2016 IEEE International Symposium on Multimedia

Feedback Control System for Exposure Optimization in High-Dynamic-Range Multimedia Sensing

Ryan Janzen and Steve Mann
University of Toronto
Department of Electrical Computer Engineering
http://eyetap.org

***Abstract*—We introduce the concept of coupled dynamic dynamic-range (D^2R) compositing – that is, assembling sensor information, such as images or audio, from multiple dynamic-ranges that are allowed to move and change over time, as lighting conditions or sound conditions change over time in their amplitude-domain properties. We propose a feedback-control method to automatically adjust multiple exposure-value settings for HDR compositing, to increase the dynamic range of a sensory process, such as video capture. The method uses a cost function to express measurement uncertainty, and a "cross-uncertainty" metric between different exposures, fed into a dynamic control system. The system is designed to asymptotically approach an optimal distribution of camera exposure control settings, under varying lighting conditions and motion, to capture an extremely high dynamic range for HDR compositing.**

***Keywords*—high dynamic range, HDR sensing, HDR video, HDR audio, HDR welding, composite dynamic range, CDR**

I. Introduction

All sensors have a limited dynamic range. The dynamic range of a sensory process can be improved by high dynamic range (HDR) compositing, where multiple exposures (images [1], audio samplings [2], *etc.*) of a physical phenomenon can be taken at different exposure settings, either using a single sensor that switches between settings in a sequence, or using an array of differently-configured sensors. Finally the data is composited into an output that covers a wider dynamic range than that of one single capture of the signal. In this way, it is possible to overcome the limited dynamic range of a camera, audio recorder, or other sensor. (Fig. 1)

Research has shown how to intelligently combine the data from multiple exposures, while accounting for the particular response function of a sensor [1] [3] [4] [5] [6] [7] [8]. For example, a camera's nonlinear response function can be determined and reversed for each exposure, and based on the result, each pixel in each exposure can be weighted according to the precision (degree of certainty) each exposure's pixel gives to the combined measurement [1][4]. Pixels whose values are near the extrema (either cutoff near zero, or saturated near the maximum value) are given the lowest possible weighting.

A composite-dynamic-range signal can then be reconstructed, after compensating for the different signal strengths entering the sensor for each exposure. We adapted this process for audio and other time-varying signals [2]. See Fig. 3.

However, the choice of exposure (*e.g.* gain) settings is important to the process, since unknown indeterminate samples or pixels may result due to lack of information about the

Fig. 1. Dynamic range compositing, to form a high dynamic range (HDR) image from multiple sensor readings (exposures), where the sensor's dynamic range is not sufficient to capture the full dynamic range of the original signal. Notice the black areas in the left image, and white saturated areas in the right image. The combined dynamic range is incorporated into the output image (bottom). This paper targets extreme dynamic range situations, such as photographs of the sun in space, video of arc-welding, or seismic waves capturing a mouse whispering during an earthquake.

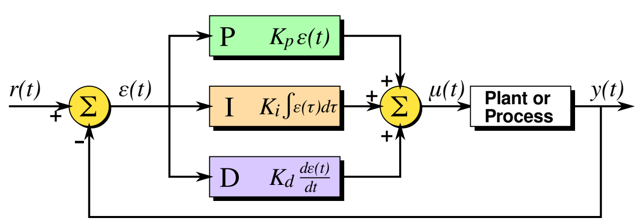


Fig. 2. Proportional-Integral-Derivative (PID) Controller: A typical configuration, where a control system is controlling a parameter $y(t)$ in a system or process, called the "plant". In this case, the "plant" we wish to control is the sensor, under the conditions of the physical phenomenon it is sensing. Negative feedback is the key principle, where an error measurement $\epsilon(t)$ is minimized in real-time. In our case, we use a metric of "uncertainty" in the sensory process, at all times attempting to track it along a reference value of zero. $\epsilon(t)$ is the error or discrepancy between this reference and the actual observation. The PID control system is designed to control the plant while promoting stability of the overall system.

quantity of light at those locations. This may occur if there are insufficient exposures, or if the specific exposures are not chosen properly. A common practice is rather unscientific: to plan for a specific dynamic range capability, and then try a number of different exposure settings until satisfied by visual inspection of the output image. Terraced lines are one

978-1-5090-4571-6/16 $31.00 © 2016 IEEE
DOI 10.1109/ISM.2016.83

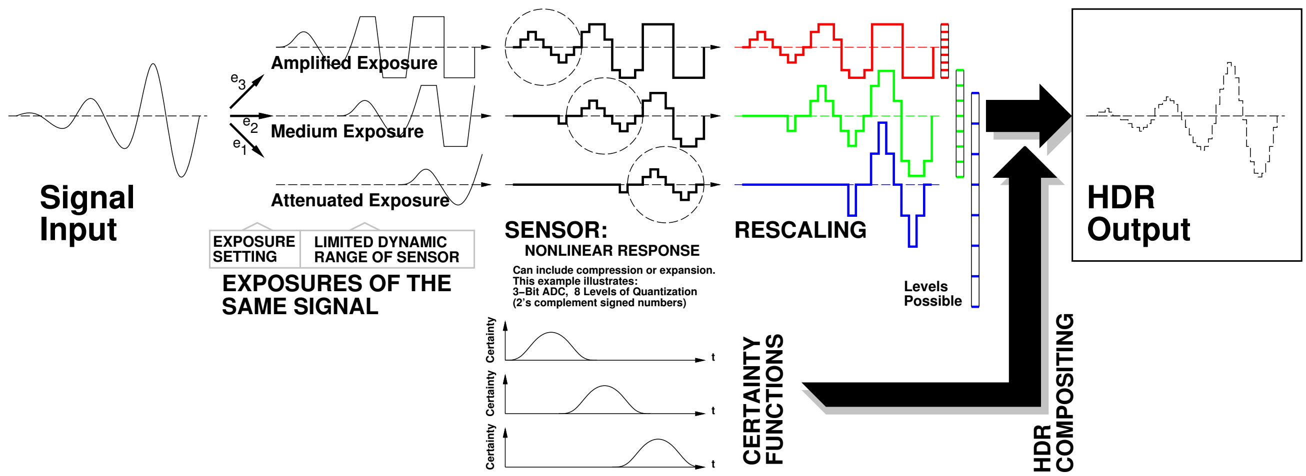


Fig. 3. Dynamic range compositing: When time-varying signals such as audio and images are composited, an important factor is any nonlinear response of the sensor. Beyond simple quantization, this can also include compression functions (e.g. $\sqrt(x)$, $log(x)$), or expansion functions (e.g. x^2, e^x). A certainty function can express the salience of information received for each numerical value of each sample.

recognizable feature in a HDR output image, associated with an undesirable gap between exposures. Features like these aid in visual inspection and improvised choices of static exposures.

A static exposure optimization method was developed for HDR compositing of time-varying signals [2]. This work found a set of constraints used to control exposure settings, based on the properties of a time-varying signal such as light or sound. This method used an "exposure packing" dynamic range to compute the values of exposure gains. However, this work did not dynamically control the exposures.

Automatic gain control (AGC) is a method to dynamically adapt to different signal strengths. However, there are two flaws with ordinary AGC: The dynamic range is only matched to one amplitude regime at a time (very strong and very weak signals cannot be sensed simultaneously), and information is lost about the original signal strength.

We could call this "Generation-1 AGC", and suggest a "Generation-2 AGC" that records the exposure gain at all points in time. Still, there is no simultaneous sensing of strong and weak signals, such as direct sunlight and dim shadows in the same image.

Automatic exposure bracketing (AEB), or auto-bracketing, uses a similar principle to AGC with HDR: A series of exposures are chosen with fixed ratios between them. Cameras currently on the market with auto-bracketing use AGC to find one middle exposure, and use a pre-set ratio to choose two or more other exposures at a fixed ratio away from the middle exposure [9][10][11].

Generation-3 AGC could be defined as multiple separate AGC units, each working separately from each other as if they were different cameras or other unconnected sensor exposures [12][6][13].

We propose a Generation-4 AGC, where a dynamic coupled control system adapts to signal conditions, and *jointly* controls a *vector of exposure values*. This builds on previous work by not only merging dynamic-range data in real-time, but also dynamically and jointly controlling a vector of *exposure settings themselves* in real-time.

II. Dynamic^2-Range Compositing

A key limitation of previous work has been its lack of adaptation to time-varying conditions. For example, an HDR video processor might use two exposures which capture a complete dynamic range in an office setting, but may be woefully inadequate when brought to a welding facility with low light levels plus extremely bright light at small points where the welding is taking place.

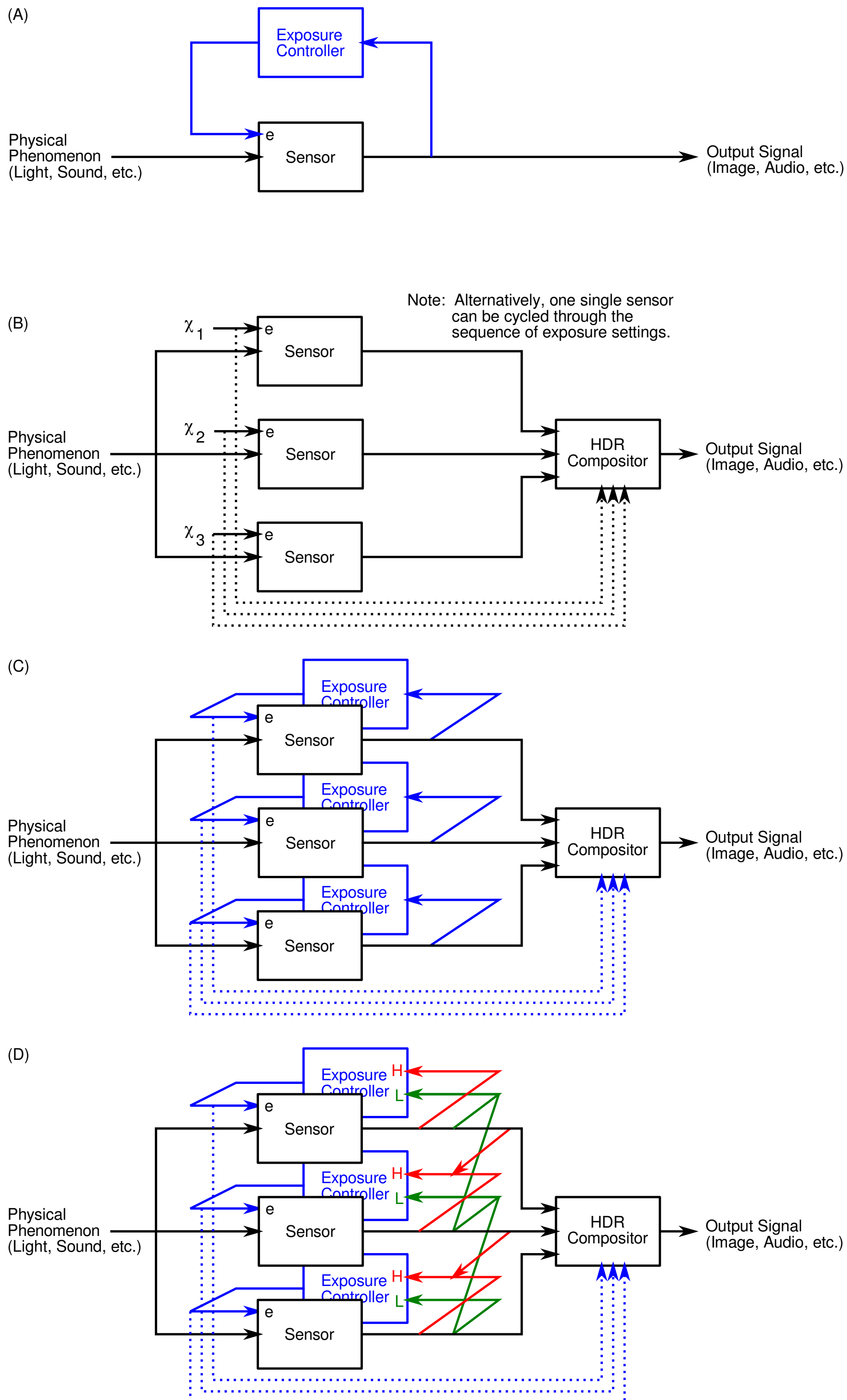
The result can be portions of an image that are all black, while the welding work is seen clearly (the case of dark exposures), or a completely white, saturated, indeterminate region where the welding takes place, and the rest of the room is visible (the case of bright exposures). This static nature of HDR exposures is illustrated in Fig. 4(B).

Instead, we wish to make a *dynamic* and *coupled* control system to control dynamic ranges of each exposure, through the exposure settings. See Fig. 4(D). Adding a feedback loop allows asymptotic approach of an optimum, and adaptation to time-varying lighting conditions (in the case of a camera) in concert with the particular nonlinear response of the camera. Furthermore, we perform *joint* control of the exposures, based on joint metrics of their real-time response to light.

First, we explain the dynamic range compositing system, then its metrics for optimization, and then the control system.

III. Linear v.s. Nonlinear Exposures

We can control each of the $e = 1...E$ exposures, each set with parameter χ_e. In a camera, for example, several different settings can be changed: 1. exposure speed; 2. aperture size (F-stop); 3. ISO via image sensor sensitivity.



Automatic Gain Control (AGC)

- Loss of information of original signal strength
- Failure to capture outlier elements in array sensing (e.g. saturated image pixels)

Typical D.R. Compositing

- Dynamic Range is covered by multiple exposures
- Predefined exposure settings
- Risk of information loss if signal becomes larger or smaller than the anticipated dynamic range.
- Risk of inefficient D.R. coverage.
- Excessive sacrifices to sample/frame rate due to inflexible exposure settings.

Isolated Dynamic Dynamic Range ($\mathbf{D^2R}$) Compositing

- Dynamic response to varying input signal conditions.
- Independent AGC units, each working separately on unconnected sensor exposures.
- Inefficient coverage of the entire dynamic range, due to uncontrolled exposure overlap.

Coupled Dynamic Dynamic Range ($\mathbf{D^2R}$) Compositing

- Dynamic response to varying input signal conditions.
- Continuous optimization of dynamic range coverage.
- Efficient coverage of the entire dynamic range, by joint, coupled control of exposure overlap.
- Each exposure balances its own uncertainty forcing functions against those of its neighbours.

Fig. 4. Combining exposure control with HDR compositing: (A) A typical autoexposure camera or sound recorder with automatic gain control (AGC) uses a feedback control to vary the exposure or gain setting. (B) Typical HDR compositing uses a fixed set of gains or exposures, to try to capture an entire dynamic range. However, these settings cannot anticipate the particular subject matter in a photo, for example, where a photo of the sky might produce many pixels which lie in a suboptimal portion of the dynamic range. (C) We add dynamic feedback control of exposures. (D) By cross-linking the exposure control, we dynamically control each exposure according to a cost function based on the response of neighbouring exposures.

To overcome the nonlinear response of a typical camera, we convert the tonal range of the image into an equivalent tonal range of physical light levels, for each pixel. The available tonal range given from an image is referred to as *imagespace* [14], and typically ranges in value from 0 to 255 for an 8-bit image, across red, green and blue channels.

Instead we calibrate a nonlinear model of the camera's response function, which converts pixel value into an estimated true quantity of light. This true, physical range of values is referred to as *lightspace* [14].

Audio systems, on the other hand, are designed to be as linear as possible in their response, unlike a camera. In this case, the exposure setting is a linear gain value, $\chi_e = g_e$, and the dynamic ranges of the exposures can be re-aligned by a factor of $1/g_e$. However, audio HDR requires harmonic certainty functions that are designed for the harmonic nature of audio signals [2], unlike image-based HDR, where the harmonics of each pixel typically do not need to be considered.[1]

IV. Optimization Cost Functions

HDR exposures could be chosen such that each of the dynamic ranges neatly abut against each other, where one exposure saturates exactly at the point when the next exposure barely senses a signal. Is that sufficient?

If the sensor's response function, $f(q)$, has low-precision anywhere in its dynamic range $[q_{\text{min}}, q_{\text{max}}]$, then a greater amount of overlap with another exposure is desirable at that point in the dynamic range.

First, we define an input signal received from the sensor:

$$\rho = f(q) \tag{1}$$

The original physical signal, q, might be at a variety of different levels (bright v.s. dark, or loud v.s. soft), but the only observable from the sensor is its output signal, ρ.

Precision of the sensor within its dynamic range can be expressed by an uncertainty function, $u(\rho)$, for each possible sensor output ρ. For example, if the sensor's nonlinear response function is smooth and continuous, the uncertainty function could be given as:

$$u(\rho) = \left(\frac{\partial f}{\partial q}\right)_{f^{-1}(\rho)} \tag{2}$$

In this case, the more coarsely spread is the sensor's response at a particular signal level, the less precision is available about the original phenomenon, and thus the less weighting is to be placed on that sample/pixel from that exposure, as opposed to the same pixel from other exposures [3]. Uncertainty is thus a measure of the degree of usefulness of each sample/pixel in an HDR composite, based on the particular exposure settings.

We compute an *uncertainty image* $U_e(x,y)$ for each exposure frame/image $I_e(x,y)$, composed of the uncertainty function of each sample/pixel. We then compute a set of optimization cost functions:

Supra-uncertainty, a cost function expressing a penalty on near-saturation, strong samples/pixels in each exposure. It influences the exposure, forcing it down to a lower amplification.

$$u_{H,e} = \sum_{x,y} \frac{I_e(x,y) - \rho_{\text{min}}}{\rho_{\text{max}} - \rho_{\text{min}}} \cdot \left(\frac{\partial f}{\partial q}\right)_{f^{-1}(I_e(x,y))} \tag{3}$$

Infra-uncertainty, a penalty on weak signals near cutoff:

$$u_{L,e} = \sum_{x,y} \frac{\rho_{\text{max}} - I_e(x,y)}{\rho_{\text{max}} - \rho_{\text{min}}} \cdot \left(\frac{\partial f}{\partial q}\right)_{f^{-1}(I_e(x,y))} \tag{4}$$

Cross-uncertainty a penalty or cost function, expressing a joint uncertainty caused by two adjacent exposures:

$$u_{C,e}(I_e, I_{e+1}) = \sum_{x,y} \min\left(U_{L,e}(x,y), U_{H,e+1}(x,y)\right) \tag{5}$$

As an intermediate step, the per-pixel *uncertainty images* can be seen in Fig. 7:
$U_{H,e}(I_e)$ supra-uncertainty image (penalty array expressing pixels whose uncertainty is caused by saturated signals)
$U_{L,e}(I_e)$ infra-uncertainty image (penalty array expressing pixels whose uncertainty is caused by cutoff, weak signals)
$U_{C,e}(I_e, I_{e+1})$ cross-uncertainty image (penalty array expressing pixels with a joint uncertainty caused by two adjacent exposures)

We control each of the exposure settings, χ_e, for the e^{th} exposure. The feedback loop is completed with the sensor capture $I_e(\chi_e, t)$ for each exposure, resulting from those exposure settings.

The control system is governed by two counterveilling influences. Each exposure setting is influenced by a push-pull mechanism from *both ends of the dynamic range* of each exposure. This signal is expressed as Δu, accounting for the difference in uncertainties between successive exposures.

$$\Delta u_e = \left\{ \begin{array}{ll} u_{C,e}(I_e, I_{e+1}) - u_{H,e}(I_e), & \text{for } e = 1 \\ u_{C,e}(I_e, I_{e+1}) - u_{C,e-1}(I_{e-1}, I_e), & \text{for } 1 < e < E \\ u_{L,e}(I_e) - u_{C,e-1}(I_{e-1}, I_e), & \text{for } e = E \end{array} \right\} \tag{6}$$

The basic idea is to compare uncertainty caused by high-valued pixels and low-valued pixels. In two adjacent exposures, this involves the cross-uncertainty of one and the cross-uncertainty of the next, except in the case of the first and last exposures, where we must use an absolute uncertainty caused by saturation in that given exposure alone.

This will be compared to a mass-spring system in a following section.

These cross-linkages implement the cross-effects illustrated in 4D. A smooth and continuous control response is made possible by accounting for the floating-point difference in uncertainties, rather than simply incrementing and decrementing the exposure setting.

V. PID Control

To complete the exposure control system, we used a PID (proportional integral derivative) control, as shown in Fig. 2

[1]Important issues with audio DR-compositing also include quantization (to which the human ears can sense much more keenly than with light), and the masking of one frequency range over another frequency range, between the dynamic ranges of different exposures (which we define as *eclipse*).

to govern the exposure change velocity, as follows:

$$\nu_e(t) = K_P \cdot \Delta u_e + K_I \cdot \int_0^t \Delta u_e(\tau) d\tau + K_D \cdot \frac{d}{dt} \Delta u_e \quad (7)$$

Finally, the exposure setting $\chi_e(t)$ is composed of the exposure control velocity ν_e, as:

$$\chi_e(t) = \int_0^t \nu_e(\tau) d\tau \quad (8)$$

VI. Physical Analogy

This camera control system can be compared to a mass-spring system in physics. The exposure setting is analogous to the position of the mass, and the uncertainty cost functions are analogous to forces on the mass.

First we imagine one force controlling one mass: This is analogous to a single exposure, with only one factor controlling it, according to uncertainty in the image. If $\Delta u(\chi)$ were to behave linearly with the negative of χ, and if we only used the "I" term of the PID controller, the velocity control function $\nu(\Delta u)$ becomes:

$$\nu_e(t) = K_I \cdot \int_0^t \Delta u_e(\tau) d\tau \quad (9)$$

and therefore, the exposure setting is:

$$\chi_e(t) = K_I \cdot \int_0^t \left[\int_0^t \Delta u_e(\tau) d\tau \right] d\tau \quad (10)$$

Therefore, the forcing function behaves according to Hooke's law, representing the physics of a spring: $F = -k\chi$. This would produce a simple harmonic motion, with amplitude A, angular frequency ω, and phase ϕ:

$$\chi(t) = A \cos(\omega t + \phi) \quad (11)$$

For two exposures, the equivalent mass-spring dynamics are:

$$m \frac{d^2 \chi_1}{dt^2} = -k\chi_1 + k(\chi_2 - \chi_1) \quad (12)$$

$$m \frac{d^2 \chi_2}{dt^2} = -k\chi_2 + k(\chi_1 - \chi_2) \quad (13)$$

This leads to two independent *normal modes* of oscillation:

$$\vec{\eta}_A = \begin{pmatrix} \chi_1^A(t) \\ \chi_2^A(t) \end{pmatrix} = c_A \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cos(\omega_A t + \phi_A) \quad (14)$$

$$\vec{\eta}_B = \begin{pmatrix} \chi_1^B(t) \\ \chi_2^B(t) \end{pmatrix} = c_B \begin{pmatrix} 1 \\ -1 \end{pmatrix} \cos(\omega_B t + \phi_B) \quad (15)$$

In our system, by using damping, the "P" term, and the compressive nonlinearity of an image sensor, we can prevent sustained oscillation of the system.

Interestingly, though, we can observe wave-like motion of the exposure settings, which is analogous to the mathematics of high-order mass-spring systems. In the theory of coupled resonators, it can be mathematically shown that as the number of masses and springs approaches infinity, the solution to the equations of motion approaches that of a travelling wave [15]. (This yields the underlying physics behind physical waves such as sound, light, and water waves!)

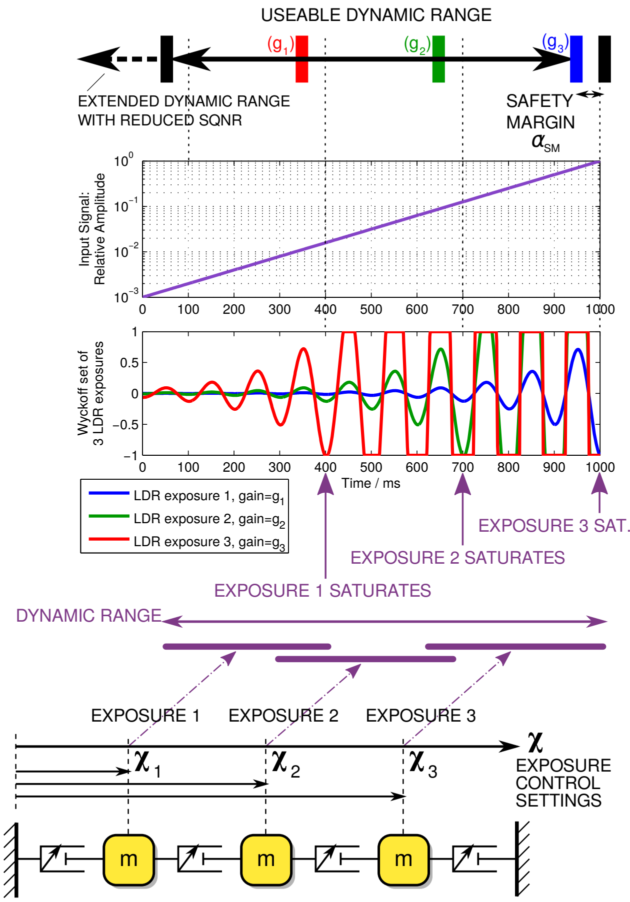


Fig. 5. Physical analogy for coupled motion of the exposure controllers. A forcing function between each of the exposures can be imagined analogously to a series of pistons (bottom of diagram) providing a forcing function between a series of masses. The forcing functions are controlled by the uncertainty of the sensed signals, resulting from the choice of exposure settings at each point in time.

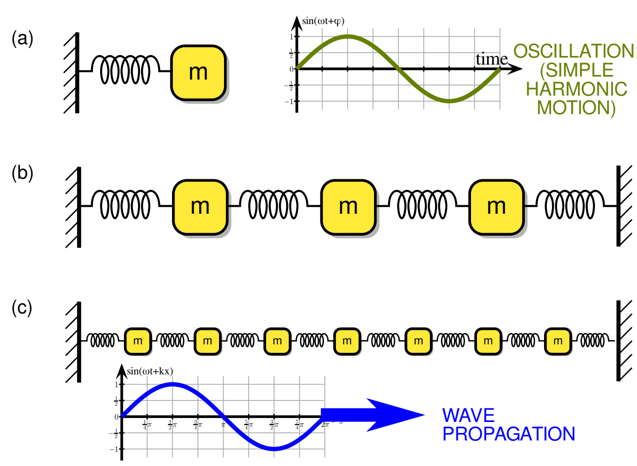


Fig. 6. Mass-spring analogy, for coupled motion of the exposure controllers. (a) Single mass-spring. (b) Analogy to a 3-exposure implementation of this system. (c) Analogy to a higher number of mass-springs. A perturbation on the masses can propagate through the series like a wave. This is analogous to a bright flash of light in the video stream, offsetting the HDR exposures until they re-adjust to optimize for the new conditions.

VII. Implementation and Results

The coupled dynamic dynamic-range (CD^2R) compositing system was implemented and evaluated using MATLAB code interfaced to a camera. To ensure accuracy of exposure alignment, we implemented the option whereby one single sensor was used, and exposures were extracted by cycling the sensor between different exposure settings.

We used an off-the-shelf USB camera image sensor, which outputted 8-bit images whose pixel-values ranged from 0 to 255 in each of the red, green, and blue channels. This provided a worst-case test of the control system due to a very limited dynamic range.

An example of the system's operation is illustrated in Fig. 7, showing the network of interconnected uncertainty metrics between the exposures. Figs. 8 and 9 show the time evolution of three exposures (dark, medium and light), under the control of the algorithm. The control system attempts to minimize the mutual uncertainty between the exposures over the combined dynamic range. It attempts to ensure that sufficient information is known about every pixel; for example, that no pixel is saturated in all three input exposures.

The final result is a composited image, where each pixel is composed of tonal information from at least one of the three corresponding input exposure pixels. In all cases, each triplet of corresponding input exposure pixels are combined by compensating for the known gain of each exposure, and the nonlinear response function of the camera for each exposure.

It is interesting to observe the wave-like propagation between exposures: the first exposure moves quickly out from its starting position, then begins pulling on the second exposure.

Finally, the system performance was compared alongside the performance of a traditional static-exposure HDR system. In Fig. 10 we can see a reduction in exposure uncertainty, aggregated across the time period of a video sequence. Our system takes time (a few video frames) to adapt to the changing conditions as the camera is moved (during which performance suffers over the duration of a few frames), but overall, it approaches a more optimal set of exposures which dynamically adapt to the changing conditions.

VIII. Conclusion

We have devised a new method for automatic exposure-setting control, to enable coupled dynamic dynamic-range (CD^2R) video compositing. Rather than an HDR system that needs to be tuned for each lighting scenario (e.g. indoors with two exposures, and then retuned outdoors with three exposures when viewing the sun or welding), the feedback control system adapts to the dynamic histogram of each exposure image, to control all the exposure settings in tandem.

The dynamic-range is "covered" by a set of exposures which are moving over time, in response to varying conditions of the original signal.

Welding vision systems, autonomous robot and spacecraft vision systems, acoustic recorders in geology/mining, and scientific cameras and signal recorders, are all particular applications of this system, requiring extreme dynamic ranges with unpredictable, nonstationary signals.

Fig. 7. Snapshot of the control system operation, captured at the instant of the 20th image frame in a video sequence.

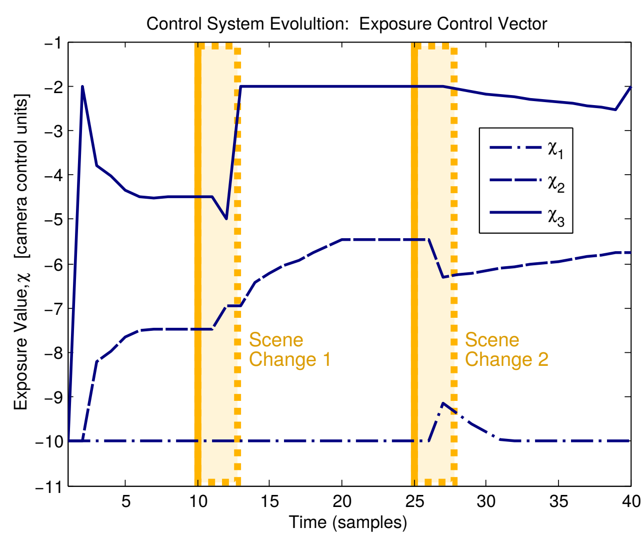


Fig. 8. Time-evolution of the exposure controller outputs, during two scene changes in a video sequence.

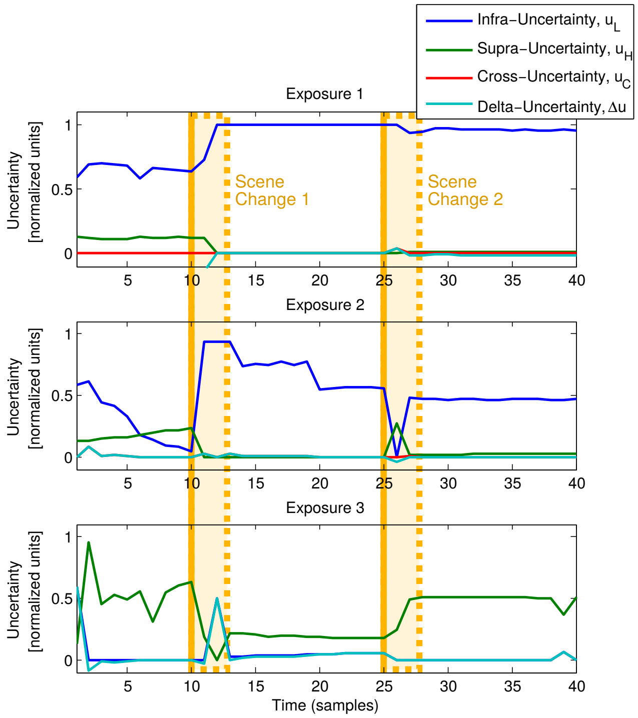


Fig. 9. Time-evolution of the exposure controller's uncertainty metrics, during the same video sequence as Fig. 8.

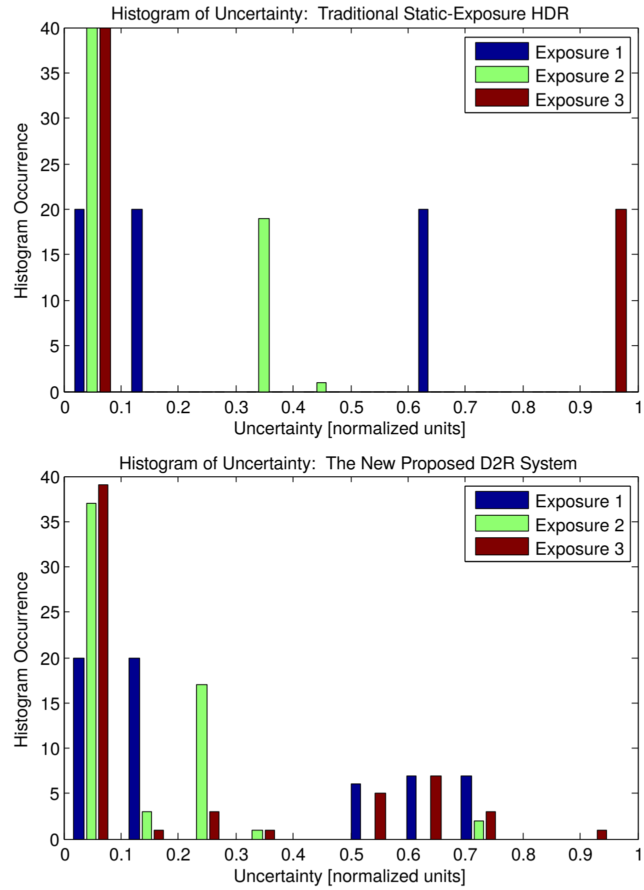


Fig. 10. Performance comparison of traditional HDR with static exposures (top), and the proposed system (bottom). Histograms show the amount of uncertainty integrated over time — that is, the time-aggregated uncertainty from non-optimal exposure responses to the given lighting conditions. The static exposure implementation suffers from poor uncertainty, seen as an emphasis on high-valued uncertainty levels on the right side of the upper histogram. In the lower histogram, the new dynamic system exhibits a locus of uncertainty between approximately 0.5 and 0.75 — with low emphasis, this is likely the brief periods of time when the dynamic system adapts to changing conditions. Note that the histogram bins are in groups of 3.

REFERENCES

[1] S. Mann and R. Picard, "Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proc. IS&T*, May 7–11 1995, pp. 422–428.

[2] R. Janzen and S. Mann, "High dynamic range simultaneous signal compositing, applied to audio," in *Proc. IEEE CCECE 2012, Montreal*, April 29 to May 2 2012.

[3] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *IEEE Trans. Image Proc.*, vol. 9, no. 8, pp. 1389–1406, August 2000, iSSN 1057-7149.

[4] ——, "Compositing multiple pictures of the same scene," in *Proc. IS&T Conference*. Cambridge, MA: The Society of Imaging Science and Technology, May 9-14 1993, pp. 50–52, iSBN: 0-89208-171-6.

[5] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *SIGGRAPH*, pp. 369–378, 1997.

[6] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," in *ACM Trans. Graphics (Proc. SIGGRAPH 2003)*, vol. 22, no. 3, 2003, pp. 319–325.

[7] E. Khan, O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," *ICIP*, pp. 2005–8, 2006.

[8] M. Robertson, S. Borman, and R. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using mult. exposures," *J. Electronic Img.*, vol. 12(2), pp. 219–228, 2003.

[9] H. Matsushima, "Camera having automatic exposure bracketing device," Jul. 26 1994, uS Patent 5,333,027.

[10] O. Satou, I. Hirai, and M. Nakata, "Auto-bracketing system," Apr. 16 1996, uS Patent 5,508,779.

[11] C. Loscos and K. Jacobs, "High-dynamic range imaging for dynamic scenes," *Computational Photography: Methods and Applications, CRC Press/Taylor & Francis*, pp. 259–281, 2010.

[12] L. E. Alston, D. S. Levinstone, and W. T. Plummer, "Exposure control system for an electronic imaging camera having increased dynamic range," Mar. 3 1987, uS Patent 4,647,975.

[13] P. Lapray, B. Heyrman, and D. Ginhac, "Hdr-artist: An adaptive real-time smart camera for high dynamic range imaging," *Journal of Real-Time Image Processing*, 2014, http://link.springer.com/article/10.1007/s11554-013-0393-7.

[14] S. Mann, *Intelligent Image Processing*. John Wiley and Sons, November 2 2001, iSBN: 0-471-40637-6.

[15] A. P. French, *Vibrations and waves*. CRC press, 1971.