform for further tasks, and this is one of our future work.

4. Applications

In this section, we explore the use of our proposal in practical applications. The framework can be applied to a wide variety of uses, this work focuses on fields that would assist public safety. Specifically, people detection and forensic study.
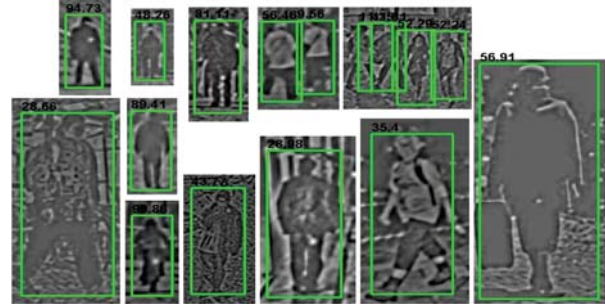
4.1. People Detection

Despite successful results in people detection [2, 6, 9], there are almost no works that addressed nighttime detection using color (RGB) images. Research in this direction is valuable as it can be the groundwork for complex tasks, such as suspicious activity detection at nighttime that would lead to identifying criminal activity. The combination of our framework with people detection algorithm can be the basis for such future works. Generally, the method can be coupled with any detector to perform person detection in the dark. In this work, we choose the ACF detector [5] as it is one of the fastest people detectors with state-of-the-art performance.

Using our proposed dataset, the images are separated into training and testing sets, where the training set has 100 positive samples of people images and 60 negative samples, while the testing set has 40 positive and 21 negative samples. We tested the people detection on $I_{R,SIFT}$ as established in Section 3.3, that $I_{R,SIFT}$ is more robust compared to $I_{R,HOG}$. Fig. 5 shows the samples of successful people detections with $I_{R,SIFT}$. It is noted that the detected people are of varying sizes, yet their details can be clearly seen and understood even by human vision.

4.2. Forensic Study

Our proposed method that "reveals" details in the dark is also a useful tool for forensics study. In investigations, image evidence is crucial and every detail within the image is vital. However, photo evidence is not always presented in the best conditions with good lighting and contrast. Especially if the evidence is a crime that happened in the dark or at night, and night vision cameras are not readily available to capture the moment. As a result, investigators would have to put in more time to process the images. Our proposed framework would assist them in processing dark images and possibly bring forward (unveil) useful details that were "hidden".

We applied our framework on all 221 images of the MyPD dataset and studied the contents. To our surprise, we
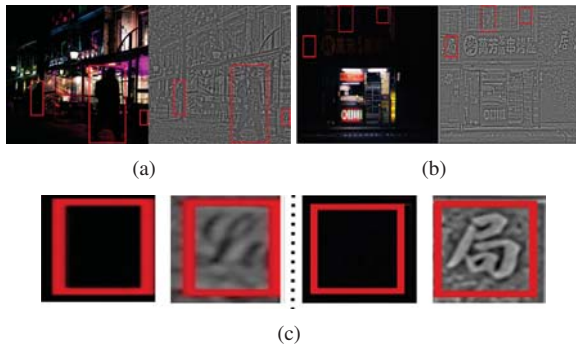
Figure 6. Example of visualization in the dark. (a) $I_{R,SIFT}$ shows a person is behind the pole at the back, the type of clothing the person at the front is wearing. (b) $I_{R,SIFT}$ shows a person at the window above the signboard, and a box like object in another window. (c) a zoom up for the image before and after our proposed method (most right box of (a) and most left box on (b)).
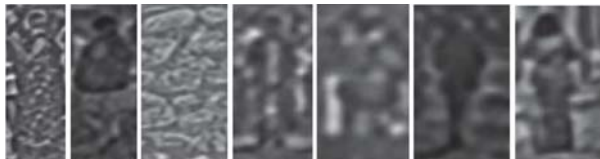

Figure 7. The correspond to the original image in Fig. 1. All except two (3rd and 5th image) contains a person.

notice that the visualizations bring details to our attention that were otherwise missed, such as the example shown in Fig. 6. In Fig. 6(a) we found a person behind the pole at the back, the type of clothing the person at the front is wearing, and writings on a sign in the darkness, while in Fig. 6(b), we noticed what seems to be a person at a window above the signboard and a box-like object in another window. Also, a Chinese character on the left of the image is clearly visible.

5. Conclusion

In this paper, we propose a novel visual analytic approach that utilizes naturally available ambient light to unveil the contrast of dark images. The proposed framework makes use of characteristics of feature descriptors to capture ambient light and employs learning based feature visualization to invert the features into visible image space. The result is an image with emphasized contrast, similar to gamma correction. It is found that SIFT is a better gradient based feature for this task in comparison with the HOG, because the resultant image contains higher degree of detail. Furthermore, we presented applications that can benefit from this framework, namely people detection at nighttime and forensic study, whereby it can serve as a basis for future nighttime applications.

References

[1] M. Allen, L. Weintraub, B. Abrams, and M. Green. *Forensic Vision with Application to Highway Safety*. Lawyers & Judges Publishing, 2011.

[2] R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection, what have we learned? In *ECCV Workshops*, pages 613–627. 2014.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

[4] E. d'Angelo, A. Alahi, and P. Vandergheynst. Beyond bits: Reconstructing images from local binary descriptors. In *ICPR*, pages 935–938, 2012.

[5] P. Dollar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *T-PAMI*, 36(8):1532–1545, 2014.

[6] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *T-PAMI*, 34(4):743–761, 2012.

[7] D. Farrington and B. Welsh. *Effects of improved street lighting on crime: a systematic review*. Home Office London, England, 2002.

[8] S. Fotios, J. Unwind, and S. Farrall. Road lighting and pedestrian reassurance after dark: A review. *Lighting Research and Technology*, 47(4):449–469, 2015.

[9] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf. Survey of pedestrian detection for advanced driver assistance systems. *T-PAMI*, 32(7):1239–1258, 2010.

[10] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[12] B. Qi, V. John, Z. Liu, and S. Mita. Use of sparse representation for pedestrian detection in thermal images. In *CVPR-W*, pages 274–280, 2014.

[13] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *ACM Multimedia*, pages 1469–1472, 2010.

[14] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba. Hog-gles: Visualizing object detection features. In *ICCV*, 2013.

[15] P. Weinzaepfel, H. Jégou, and P. Pérez. Reconstructing an image from its local descriptors. In *CVPR*, pages 337–344, 2011.

[16] B. Welsh and D. Farrington. Improved street lighting. In *Preventing Crime*, pages 209–224. Springer New York, 2007.

[17] X. Zhao, Z. He, S. Zhang, and D. Liang. Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification. *Pattern Recognition*, 48(6):1947 – 1960, 2015.