Non-Uniform Integration of TDCI Captures

Paul Eberhart and Henry G. Dietz; University of Kentucky; Lexington, Kentucky

Abstract

TDCI (Time Domain Continuous Imaging) is a system for image capture and representation in which scene appearance is modeled as a set of continuous waveforms recording the changes in incident light at each pixel over time. Several of the advantages of TDCI are related to the ability to set exposure parameters after-the-fact, rather than at the time of capture. These exposure parameters can be far more complicated than are physically realizable in a conventional camera, or reasonable to design without the ability to repeatedly expose the same scene. Previous TDCI experiments have performed relatively traditional integration; this work explores a pair of related exposure behavior enabled by TDCI - the non-uniform integration of incident light into an image along the axes of both the time and space.

This paper details a proof-of-concept implementation which ingests video frames and re-exposes images from the resulting sampled light with user-specified spatially and temporally non-uniform gain.

Introduction

In conventional photography, two major parameters govern the generation of images from the incident light passing through the lens; the sensitivity of the photosensitive medium (the film speed) and the interval of time the photosensitive medium is exposed to incoming light (the shutter speed). These two parameters are traditionally pre-determined before exposure, and fixed for the entire frame and the entire interval of exposure, as they are set by the chemical properties of the film and physical motion of a shutter, respectively. This imposes several limitations on scenes which can be photographed without significant information loss, in particular restricting the dynamic range of scene brightness and speed of moving features in the scene. Digital cameras are, by convention, restricted to the same uniform, pre-set sensitivity and interval behavior as film cameras, though this convention is not necessarily mandated by their physical properties.

Recently, some researchers have begun to reconsider the static-frame assumptions carried over from film in digital cameras. Non-frame-based capture methods, such as Time Domain Continuous Imaging (TDCI) [1] or event cameras [2] instead record changes to the light arriving at the sensor over time, leaving the integration into images to a later, separate, computational step. This work builds on existing TDCI tooling; while no cameras that directly emit a TDCI stream have been constructed, TIK [3] provides a tool to create a TDCI model of the incident light from scene by ingesting frame-based input, such as a video. These virtual TDCI captures can then be re-exposed with arbitrary parameters. Decoupling the capture and exposure (integration) steps of photography allows arbitrary functions for the admission of light over time and sensitivity to light over space to be imposed on the process, repeatedly if desired. This new ability simplifies and improves the execution of a number of existing photographic techniques, as well as enabling a variety of new techniques which may prove useful for both artistic and scientific imaging.

Exposure Interval

In a conventional frame-based camera, the photosensitive element - film or digital sensor - is exposed to light for a fixed interval by the opening and closing of a shutter. In the archetypal camera this is accomplished with the use of a focal-plane shutter; a pair of light blocking curtains which slide over the sensor one after the other, with the intervening time comprising the shutter speed.

With a focal plane shutter, and a relatively long exposure, a first curtain opens, allowing light through the lens to reach the sensor, and some time later the second curtain closes, cutting off the exposure. This scenario provides a very good approximation of the whole sensor being exposed for the same time interval, but that is not always desirable. In scenes with very large variations in brightness, exceeding the dynamic range of the sensor, a uniform exposure time will over or under expose parts of the scene. That is, if the bright areas of a scene are properly exposed to capture a maximum amount of detail, the darker areas may be underexposed and simply appear dark, losing information in that part of the scene. Conversely, if the darker areas of a scene are properly exposed, the brighter areas may be over-exposed and saturate, losing information about that part of the scene.

For very short exposure intervals, a focal plane shutter allows light to strike the sensor by by moving both curtains at the same time, releasing them with an offset less than their individual travel time, thus traversing a slit between the first and second curtain across the sensor. This scenario is a poorer approximation of the entire sensor being exposed for the same interval - while the amount of time each sensel is exposed for approximately the same amount of time, the area along the leading edge of the frame is exposed at an earlier absolute time than the area at the trailing edge of the frame. This offset can result in smearing or other artifacts, particularly if objects are moving in the scene fast enough to create substantial displacements during the shutter interval. Other shuttering methods, such as irising leaf shutters or various electronic dump-and-readout schemes are also employed and create their own distinct artifacts in the resulting image, several of which are detailed in [4].

Some photographers may alter the exposure interval in intentional ways to create specific desired effects. The best-known of these alterations is to shoot multiple exposures on the same photosensitive frame, effectively compositing the incident light during multiple images into a single photo. These techniques, however, are physically complicated to set up, difficult to predict the results of, and require that the setup be executed perfectly at the time of image capture(s). The difficulty of physically realizing complicated exposures, and simplicity of achieving similar effects by computationally compositing images after the fact mean these effects are already most often accomplished by post-processing.

Non frame-based capture schemes like TDCI allow photographers to avoid these problems with exposure by computationally integrating the incident light over one or more interval(s) to create the final image. This means different portions of the scene are not competing for exposure parameters, as they are being sampled independently. Likewise, even the possibility of shuttering artifacts is eliminated, since if the sensels are continuously independently sampled, there are is no correlated scan pattern which could produce artifacts. Most importantly, computational integration separates the processes of sampling the scene and exposing the image, so once sampled, the same interval of incident light can be exposed over and over to produce images, allowing the photographer to tweak the exposure parameters repeatedly, viewing the resulting image and adjusting until the desired effect is achieved.

Film Speed

In a conventional fame-based camera, the sensitivity to incident light (gain) is set as a whole-frame parameter, referred to as "Film Speed" for historical reasons. In an actual film camera, this gain is set by the photo-chemical sensitivity of the film being used, measured in modern times with the ISO 5800 system. In a digital cameras, the gain is determined by the "ISO Setting" in the camera. This setting's properties are specified by analogy to the behavior of film in ISO 12232. In either case, this setting is fixed for the entire frame, for the entire interval of exposure. This whole-scene gain setting is often undesirable as it limits the dynamic range which can be represented in a single capture. Much as for the exposure interval, setting the gain to suit one part of a scene will often leave other parts dramatically under- or over-exposed, losing information about those areas as they saturate or fail to fill in details.

When performing integration computationally after the fact, there is no reason the gain must be uniform for the entire scene. In this scenario, a photographer can specify different gains, or gain functions as above, for different portions of the scene such that each portion of the scene is exposed as desired. There is significant precedent for the desirability of such a feature, as a number of "tricks" allow modern digital cameras to evade whole-scene exposure settings, albeit with significant caveats. First, modern digital cameras with computer-controlled optical paths often have built-in support to take rapid bursts of exposures while automatically varying exposure settings. This method is typically referred to as "HDR Bracketing", as it allows capture of a High Dynamic Range image by "bracketing" - taking exposures with either the aperture or shutter speed varied in steps around an estimated center value. This allows photographers to extend the dynamic range of a processed image by taking the series of frames exposed differently, and ideally correctly for different regions of the scene, and composing them in photo editing software after the fact. This method suffers from several serious limitations. The most obvious limitation is that because the exposures are taken successively, any motion in the scene or camera will cause the successive frames to not line up perfectly, creating artifacts in the composite image. More fundamentally, images generated by HDR bracketing still require that every part of the scene is properly exposed in at least one of the series of frames, requiring the photographer (or camera), to properly estimate the required number and range of exposure settings at the time of capture.

A second extant technique for cheating whole-sensor gain with modern digital cameras is that the sensor ISO setting is often applied, all or in part, as a digital multiplier in the post-processing [5] after the sensor has been read out, rather than by changing the behavior of the sensels. This property is, lately, referred to as "ISO Invariance" or "ISO-Less shooting". This after-the-fact gain function means information is only gained or lost based on the ISO setting during the image processing pipeline, not during capture. As a result, if images are captured at an ISO invariant camera's base ISO, all the scene information the sensor is capable of capturing will be captured and retained, albeit typically with less-than-pleasing brightness. The image can then be brightened in post-processing, essentially applying digital gain later when the photographer has the advantages of time, multiple tries, and additional compute power to make superior decisions about the gain factor. The gain can also be spatially non-uniform - selectively brightening or darkening parts of the scene is an extremely common post-processing manipulation. If the camera used was ISO Invariant this practice is effectively equivalent to selectively changing the sensitivity of the capture device. This technique, however, does not extend the range of the captured data beyond what the sensor can represent for a fixed interval, and may actually shrink it if the brightening or darkening range-clips any pixels. Employing this technique also complicates selecting an appropriate shutter speed, since the captured image will be intentionally under-exposed at the time of capture.

By leveraging TDCI encoding, similar effects can be produced while avoiding many of the disadvantages of the methods that rely on traditional photographic modes. Some methods for using TDCI processing to generate images with larger dynamic range and fewer artifacts were previously explored in [6], but the techniques in that work retained the practice of using a single uniform sensitivity and exposure time. Specifically, if the incident light is recorded as a waveform and sampled after the fact, all of the differently-integrated regions can be integrated with time centered at the same instant, avoiding the issue of artifacts due to changes in the scene between sequentially-shot bracketed exposures. Even better, recording the incident light variation rather than a series of exposed images removes the requirement that the parameters for each of the constituent exposures be pre-selected. Since there are no pre-determined exposures under this scheme, there is no danger of clipping regions of the scene due to lack of data or saturation. This allows the selection of the parameters with which regions of the scene are exposed to be done after the fact, as many times as is necessary, until precisely the desired exposure parameters for each part of the scene are found.

Non-Uniform Over Time

When performing integration of incident light computationally after capture, there is no restriction that the virtually admitted light must be uniformly "exposed" over a single interval as with a physical shutter. Integration gain functions can be specified which simulate mechanically implausible shutter behaviors with only a small amount of extra difficulty. For example, the gain function can have multiple distinct peaks, producing an effect analogous to multiple exposures. Even less physically realizable, the gain function can slope, vaguely physically analogous to imposing a time-varying neutral density filter over the lens, or (somewhat less precisely, as slope variations will not affect depth of field) irising the aperture during the exposure. It is both reasonable and desirable to integrate with gain functions that could not be practically realized by a mechanical means.

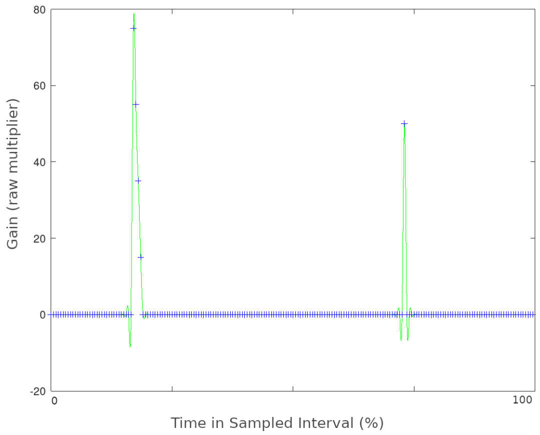

__Figure 1.__ Example Integration Function

__Figure 2.__ Capture of riding mower integrated with the function shown in Fig. 1

In the prototype implementation, and likely going forward through subsequent developments, the integration gain function is represented as a composite spline. Users familiar with image editing will, whether they know it by that name or not, posses at least a passing familiarity with composite Bezier curves, used for drawing arbitrary lines in a wide variety of image editing applications, or (Centripetal) Catmull-Rom splines, also used in image editing for specifying color curves in many photographic editing tools, and modelling camera motion in video processing. Initially, Centripetal Catmull-Rom splines seemed particularly appealing because they are straightforward both to visually manipulate, and to compute the value of at arbitrary position, and already widely used in imaging applications. Furthermore, Centripetal Catmull-Rom are inherently smooth and non-looping (mathematically; twice differentiable), making them immune to ambiguities or discontinuities. Unfortunately, a single Centripetal Catmull-Rom spline can represent only a very limited family of functions, which does not include many trivially-interesting cases.

A more general representation of temporal gain functions is therefore required; current experiments use normal cubic splines [7] as a highly-flexible representation which retains the desirable property of always being twice differentiable. Using this scheme, each the integration gain function is specified as a series of control points. To use the integration gain function, the domain of the specified control points is mapped to the interval of recorded light (allowing for arbitrary granularity), such that the gain to be applied to the incident light at time $t_{current}$ into the sampled interval is the value of the normal cubic interpolation at $t_{current}/t_{max}$.

While the natural behavior for a computer scientist is to specify the temporal integration gain function as an equation or series of control points, perhaps over a unit interval, for most users this will be extremely awkward. An elegant interface would present the user with an initially horizontal line that they can interactively modify by clicking and dragging to add and modify control points. The integration scale factor for each instant is then the height under the curve at that distance into the interval, which is mathematically straightforward, readily visually representable, and leverages existing mental models likely to be available to those accustomed to image editing tools. The horizontal (x) axis of this function represents time, to be stretched over whatever integration interval is selected. The vertical (y) axis of this function represents the instantaneous gain to be applied to the incident light at time $(t/interval) * fn_{max}$.

1 shows an integration gain function representing a double exposure, with the first exposure ramping very quickly to very high gain then somewhat more slowly tapering, and the second very quickly flicking on and off to a smaller maximum gain. 2 shows the result of applying that integration function to a capture of a passing riding lawn mower, rendered from 30FPS video. Note that the resulting image contains two displaced images of the mower. The first mower image corresponds to the first spike in the integration gain function. It is mostly opaque, picks up suddenly with a sharp leading edge, then slowly fades away in a smear as though it were moving quickly relative to the shutter speed. The relative opacity is because the majority of the light energy integrated at those locations in the frame come from the first spike, while the sudden appearance and slow taper are the result of the shape of that first spike. The second mower image, further down the row, corresponds to the second spike. It is relatively sharp and un-smeared because the width of the spike is short compared to the speed the mower was moving. However, it appears relatively translucent because the majority of the light energy integrated at its location was contributed by the background in the earlier portion of the exposure, rather than the time when the mower was at that position.

One interesting detail of arbitrary integration functions is that it is perfectly possible to specify *negative* gain for some portions of the interval. This behavior would be physically analogous to the sensor subtracting the contribution of incident light during portions of the exposure, rather than adding it. This is not something that is physically realizable in a conventional camera, but is very useful creatively for tasks such as subtracting static features from a scene. Partially negative gain functions also provide a ready way to provide even average brightness for differently-integrated parts of the scene to compensate for intervals with particularly high gain applied.

Non-Uniform Over Space

Non-uniform integration over space is analogous to creating a scene-specific piece of film whose sensitivity to light varies across it's surface, or the practice or selectively lightening or darkening portions of a scene in post-processing. While applying different gains to different portions of the scene in order to properly expose each is obvious, specifying the spatial regions on which to apply the different gains is somewhat mechanically awkward. Extended discussions on the matter resulted in several unappealing options - specifying by mathematical function (awkward for the user), specifying by rectangular region (restrictive), specifying by arbitrary polygons (complicated and restrictive), or specifying by bucket-fill algorithm (computationally difficult, self feedback problems) - and one promising avenue.

The promising method is to use a mask, drawn as a bitmap of the same resolution as the capture source, colored with a different pixel value for each region to be processed. This covers all the functionality of arbitrary geometry or function-determined regions by offering users a simple, portable, well-known format to generate their complex masks in, while allowing straightforward use cases to simply draw their desires in a basic image editor. A bitmap mask for region definition also allows for straightforward batch processing, either by applying the same generated or otherwise pre-prepared mask for multiple exposures, or enabling the use of an external tool to perform higher-level per-frame functions, such as object tracking, to generate sequential video frames from a TDCI stream with specific exposure properties for different objects in the scene.

The format currently being used to express these spatial exposure masks is a 8-bit P2 PGM with the same spatial size as the TDCI stream to be exposed. Each of the 255 gray levels possible in the format represents a distinct region, and the value of each pixel in the mask PGM specifies which exposure region to apply to the corresponding pixel in the input stream. This way, a simple gray-scale mask can be generated where each pixel in the mask is tagged with the encoded value of the gray level. Each gray level is then assigned a particular integration function. This provides a number of regions far in excess of any easily-conceived practical application, avoids forcing users to deal with any specialty tools or mathematical specifications, and is extremely straightforward for software to both generate and ingest. Each numbered region is then assigned an integration function with which to "expose" that portion of the image.

This technique is quite general. It is possible to not only vary the integration parameters for portions of the scene, analogous to existing HDR techniques, but to directly produce temporally composite exposures. In such an exposure, the output image is generated from sections which spatial sections of the image are are integrated from temporally differently-centered and possibly non-overlapping sub-sections of the sampling interval, opening a wide range of options. One simple application for this combined case might be selecting independent intervals for each face in a group picture to give each pictured individual open eyes and pleasing expressions, even though though they did not happen at the same time. Another application could be masking a moving object from the background in a scene, and integrating the moving part with a short, sharp-edged, high-gain integration function, and the background with a longer, shallower, lower-gain integration function on the same center, yielding an image with a sharp object with a minimum of motion blur on an apparently well-lit, detailed background.

Including a simple editor which would provide a transparent overlay of a frame preview and some basic drawing tools to generate region mask bitmaps inside the TDCI exposure tool is an obvious nice-to-have, but is not in the critical path for demonstrating the technological features. This is especially reasonable as most users likely to be using complicated exposure behaviors are likely to already have a deep familiarity and established workflow with their image editing tool of choice, and staying out of their way may even be the better choice in general.

To provide a minimal illustration for the effects possible with this mechanism, 4 contains a simple top/bottom split mask, specifying the two regions on which to apply the two integration functions in 3. This mask and function are then applied to a 240FPS video of a foam penguin and foam rock swinging like a pendulum while attached to the same string, resulting in 5. Note that the penguin, integrated with the function with two wider peaks, appears in two places, with a relatively large amount of motion blur indicating the two longer intervals of integration, while the rock, integrated with the single very narrow peak, appears only once, and relatively sharp-edged, corresponding to the single narrow peak. Also note that the offsets between the peaks and the appearance of blurring in the image provide a tell that the images of the penguin appear from light contribution during the right-to-left traversal, while the light contribution from which the rock is taken is from the left-to-right return trip, despite appearing "between" the two images of the penguin.

The Prototype

The current proof-of-concept implementation is in the form of an Octave script which can ingest a sequence of video frames to simulate continuously sampled incident light, and apply user-specified spatial and temporal non-uniformities to the integration of that light. This proof-of-concept implementation serves primarily as a testbed for algorithms and representations, as well as an easy way to experiment with the effects which can be produced, and is not tuned to be particularly fast or high quality.

In this prototype, spatial non-uniformity is specified by an 8-bit PGM mask as proposed above. Each region to be integrated is assigned a unique gray value, and each gray value is mapped to a corresponding temporal gain function by a simple table of gray value:function index correspondences. Directly encoding the index of the gain function to be used as the gray value was rejected, as numerically adjacent gray vales are indistinguishable to a human observer, and storing the control point vectors in a sparse representation adds more complexity to the prototype than simply re-mapping the indexing.

Likewise, in this prototype, integration gain functions for temporal non-uniformity are specified by a series of p user-supplied control points per function. These control points are interpolated with a normal cubic spline to give the gain to be applied at time t into a sampled interval of length t_{max} by evaluating the interpolation at $(t/t_{max}) * p$. This means every integration gain function is mapped to the entire input interval, so the granularity of the function can be increased by inputting a larger number of control points, and the position inside the sampled interval can be accomplished by zero padding, without the addition of any other constructs. An arbitrary number of control points can be specified

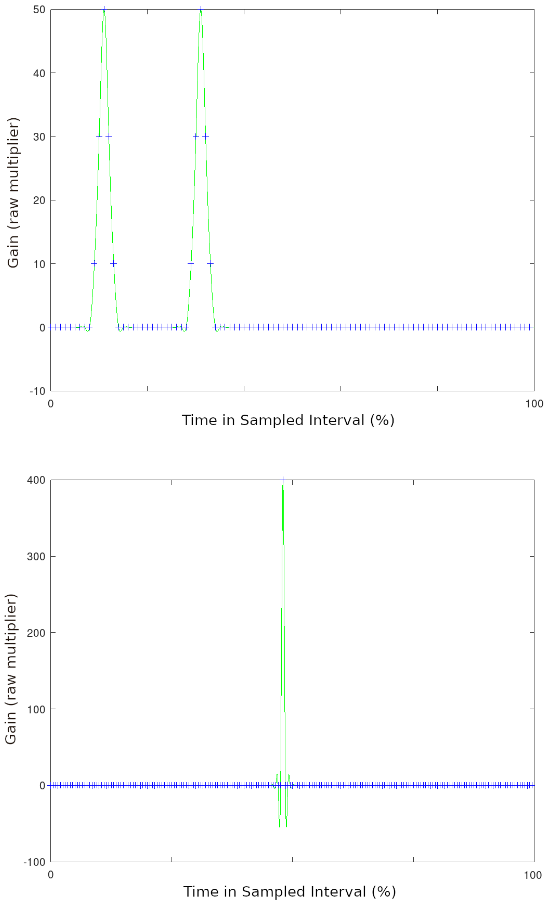

Figure 3. *Two integration functions*

to produce an approximation of any desired integration function, and the multiple functions specified for multiple regions are not required to contain the same number of control points.

The approximation to incident light to be integrated is generated by extracting successive from a video sequence, and treating each as an interval of contributed light for $1/framerate$ seconds. Integration is performed by summing the contribution of each input frame, multiplied by the average value of the mask-specified gain function for the interval represented by the frame, and subsequently dividing this weighted sum image by the number of frames integrated over to normalize the exposure.

This initial prototype does not attempt to interpolate between samples, as was done in the TIK TDCI testbed [3], but a version built on top of TIK is currently under development based on the algorithms demonstrated in this work, and should result in higher-quality output with less dependency on the frame-rate of the video input, and much better performance.

Results

This proof-of-concept implementation has demonstrated the feasibility of using the TDCI paradigm to generate virtual exposures with physically impractical exposure parameters. This mechanism shows promise both for a variety of creative applications, as well as potential for use in scientific applications. Specifically, using this technique to perform HDR stacking with different exposure parameters for different parts of the scene, all centered on the same instant in time allows a photographer to (iteratively) develop the dynamics of an exposure without the common problems of stitching errors from scene motion or incorrectly pre-parameterized exposure settings. Likewise, the function-driven integration can be used as a more flexible alternative to stroboscopic photography for capturing and analyzing motion, by sampling a scene then imposing a pulse train exposure function until the desired effect is achieved.

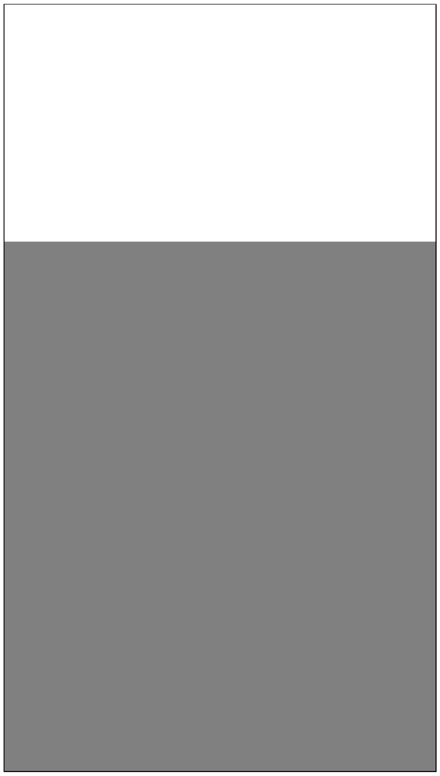

Figure 4. *Mask specifying two areas of the scene to integrate with different functions.*

Visualizing the effects achieved by these methods, much less their compelling applications, is still rather difficult, since many of them are not achievable with any physically realized camera. So far, the most effective way of visualizing the tool's behavior is to imagine a camera with a focal-plane shutter consisting of an extremely transmissive, extremely fast, LCD of resolution high enough that its dot size is not the limiting factor in the optical path. Once can then think of displaying the imposed spatial and temporal functions on this screen in the optical path and - more or less - predict the properties of the resulting image.

Future Work

The current implementation suffers from several deficiencies. The performance, in terms of both quality of temporal interpolation and the amount of time required to render an exposure, is rather disappointing. However, independent of those factors, the existing prototype has proven the viability of the algorithms and techniques. Based on that promise, a second generation prototype built atop the TIK TDCI testbed [3] is already underway.

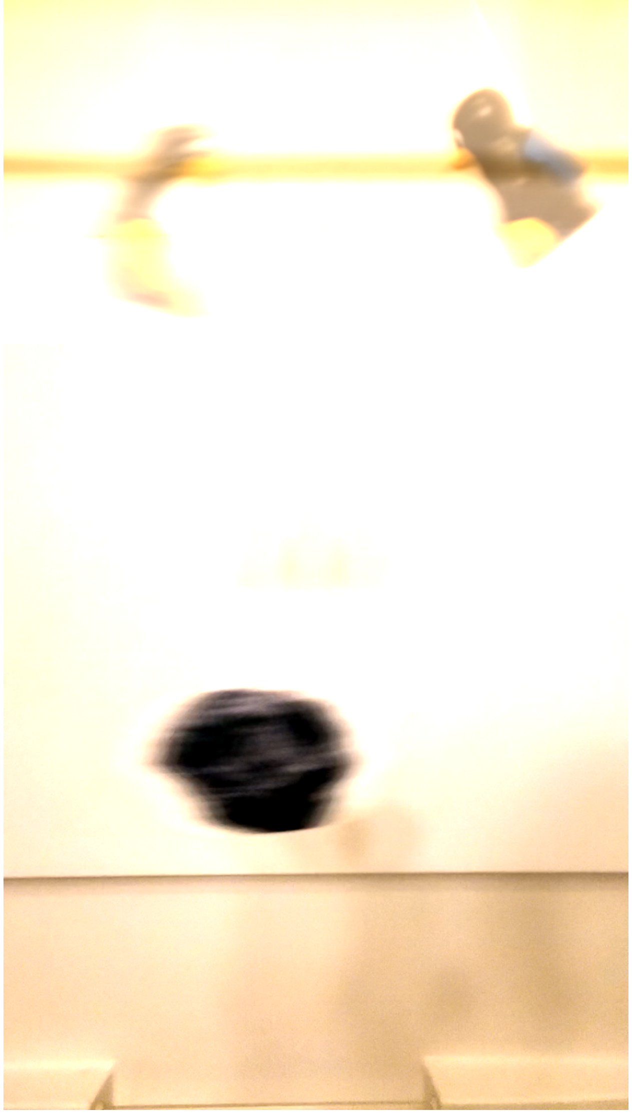

Figure 5. *Capture of a single pendulum carrying a foam penguin and rock, exposed with the mask and functions from fig. 3 and 4*

Unfortunately, this second iteration is not fully working at time of publication.

One major problem not following from the limitations of the existing prototype is that specifying functions an exposure masks is rather awkward. Textual lists of control points and hand-crafted PGM masks are adequate to verify the functionality of the algorithms, but more flexible, rapid, and generally user-friendly tools will be required for general use. This will likely be accomplished via a graphical tool backed by a specification language which can define all the necessary properties of an exposure in a standard human- and machine- manipulable way, but the process is only now being explored in enough detail to anticipate the requirements of such tools.

Perhaps the most interesting feature of this work is that because arbitrarily spatially and temporally non-uniform exposures have not previously been available to photographers, early proposed uses for the tool have all been by analogous to existing techniques. As more users gain more experience with this mode of photography, new and interesting techniques are already beginning to suggest themselves, but are not yet well-understood.

Bibliography

[1] H. G. Dietz, “Frameless, time domain continuous image capture,” in *Electronic Imaging 2014*, vol. 9022, 2014, pp. 7–12. DOI: 10.1117/12.2040016. [Online]. Available: http://dx.doi.org/10.1117/12.2040016.

[2] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, “Event-based vision: A survey,” *CoRR*, vol. abs/1904.08405, 2019. arXiv: 1904.08405. [Online]. Available: http://arxiv.org/abs/1904.08405.

[3] H. Dietz, P. Eberhart, J. Fike, K. Long, C. Demaree, and J. Wu, “Tik: A time domain continuous imaging testbed using conventional still images and video,” *Electronic Imaging*, vol. 2017, no. 15, pp. 58–65, 2017, ISSN: 2470-1173. DOI: doi:10.2352/ISSN.2470-1173.2017.15.DPMI-081. [Online]. Available: http://www.ingentaconnect.com/content/ist/ei/2017/00002017/00000015/art00010.

[4] H. Dietz and P. Eberhart, “Shuttering methods and the artifacts they produce,” *Electronic Imaging*, vol. 2019, no. 4, pp. 590-1-590–7, 2019, ISSN: 2470-1173. [Online]. Available: https://www.ingentaconnect.com/content/ist/ei/2019/00002019/00000004/art00010.

[5] H. G. Dietz and P. S. Eberhart, “Iso-less?” In *Electronic Imaging 2014*, vol. 9404, 2015, pp. 94040L-94040L–14. DOI: 10.1117/12.2080168. [Online]. Available: http://dx.doi.org/10.1117/12.2080168.

[6] H. Dietz, P. Eberhart, and C. Demaree, “Multispectral, high dynamic range, time domain continuous imaging,” *Electronic Imaging*, vol. 2018, no. 5, pp. 409-1-409–9, Jan. 28, 2018, ISSN: 2470-1173. DOI: doi:10.2352/ISSN.2470-1173.2018.05.PMII-409. [Online]. Available: https://www.ingentaconnect.com/content/ist/ei/2018/00002018/00000005/art00012.

[7] R. Burden and J. Faires, “Numerical analysis,” in. Cengage Learning, 2010, ch. 3, pp. 144–153, ISBN: 9780538733519. [Online]. Available: https://books.google.com/books?id=zXnSxY9G2JgC.