https://doi.org/10.2352/ISSN.2470-1173.2020.4.MWSF-216

Camera Unavoidable Scene Watermarks: A Method for Forcibly Conveying Information onto Photographs

Clark Demaree and Henry Dietz; Department of Electrical and Computer Engineering, University of Kentucky; Lexington, Kentucky

Abstract

When a scene's image rights need to be protected e.g. a stage performance, it is valuable to use human imperceptible methods to forcibly add markers to a camera's perception of the scene regardless of the camera's precise location, focus distance, or shutter speed. This work expands upon extant methods for adding human imperceptible, camera perceptible markers to scenes but does so with the assumption that the photographer will take natural steps to avoid capturing the markers. The proposed method utilizes a combination of a traditional method of adding an image to the scene, and projections from the scene onto the camera's entrance pupil. This method is intended to function even when the target camera utilizes an IR filter and has a shutter speed 1/60 s. More interestingly, the combination of traditional images with projecting onto the camera allows this method to not be reliant upon knowledge of the camera's focus settings, or the precise location of the camera. It is, however, marred by numerous other requirements which make the method unreliable.

Introduction

Traditionally, watermarking has focused on images and videos rather than scenes. There is, however, a similar desire to protect rights to physical scenes such as architecture, stage performances, etc. The methods by which these rights are violated by bad actors, however, obviously differ from the typical copy-paste violations of digital image rights. The protections, therefore, must account for this difference; "watermarks" must be added to images as they are being taken by a bad actor's camera.

Most features about bad actors violating rights to these scenes are evident; It can be assumed that a bad actor who intends to photograph/film a scene and distribute the resulting images/video will adjust the camera's aperture, shutter speed, and sensitivity. It is also assumed that none of these factors are precisely known but fall within known ranges.

There are two key differences between camera and human perception that motivate this work - flicker perception particularly for red and near infrared light, and out-of-focus point spread functions.

Obviously, when a camera's shutter speed is much shorter than the period of a flickering image, the camera is likely to only capture one state of the image or to capture significantly more of one state than the other. For humans, the frequency at which flickering light appears to be steady, the critical flicker fusion threshold (CFF) is lowest for deep red and near infrared light. A method for determining critical flicker fusion threshold under the conditions used in this work was not identified [1]. The critical flicker fusion threshold is however, less than 60 Hz for $\lambda = 670$ when the luminance is 1 cd/m2 and the image covers 19° visual angle and under the ideal circumstances of healthy adults viewing in a darkened room. When the area and luminance are both decreased by a factor of 10, however, that value can fall below 15 Hz [1] [2].

If an image is flickered with its inverse, it will appear to humans to be the average of the two and carry no information [2]. Flickering forms have been used to protect against unwanted photography and videography, but this has been limited to 2D media [3].

More critically, the out of focus point spread function of a camera takes the approximate form of the virtual image at the entrance pupil of the camera [4]. A similar effect was used by MIT Media Lab for their "bokode" designs which used a projector creating an image on the sensor of a camera focused to infinity [5]. This is not, however, useful if the photographer is trying to avoid capturing the projected image in his or her photograph. This work focuses on forcing these kinds of figures into photographs when the photographer is attempting to prevent the figures from appearing.

We make the following contributions:

- We propose a mechanism for using flickering images to introduce figures to a scene that are predictably visible to cameras but have minimal effect on human perception of the scene.
- We introduce a second mechanism, EPP (Entrance Pupil Projection) which introduces figures to a scene that are visible to cameras even when the camera focus settings are not precisely known.
- We evaluate the expected performances of both mechanisms.

Projecting Flickering Images

The obvious way to covey IP information is to put up a sign or poster or to project that information onto a flat surface. This will, however, disrupt the viewer's experience of the scene. We can replace the static image with a flickering one to reduce its visibility to humans. This certainly is not the first use of flickering figures to interfere with camera capture but is unique in the desire to, rather than simply disrupt an image, force information onto it. The minimum rate to achieve this effect is the critical flicker fusion threshold. When an image is flickered fast enough, it appears to a human to be in steady state. When we alternate an image with its inverse as shown in figure 1, it appears to be blank.

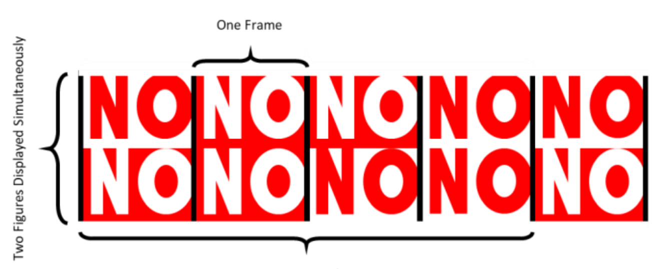

Figure 1. projected image changing over time as figures are inverted with different phases

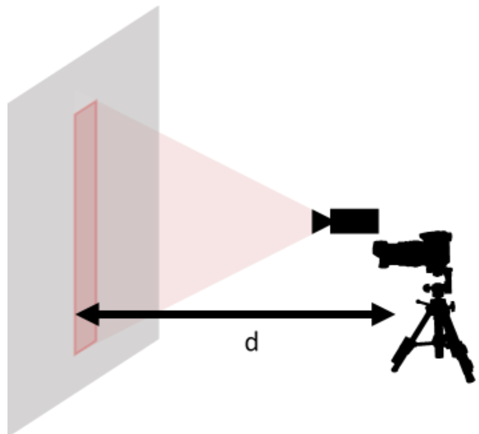

Figure 2. flickering image projected onto surface with camera focused on the surface

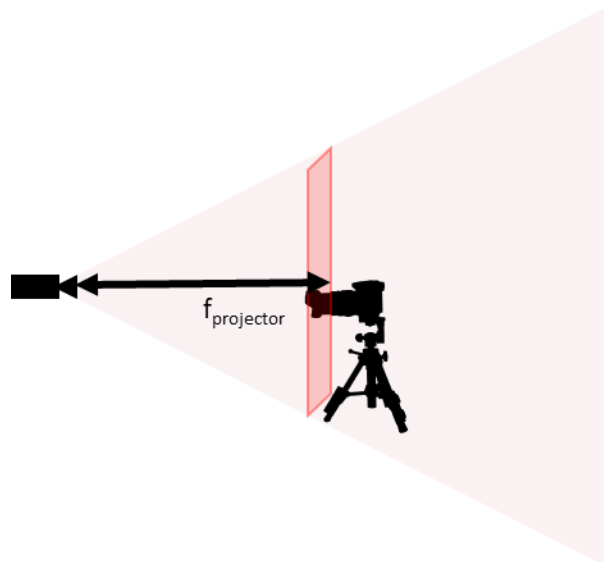

Figure 3. a projector projects image on cameras entrance pupil and the camera is not focused on the projector

The shutter speed must be short enough that it only captures a fraction of the flicker period. Otherwise, it may capture both flicker states equally and not carry any information If only one flickering figure is used, there is always a chance that the shutter captures a transition between states. To prevent this, we can use multiple figures that are out of phase. When we have two figures as shown in figure 1, at least one figure is in steady state for any length of time less than a quarter of the flicker period. When many figures period. This means that with many figures and a flicker frequency significantly less than 60 Hz, we can be effective with shutter speeds as slow as 1/30 of a second.

In practice, once a frequency is selected, the figures are arranged into an image to be projected and projected. The projector obviously must be able to alternate between the image with its inverse as seen in figure 1 where each figure is inverted between being white and red and being red and white. In practice, however it is more appropriate to use black and red so that the total light being used is limited or possibly black and white so that the projection does not affect the color of the scene.

The basic implementation of this is shown in figure 2. The projector simply creates the image of flickering figures on a surface and the camera is focused on that surface.

Entrance Pupil Projection

The flickering image only works if the offending photographer's camera has it in focus. Because the system cannot depend upon the photographer focusing on any one point, a second mechanism is needed. Like bokodes, EPP allows the camera's position to change but ensures that a figure will be visible to the camera but not to a viewer. Both of these involve projecting a virtual image onto a camera lens so that it appears in the place of the typical out of focus disk.

This method differs by focusing from the source side rather than the camera side. As a result, it is not necessary that the camera to have a specific focus distance. It is necessary that the camera does not have the projector in focus and that the camera's entrance pupil to be close to the focus plane of the projector. In implementation, a projector with an aperture of a few mm projects a repeating pattern of small figures at the region of a scene where a camera is expected to be. We see this in figure 3. In this case, it is assumed that the projector is focused to the camera and the camera is not focused to the projector. The light enters the entrance pupil of a camera at that position and appears as a figure on the sensor as seen in figure 2 where figures of the letters UK are projected. It is important that the projected images have a size less than half the entrance pupil size so that it is ensured that at least one figure is completely captured. We see this in figure 4 where multiple UK figures are partially visible.

If the exact region where a camera will be in unknown, multiple projectors can be used to cover a wider range. In this work, it is assumed that the figures are static but in practice, they could be dynamically changed if the projector used allowed dynamic control of them.

Figure 4. Figure of letters "UK": visible in out of focus disk shot with Sony a7

Combining Flickering image and EPP

When multiple projectors are used and focused to different distances, when the camera's position moves so that the camera no longer captures one figure sharply, the camera captures a different figure sharply as shown in figure 5. When this is used, a

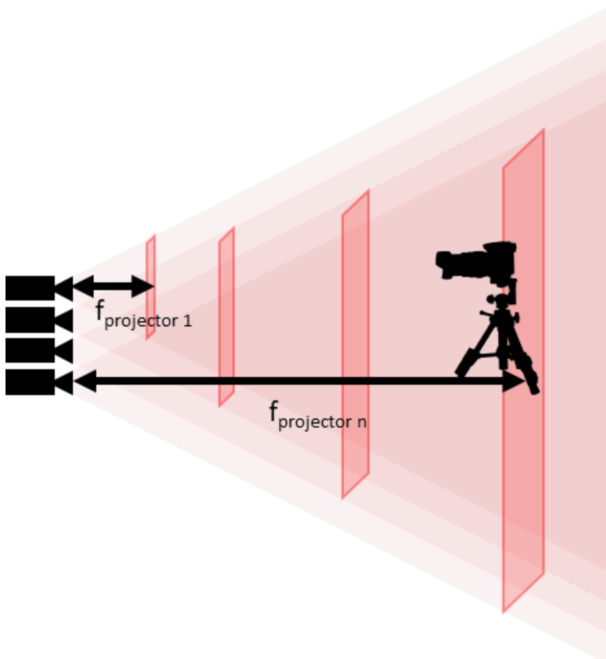

Figure 5. many projectors covering a range of possible distances to the camera by each being focused to different distances

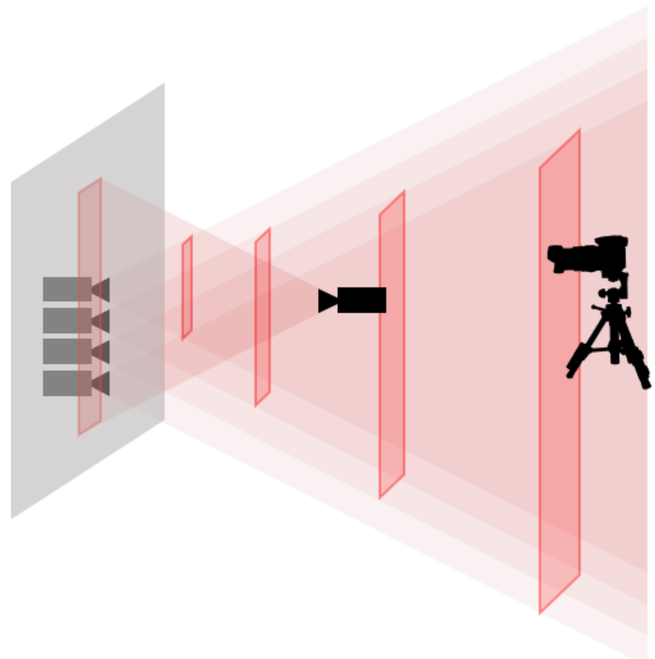

Figure 6. a range of distances and focus settings accounted for by a combination of methods

wide range of camera distances but the system requires that the camera is not focused to the projector.

As was stated previously, the flickering image fails when the camera does not have it in focus and EPP fails when the camera has it in focus. As a result, if EPP projectors are placed in the same plane as a flickering image, when the camera focus causes the flickering image to be too blurred to be read, the EPP figure is visible. This is seen in figure 6 where a range of distance to the camera and focus settings are accounted for by the system.

Failure Cases

There are several cases that can cause the system presented in this work to underperform or to fail entirely. It is clear that from the nature of protection systems, having such simple loopholes available to photographers makes the system unreasonable in its current form. One challenge that affects both mechanisms is exposure level. Particularly in the case of flash photography, exposure level can be challenging to predict which can lead to over or underexposure of the figures.

There are several obvious risks. The dependence upon figures being in focus means that if a synthetic blurring is used, the watermarks may be obscured automatically. When the camera is pointed in a direction where the figures are not in frame, they will simply not be visible. If the images used are in the deep red, a color filter can reduce the camera perception of the figures.

Caveats for Flickering Images

The greatest concern with flickering images is that there is great variation between individuals for flicker perception and so a system which is imperceptible to most viewers may be perceptible to some. Additionally, the use of CFF as the limit of human perception is only appropriate when the viewer does not have eye motion or blinking. This does not mean that flickers at frequencies may be intermittently perceptible. Those flickers, however, may not be as disruptive as flickers with frequencies below the CFF and may be appropriate for some applications.

Additionally, because of the high frequency of the image flicker, temporal shutter artifacts can occur. The exact artifacts are dependent upon the shutter method [6] but the one of greatest concerns is shutters which capture different parts of the scene at significantly different times which could cause all of the figures to be captured in a transition state. It should however be noted that this is quite unlikely because spaciotemporal shutter artifacts decrease in significance as the shutter speed becomes slower [6].

Failure Cases for Entrance Pupil Projection

By its nature, EPP depends upon only one entrance pupil being used by a camera and that it be an appropriate shape and size to capture a figure. As a result, it is ineffective when multi-lens and multi-camera systems are used. Additionally, it fails when a mirror lens is used because mirror lenses have ring shaped entrance pupils [5]. Another issue is when the user is able to eliminate or alter bokeh in certain areas such as when light field cameras are used.

Evaluation of Flickering Images

A key measure of the camera perception of the figure from a flickering image is the contrast between the two color values used to form the figure. In figure 1, the values white and red are used. This can be treated as a simple digital signals problem of capturing a pulsing signal with a duty cycle of 50% When n figures with equally spaced phases are used in a flickering image, the maximum length of time in which it is assured that at least one figure is in steady state is

$$t_{shutter\,\max contrast} = T_F \frac{n-1}{2n} \quad (1)$$

Where T_F is the flicker period and $t_{shutter\ max\ contrast}$ is the maximum shutter speed without a risk of contrast loss. We see the relationship between the number of figures used and the maximum shutter speed without a risk of contrast loss in figure 7. When n is high, the maximum shutter speed approached half the flicker period.

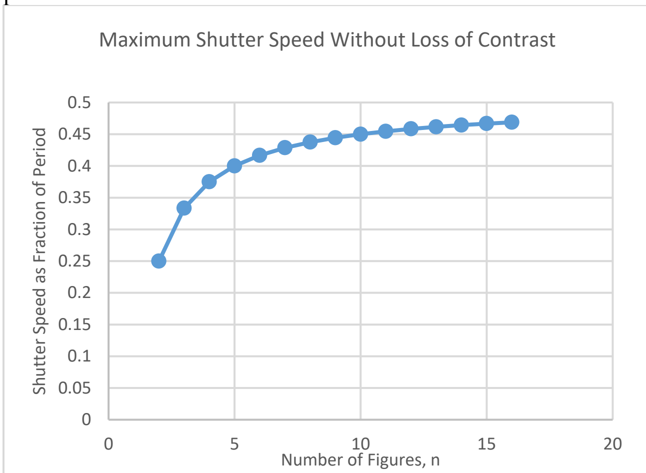

Figure 7. Relationship between n and maximum shutter speed without loss of contrast

When the shutter speed exceeds this value but is less than T_F, the contrast decreases. Trivially, for any $t_{shutter} > t_{shutter\ max}$, at least one figure state will be captured for a time, t_{figure}, of at least $t_{shutter\ max}$.

To determine the worst-case scenario for a decrease in contrast, we examine the capture temporally. The worst-case scenario is that the capture is perfectly centered upon a length of $t_{shutter\ max\ contrast}$ that begins or ends at the same time that the figure's state is inverted. When this occurs and $t_{shutter} \leq T_F \frac{n+1}{n+2}$, the additional time beyond $t_{shutter\ max\ contrast}$ is split evenly between the two states. This means that the ratio of time spent capturing one state, t_{figure}, to time spent capturing its inverse, $t_{inverse}$, becomes

$$\frac{t_{figure}}{t_{inverse}} = \frac{1}{2} + \frac{1}{2}\frac{t_{shutter\ max\ contrast}}{t_{shutter}} \tag{2}$$

If $t_{shutter} > T_F \frac{n+1}{2n}$, then even in the worst-case scenario, the camera captures on complete state of a figure. As a result, unless $t_{shutter} > T_F$, in which case parts of multiple periods are captured, any time that is not spent capturing that figure is spend capturing the inverse. Naturally,

$$t_{figure} = \frac{T_F}{2} \text{ and } t_{inverse} = t_{shutter} - \frac{T_F}{2} \tag{3 and 4}$$

The contrast loss is simply defined by

$$\frac{figure\ contrast}{single\ state\ contrast} = \frac{t_{figure} - t_{inverse}}{t_{figure} + t_{inverse}} \tag{5}$$

Combining these equations, we get

$$\frac{figure\ contrast}{single\ state\ contrast} = \begin{cases} 1 \text{ for } t_{shutter} \leq T_F \frac{n-1}{2n} \\ \frac{T_F \frac{n-1}{2n}}{t_{shutter}} \text{ for } T_F \frac{n-1}{2n} < t_{shutter} \leq T_F \frac{n+1}{2n} \\ \frac{T_F}{t_{shutter}} - 1 \text{ for } T_F \frac{n+1}{n+2} < t_{shutter} \end{cases} \tag{6}$$

This is depicted in figure 8. We see that with increasing n, the contrast improves for some but not all shutter speed to flicker frequency ratios. Additionally, we see that the improvements are dramatic as we go from $n = 2$ to $n = 3$ but the improvements from increasing the number of figures is diminished when the number of figures is already high.

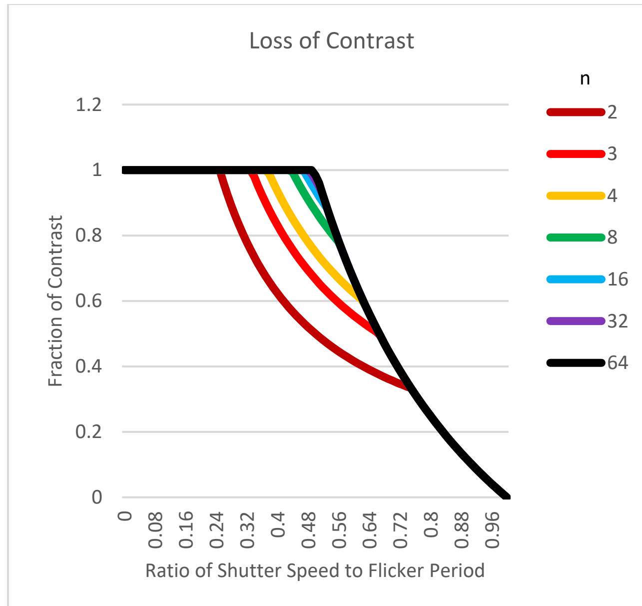

Figure 8. Loss of contrast of flickering image for various values of n

Evaluation of Entrance Pupil Projection

The entrance pupil projection performance is limited primarily by the depth of focus of the figure and by the exposure level of the figure. The average luminance is described by the amount of light from the projector that passes through the entrance pupil and reaches the sensor divided by the size of the out of focus point spread figure on the sensor. This is a simple description but because numerous factors affect the size of the out of focus point spread figure, predicting the exact exposure level without knowing the size of the out of focus point spread figure is challenging. Different cameras can have significantly different out of focus point spread figure behaviors but generally the diameter of the figure is proportional to the aperture diameter and is inversely but not linearly related to how close to the focus plane the point source is. Because the exposure level of the figure is dependent upon its size. if the figure is too small, it may become overexposed. The opposite, where a large figure is too dim to be seen, is also possible but because there is a strict limit to how large these can get, it's not as great of a concern.

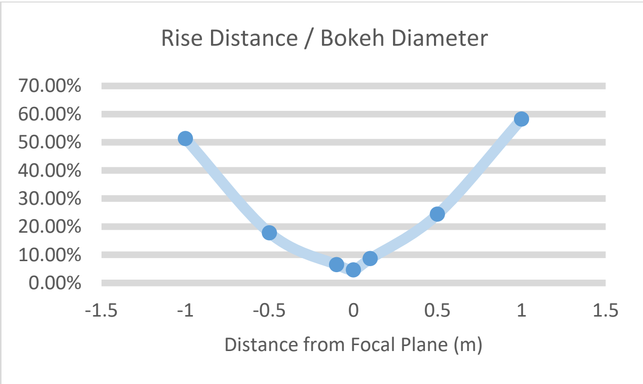

Figure 9. Loss of sharpness with entrance pupil distance from focus plane – Note that negative values are closer to projector and positive values are further from projector

The other limitation, depth of focus proved to be greater challenge in experimentation. To test the range of focus distances, a projector was focused to 10 m in front of it with a pattern of wide vertical lines. A Sony a7 camera was then placed so that the projector was in the approximate center of its field of view. The camera was focused to infinity to increase the out of focus point spread function to maximum size and photos were shot at the focus plane, and multiple distances in front of and behind the focus plane. It was found as expected that as the camera moves away from the focus plane, the figures blur.

The photos were then inspected to calculate the sharpness of the figures. Because the figures are small, finding sharpness using a spatial frequency style test was not an option. It was for this reason the figures projected were thick vertical bars. The rise distances of the bars were then determined, and the results are shown in figure 9. The rise distances are given in terms of their relationship to the out of focus point spread figure diameter because that is what determines the minimum feature size of a figure.

By its nature, EPP depends upon only one entrance pupil being used by a camera and that it be an appropriate shape and size to capture a figure. As a result, it is ineffective when multi-lens and multi-camera systems are used. A similar problem prevents it from being affected when light field cameras are used. Additionally, it is expected to fail when a mirror lens is used because mirror lenses have ring shaped entrance pupils.

Conclusions

The most essential conclusion of this work is that the mechanisms presented are not, in their current states, reasonable solutions for bringing watermark protections into large scenes. Despite the admitted lack of robustness, the successes of these mechanisms suggest that they are feasible when the specific camera requirements and resource requirements for successful performance are met. EPP likely has applications beyond those discussed in this work. Additionally, this work demonstrates the importance of accurate CFF predictions which cannot be made from the extant research.

References

[1] A. Eisen-Enosh, N. Farah, Z. Burgansky-Eliash, U. Polat, and Y. Mandel, “Evaluation of Critical Flicker-Fusion Frequency Measurement Methods for the Investigation of Visual Temporal Resolution,” Scientific Reports, vol. 7, no. 1, 2017.T. Jones, "Sample Journal Article," Jour. Imaging Sci. and Technol., vol. 53, no. 1, pp. 1-5, 2009.

[2] D Hecht D Shlaer “Intermittent stimulation by light. V. The relation between intensity and critical frequency for different parts of the spectrum”. 19, 965–979. https://doi.org/10.1085/jgp.19.6.965 Journal of General Physiology, 1936

[3] I. Suzuki, S. Ando and Y. Ochiai, "70-1: “Unphotogenic Light”: Evaluation and Detail of the High-Speed Projection Method to Prevent Secret Photography with Small Cameras", SID Symposium Digest of Technical Papers, vol. 49, no. 1, pp. 930-933, 2018. Available: 10.1002/sdtp.12228.

[4] H. G. Dietz, “Out-of-focus point spread functions,” Digital Photography X, Jul. 2014.

[5] A. Mohan, G. Woo, S. Hiura, Q. Smithwick and R. Raskar, "Bokode", ACM Transactions on Graphics, vol. 28, no. 3, p. 1, 2009. Available: 10.1145/1531326.1531404.

[6] H. Dietz and P. Eberhart, “Shuttering methods and the artifacts they produce,” Electronic Imaging, vol. 2019, no. 4, 2019.

[7] A. Mohan, G. Woo, S. Hiura, Q. Smithwick and R. Raskar, "Bokode", ACM Transactions on Graphics, vol. 28, no. 3, p. 1, 2009. Available: 10.1145/1531326.1531404.