An Adaptive Framework for Image and Video Sensing

Lior Zimet*, Morteza Shahram†, Peyman Milanfar‡
Department of Electrical Engineering, University of California, Santa Cruz, CA 95064

ABSTRACT

Current digital imaging devices often enable the user to capture still frames at a high spatial resolution, or a short video clip at a lower spatial resolution. With bandwidth limitations inherent to any sensor, there is clearly a tradeoff between spatial and temporal sampling rates, which can be studied, and which present-day sensors do not exploit. The fixed sampling rate that is normally used does not capture the scene according to its temporal and spatial content and artifacts such as aliasing and motion blur appear. Moreover, the available bandwidth on the camera transmission or memory is not optimally utilized. In this paper we outline a framework for an adaptive sensor where the spatial and temporal sampling rates are adapted to the scene. The sensor is adjusted to capture the scene with respect to its content. In the adaptation process, the spatial and temporal content of the video sequence are measured to evaluate the required sampling rate. We propose a robust, computationally inexpensive, content measure that works in the spatio-temporal domain as opposed to the traditional frequency domain methods. We show that the measure is accurate and robust in the presence of noise and aliasing. The varying sampling rate stream captures the scene more efficiently and with fewer artifacts such that in a post-processing step an enhanced resolution sequence can be effectively composed or an overall lower bandwidth for the capture of the scene can be realized, with small distortion.

Keywords: Adaptive Imaging, Varying Sampling Rate, Image Content Measure, Scene Adaptive, Camera Bandwidth

1. INTRODUCTION

Imaging devices have limited spatial and temporal resolution. An image is formed when light energy is integrated by an image sensor over a time interval. The minimum energy level for the light to be detected by the sensor is determined by signal to noise ratio characteristics of the detector.[4] Therefore, the exposure time required to ensure detection of light is inversely proportional to the area of the pixel. In other words, exposure time is proportional to spatial resolution. This is the fundamental trade off between the spatial sampling (number of pixels) and the temporal sampling (number of images per second). Other parameters such as readout and analog to digital conversion time as well as sensor circuit timing have second order effect on the spatio-temporal trade-off. Figure 1 is an example of the spatio-temporal sampling rate tradeoff in a typical camera (e.g. PixeLINK PL-A661). The markers along the graph are typical sampling rates used by digital image sensors for different applications. The parameters of the tradeoff line are determined by the characteristics of the materials used by the detector and the light energy level. A conventional video camera has a typical temporal sampling rate of 30 frame per second (fps) and a spatial sampling rate of 720×480 pixels, whereas a typical still digital camera has spatial resolution of 2048×1536 pixels.

The minimal size of spatial features or objects that can be visually detected in an image is determined by the spatial sampling rate and the camera induced-blur. The maximal speed of dynamic events that can be

*Email: lior.zimet@zoran.com
†Email: shahram@ee.ucsc.edu
‡Email: milanfar@ee.ucsc.edu

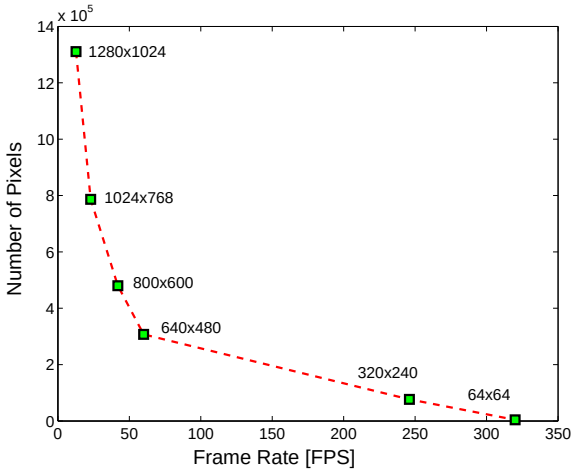


Figure 1. Typical tradeoff of spatial vs. temporal sampling in imaging sensor

observed in a video sequence is determined by the temporal sampling rate.[1] We define the sensor operating point as the pair of {spatial sampling rate, temporal sampling rate} at which the sensor is operating. A non-adaptive sensor is normally set to a fixed operating point, which does not depend on the scene. Therefore, the data from the sensor can be spatially or temporally aliased due to insufficient sampling rate.

Insufficient temporal sampling rate will introduce motion based aliasing. Motion aliasing occurs when the trajectory generated by a fast moving object is characterized by frequencies which are higher than the temporal sampling rate of the sensor. In this case, the high temporal frequencies are folded into the low temporal frequencies. The observable result is a distorted or even false trajectory of the moving object[1] (e.g. wheels on a fast-moving cart appearing to rotate backwards in a film captured at typical video rate). Meanwhile, insufficient spatial sampling rate will remove details from the image and introduce visual effects such as blur and aliasing.

Now, instead of relying on a single point on the spatio-temporal tradeoff curve (Figure 1), we could adapt the sensor to run at an operating point that is determined by the scene. An adaptive sensor would have the ability to change its operating point according to a measure of the temporal and spatial content in the scene. It therefore captures the scene more accurately and more efficiently with the available bit-rate or sensor memory or communication capabilities. The design of such a novel sensor can also be informed by user preferences in terms of acceptable levels of spatial or temporal aliasing or other factors. A Block diagram of an adaptive sensor is shown in Figure 2.

The spatial and temporal dimensions are very different in nature, yet are inter-related through the sensors capabilities. In the proposed adaptive sensor architecture we measure the spatial and temporal content separately to determine the required sampling rate for the current scene. We developed a robust, computationally inexpensive, measure for the scene content. The measure works in the spatio-temporal domain as opposed to the traditional frequency domain. We show that the measure is usable in the presence of noise and aliasing. Section 2 of the paper describes the content measure along with other possible methods.

The adaptive sensor measures the scene content continuously for every incoming frame. The required sampling rate is then determined from this measure. The required sampling rate can sometimes be out of the sensor's capabilities and a projection to the nearest possible operating point in the sensor's operating space is required. The conversion from the content measure to sensor operating point is discussed in Section 3. Using a feedback loop the sensor is reconfigured to the new operating point. The closed loop operation is described in Section 4.

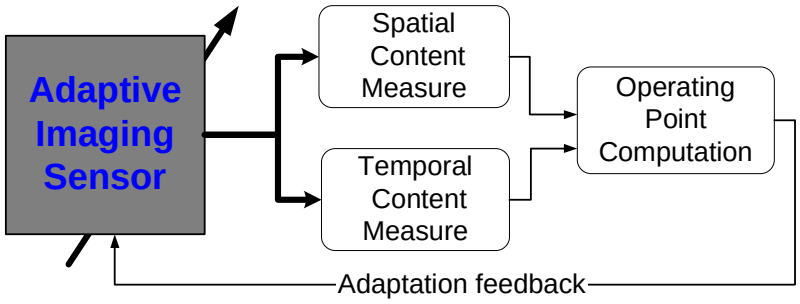


Figure 2. An adaptive sensor block diagram

Another important aspect of a sensor's capability is the data transmission bandwidth at its output. A fixed sampling rate of the sensor determines a fixed bit-rate at the output assuming no compression involved. But in many cases such as static scenes or scenes with very little details the said bandwidth is not utilized efficiently. In the adaptive framework, the sensor determines the required sampling rate and can therefore either reduce the bit-rate to the minimum necessary, or use the available bandwidth to increase the spatial sampling rate at the expense of the temporal sampling rate and vice versa.

The video sequence at the output of an adaptive sensor is a set of frames at varying spatial and temporal sampling rates. This three dimensional data cube represents the scene as was sampled by the sensor after adaptation. If the sensor sampled the scene as dictated by the content measure, this cube of data, in the ideal case, should include sufficient information to restore a high resolution spatio-temporal sequence. In some cases, the sensor operating space will limit the sampling rate and a bias towards the temporal or the spatial sampling rate has to be introduced. In both cases of ideally sampled and under sampled scene, a restoration is possible using methods such as space-time super-resolution[1, 13] and motion compensated interpolation[14] as will be further discussed in Section 5.

The use of imaging sensors at different operating points is the basis of other related work. Ben-Ezra and Nayar[2] have used a hybrid sensor configuration to remove motion blur from still images. In their approach, one sensor works in a high temporal, low spatial sampling rate operating point to capture the motion during image integration time. A second sensor acquires the image in high spatial sampling rate and uses the motion information from the first sensor to deblur the image. Lim[3] has employed very high temporal sampling at the expense of spatial sampling to restore a high resolution sequence. Other related work are from the voice recognition field.[5, 6] Here, the use of variable frame rate (VFR) is applied to speech analysis. The frame rate is determined by a content entropy measure on the recorded audio signal.

2. VIDEO CONTENT MEASURE

The purpose of the content measure is to quantify the spatial and temporal information in the scene. By spatial information we mean details or spatial frequency content. Temporal content information is the change along the time axis or temporal frequency content. Accurate measurement of such detail will allow us to determine the required spatial sampling rate and to adjust the imaging sensor accordingly. Traditional methods use frequency domain analysis to measure the frequency content of the image sequence. The frequency domain measure becomes unreliable with the existence of noise. Entropy measures as used in the speech recognition application[6] are computationally inexpensive but seem not to be robust in terms of accuracy for video data.

Other methods[11] are based on Shannon's information theory and provide metrics for quality assessment and visualization.

In the adaptive sensor framework,the objective is to keep the computational requirements minimal so that simple and cost effective implementation is possible. The chosen quantitative measure needs to be robust and accurate in the presence of noise and aliasing. In this section we first present the frequency domain and entropy methods for content measure and their characteristics with noise and aliasing. Noting their shortcomings, we then suggest a content measure in the spatial domain that is computationally inexpensive and can work robustly in the present of noise and aliasing. The content measure is first presented for the spatial case. An extension of the suggested measure for the temporal case is described towards the end of the section.

2.1. Frequency Domain and Entropy Methods

Measuring the spatial content in the frequency domain is naturally translated to a two-dimensional fast Fourier transform (FFT) of the image. The image content is determined to be at the frequency where, say, 99% of the total energy under the spectrum is captured as depicted in Figure 3 for a one dimensional signal. We can define

$$\mathbf{F} = \mathrm{FFT}_{2D}(\mathbf{X}) \tag{1}$$

where $\mathbf{X}$ is a matrix with N pixels presenting the luminance values of the image and $\mathrm{FFT}_{2D}(\cdot)$ is a two dimensional matrix that represents the energy level of the image in the frequency domain. The content measure finds the frequency where most of the total energy in the FFT_{2D} matrix has been captured. Assuming the center of the matrix $\mathbf{F}$ is the DC bin and it has N elements, the content figure is the index such that $\gamma \sum_{i=0}^{N} |\mathbf{F}|_i^2$ has been integrated from the center pixel out. γ is a number close to 1 that determines the point where the frequency energy has significantly dropped.

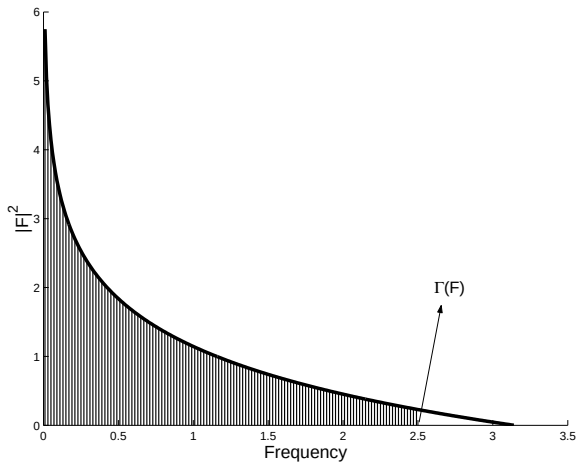


Figure 3. $\Gamma(\mathbf{F})$ operator for a one dimensional signal

In the absence of noise, this measure gives an accurate figure for the content in the image. However, the two-dimensional FFT operation is sensitive to noise. We synthesized a sequence of images for the evaluation of the content measure with respect to spatial bandwidth and noise. The sequence was composed of spatial zoneplate images (Figure 4) with frequency content from DC up to a certain known value. The sequence is composed such that the frequency bandwidth of a consecutive zoneplates in the sequence is linearly increasing and all images were sampled above their respective Nyquist rate. Figure 4 is an example of four zoneplate images from the simulation with different frequency content.

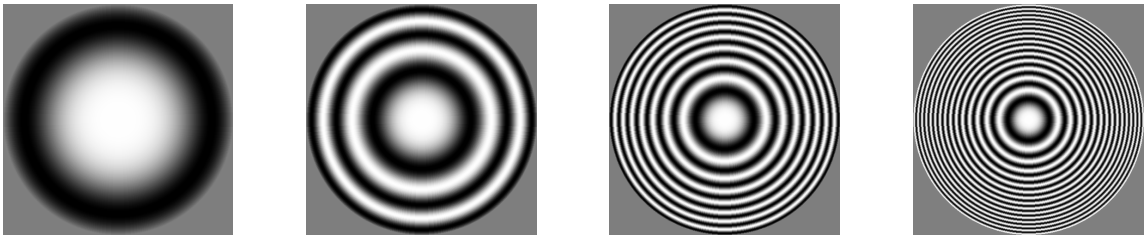


Figure 4. Zoneplate images used for the evaluation of the content measure

Figure 5 is the simulation results of the frequency domain content measure on the synthesized sequence. The solid line is the content measure of the clean images and it strongly corresponds to the linearly increasing bandwidth of the sequence. The dashed line is the content measure for the same sequence with added white Gaussian noise (WGN) with standard deviation of 20. It is clear that the noise distorts the content measure in a non-linear way such that it does not reflect the image content correctly and makes compensation rather difficult.

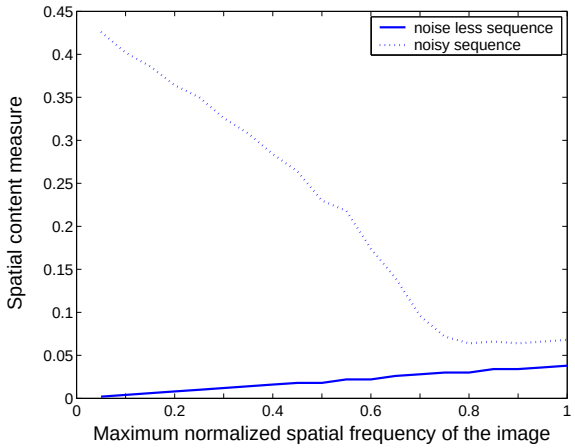


Figure 5. Frequency domain spatial content measure

Entropy methods for determining signal properties have a wide variety of forms. The entropy of a random variable is defined in terms of its probability density and can be shown to be a good measure of randomness or uncertainty. Several authors have used Shannon's entropy[7,8] and threshold-based entropy to measure the spatial content of an image. Simulation shows that entropy measure can produce results that are correlated to the image content with higher robustness to noise than then the frequency domain measure. However, the measure is not robust, nor generally useful as it is computed from the entire ensemble of pixels in the considered image without reference to the relative position of the neighboring gray values. That is, if the pixel gray values at various (or all) positions in a given image are randomly swapped with values at other pixel positions, the very same entropy measure still results. Therefore, it is impossible to relate the scalar output of the measure to the actual content.

2.2. Proposed Measure of Content

For a natural image it has been experimentally shown that the differences between adjacent pixel values mostly follow the Laplacian probability density law. Besides we can reasonably assume that in practice these differences are independent from each other.[9] By employing this significant observation, we suggest a

methodology to measure the spatial and temporal content. In the proposed framework, obtaining a figure for the content in spatial and temporal domains can be translated to a window operation using ℓ_1-norm as follows. Let $\mathbf{X}$ denote the (say raster scan) vectorized notation of the acquired image with elements $x_{i,j}$. Based on the above statistical model, we first utilize the following nonlinear ℓ_1-based filter[13, 15] applied to each pixel in the image

$$z_{i,j} = \sum_{m=-p}^{p} \sum_{l=-p}^{p} \alpha^{|m|+|l|} \left| x_{i,j} - x_{i-l,j-m} \right|, \tag{2}$$

where the weight $0 < \alpha < 1$ is applied to give a spatially decaying effect to the summation, effectively giving bigger weight to higher frequencies. $z_{i,j}$ or (in vector form $\mathbf{Z}$) is directly related to the (log-)likelihood of the image according to the assumed statistical model. To obtain a reasonably robust content measure, one can think of first finding the histogram of $\mathbf{Z}$ (call the value of this histograms $\mathbf{p}_k$ at bin $k = 0, 1, \cdots, M-1$) and then finding the value of the histogram bin (l) such that

$$\sum_{k=0}^{l} \mathbf{p}_k \geq \eta \sum_{k=0}^{M-1} \mathbf{p}_k \tag{3}$$

where η denotes the percentage of the total area under the curve we want to contribute in computing the content (for example 96%). Since computing the histogram in real-time is computationally taxing, a reasonable alternative can be employed based on the Chebyshev inequality,[10]

$$p(|\xi - \mu_\xi| \geq c\sigma_\xi) \leq \frac{1}{c^2} \tag{4}$$

where μ_ξ and σ_ξ denote the mean and variance of the random variable ξ and $p(\cdot)$ is the probability. From the Chebyshev inequality, we can determine the coefficient c based on the value of η. As an example for $\eta = 0.96$, we have $c = 5$. Next, we compute the mean and variance over the ensemble of elements of $\mathbf{Z}$ ($\mu_\mathbf{Z}$ and $\sigma^2_\mathbf{Z}$). Finally, the content measure denoted by $\rho(\mathbf{Z})$ is obtained by

$$\rho(\mathbf{Z}) = \mu_\mathbf{Z} + c\sigma_\mathbf{Z}. \tag{5}$$

The proposed ℓ_1-based operation is computationally inexpensive and proves to perform well as compared to frequency domain and entropy measures. We characterize the spatial ℓ_1-norm measure with respect to additive white Gaussian noise, measure correlation to the content bandwidth, and analyze its behavior with the presence of aliasing. Figure 6 is the simulation results of the ℓ_1-norm content measure on a synthesized sequence with a known frequency content. The sequence is the same one synthesized for the frequency domain measure in Section 2.1.

The solid line is the content measure of the synthesized sequence with no added noise. The measure behavior is monotonically increasing and strongly correlates to the linearly increasing bandwidth of the sequence. Figure 6 also shows the behavior of the ℓ_1-norm measure for added WGN with different variance (σ^2). As opposed to the frequency domain measure with added noise, the behavior of the ℓ_1-norm measure conserves the ratio of high and low content and can be compensated for rather easily, assuming σ^2 is known [§].

The compensation is done by characterizing the gap between the noisy measure and the pure measure for each σ using a polynomial fit. The polynomial is then used to remove the bias from the measure. Experiments with real video data show that this compensation method can remove the bias such that the compensated measure is consistently within 10% of the noise-less measure.

[§]Assuming readout to be the only source of noise, the value of σ^2 can be characterized offline (and hence assumed "known") for a given sensor at a particular operating point.

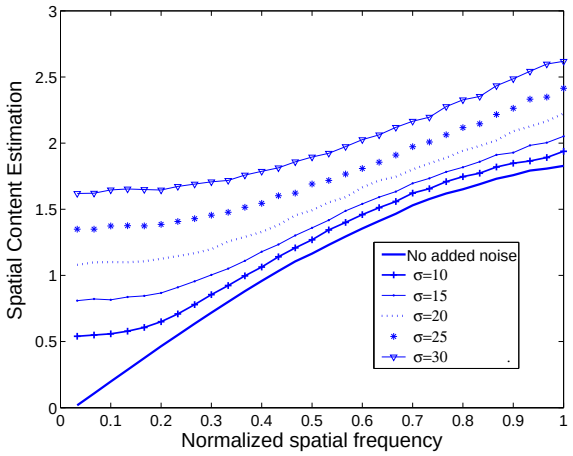


Figure 6. ℓ_1-norm spatial content measure with added WGN

As further discussed in Section 3, aliasing effect in the ℓ_1-norm measure was evaluated by down-sampling the synthesized sequence to introduce aliasing. Simulation results show that the measure saturates as soon as aliasing is introduced so that high to low content ratio is still kept. This is an important characteristic for a content measure since the adaptive sensor may run at any point in time in an operating point that introduces aliasing.

2.3. Temporal content measure using ℓ_1-norm

Measuring of the temporal content in a video sequence is the companion problem to the spatial content measure in an image. Here we quantify the temporal information in the scene. The same ℓ_1 norm method as in the spatial case can be used on a one dimensional window along the time axis in the following form:

$$q_{i,j,t} = \sum_{k=-p}^{p} \beta^{|k|} \left| x_{i,j,t} - x_{i,j,t+k} \right| \tag{6}$$

where the operation is performed on a window of duration 2p time samples and the scalar weight $0 < \beta < 1$ is applied to give a temporally decaying effect to the summation, effectively giving bigger weight to higher temporal frequencies. The content figure is given by $\rho(\mathbf{Q}_t)$ where $\mathbf{Q}_t$ is the matrix notation for $q_{i,j,t}$ (at frame t) and the same $\rho(\cdot)$ operator as defined in Section 2.2 is used to compute the temporal content measure.

Figure 7 is the simulation results of the temporal ℓ_1-norm content measure on synthesized sequences. It also shows the behavior of the measure to added WGN with different variance levels. The synthesized sequences were composed with a known temporal frequency content by changing the pixels value along the time axis using a sinusoid. Figure 8 is an example of pixels value along the time axis from four different sequences. Each sequence has different temporal content according to the sinusoid being used. The measure characteristics with respect to additive noise, correlation to the content bandwidth, and behavior with the existence of aliasing, are similar to its counterpart in the spatial domain.

3. SENSOR OPERATING POINT

The sensor operating point (SOP) is defined as {number of pixels per frame, frame rate} point in the feasible space of the sensor as depicted, for example, in Figure 1. The sensor's operating space is different from sensor to sensor and may not be smooth due to physical limitations. The required operating point (ROP) is defined as the {number of pixels per frame, frame rate} set as dictated by the scene. In other words, the ROP is the minimum required temporal and spatial sampling rates that avoid aliasing or allow for full restoration of the

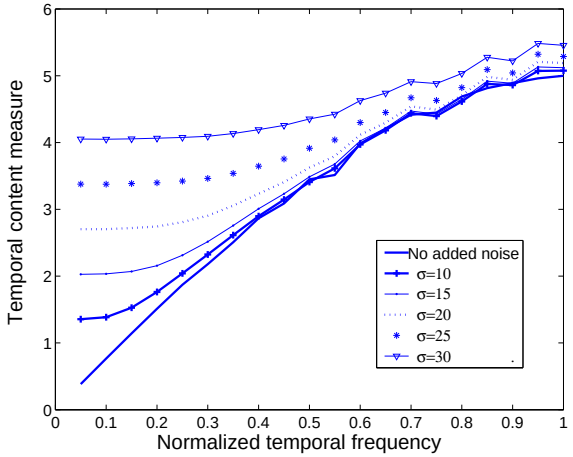


Figure 7. ℓ_1-norm temporal content measure with added WGN

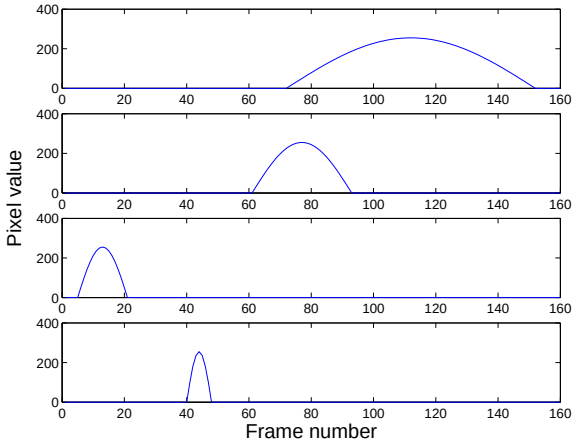


Figure 8. Pixels value along the time axis of synthesized sequences for the evaluation of the temporal content measure

video sequence by post-processing. In this framework we adapt the SOP to be as close as possible to the ROP, adapting the imaging process to the sensor's capabilities and the scene.

The ROP is derived from the spatial and temporal content measures. Ideally, the spatial content measure would be computed by a spatially high resolution sensor to get an accurate non-aliased measure. Similarly, the temporal content would be ideally measured by a high frame rate sensor such that the temporal information in the scene is measured accurately. The actual imaging would be done by a third sensor that is running at varying operating points. This three-sensor configuration may be too expensive in practical applications and requires relatively complicated optics. In practice we would like to measure the content in the scene using the same sensor that is used for imaging. This may reduce the accuracy of the content measure due to possible aliasing. Using the ℓ_1-norm content measure, the single sensor accuracy problem does not have a big affect on the closed loop operation described in Section 4.

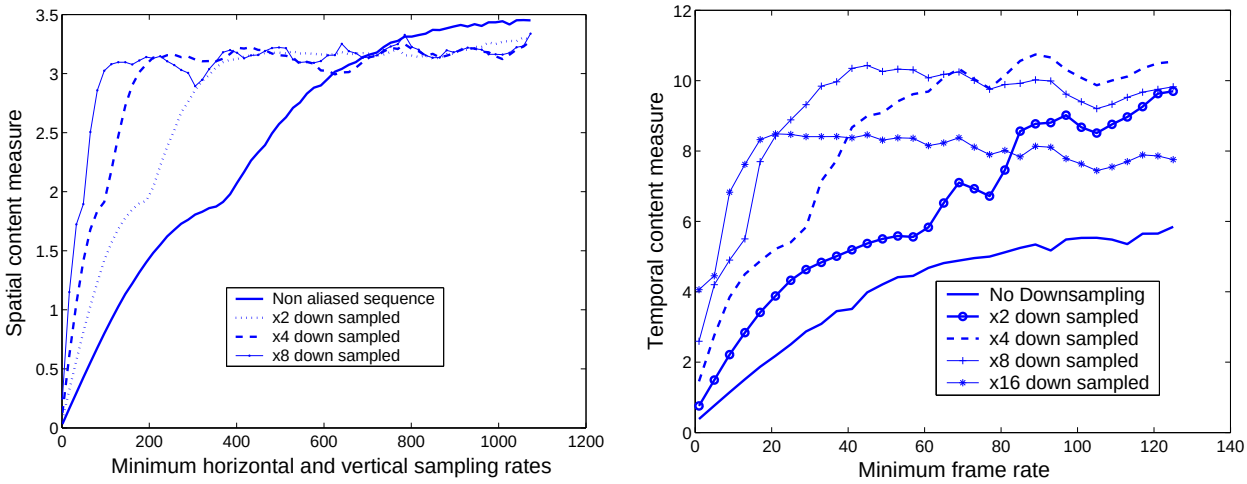
The spatial and temporal content measures produce two scalar figures for each frame of the video sequence. The content measure is the output of the ℓ_1-norm operation and does not have a direct relation to the required sampling rate (ROP). The conversion from content measure to ROP is done through the use of synthesized video sequence with known spatial and temporal bandwidth.

We show an example of the operating point computation where we take synthesized video and compute

the content measures from it. For the spatial conversion, the sequence is composed of zoneplate images as described in section 2.2 but with single frequency content for maximum accuracy. The temporal operating point computation is done through a similar concept in the time axis. The synthesized sequences for the temporal case are described in Section 2.3 and shown as example in Figure 8.

Since the characteristics of the conversion are different from one operating point to another, we use a separate conversion look-up tables for each operating point. Figure 9 is the simulation results for spatial and temporal content measures conversion to required sampling rate. The simulation for creating the conversion uses a high sampling rate non-aliased sequence as the baseline and creates lower sampling rate sequences by down sampling. The down-sampled sequences introduce aliasing as expected.

As shown in Figure 9(a), the spatial content measure for an aliased image will saturate the ℓ_1-norm operator, indicating that the current sampling rate is insufficient. This characteristic is essential for the closed loop operation of the adaptive sensor as we will describe in Section 4. The temporal conversion in Figure 9(b) has the same saturation characteristic with additional bias affect due to the global motion effect of temporal down sampling. If an imaging sensor supports continuous operating points within its operating space, then Figures 9(a) and (b) are the actual conversion to the required sampling rate. If the sensor supports only discrete points within its operating space, we can create a look-up table for the conversion by setting thresholds in the conversion curves at these discrete points.



(a) Spatial conversion

(b) Temporal conversion

Figure 9. Content measure to sampling rate conversion

4. CLOSED LOOP OPERATION

Adaptive imaging can be depicted as a control system that tracks the scene through measuring its content. As illustrated in Figure 10, the sensor's output is the system output as well as the feedback loop mechanism's input. The content measure is converted to the required operating point using look-up tables that were prepared according to the imaging sensor capabilities as described in Section 3.

The computed operating point may or may not be within the sensor's operating space. Therefore, an additional stage of projecting the required operating point to the sensor operating space is required. The projection is done in the spatial and temporal domain separately and in many cases it may not lead to a feasible point within the sensor's space. In these cases, additional input from the post processing engine or the user can impact the final sensor's operating point by balancing the point towards higher spatial or temporal sampling rate.

Finally the error in the system between the current operating point and the computed one is determined and fed back to the sensor through a feedback filter. The filter in the feedback loop is effectively smoothing the feedback response and keeps the system from diverging. In a sensor with discrete operating point, the filter can be as simple as restricting the change in the operating point to the nearest one in any direction. For a continuous operating space sensor the filter can perform a smoothing operation as follows,

$$\mathrm{SOP}(t) = \mathrm{SOP}(t-1) + \Delta\mathrm{SOP}_c(t) \tag{7}$$

where SOP_c is the SOP that was calculated at time t and Δ controls the amount of smoothing on the operating point behavior. The sensor operation starts in a middle range operating point and converges within few frames. As described in Section 3, the ROP to sampling rate conversion will saturate whenever the content is aliased for the current operating point. This important characteristic of the conversion will ensure the convergence of the system since the saturated value will drive the sensor to a higher sampling rate until no aliasing occur and the content measure is accurate again or until the sensor has reached its limits.

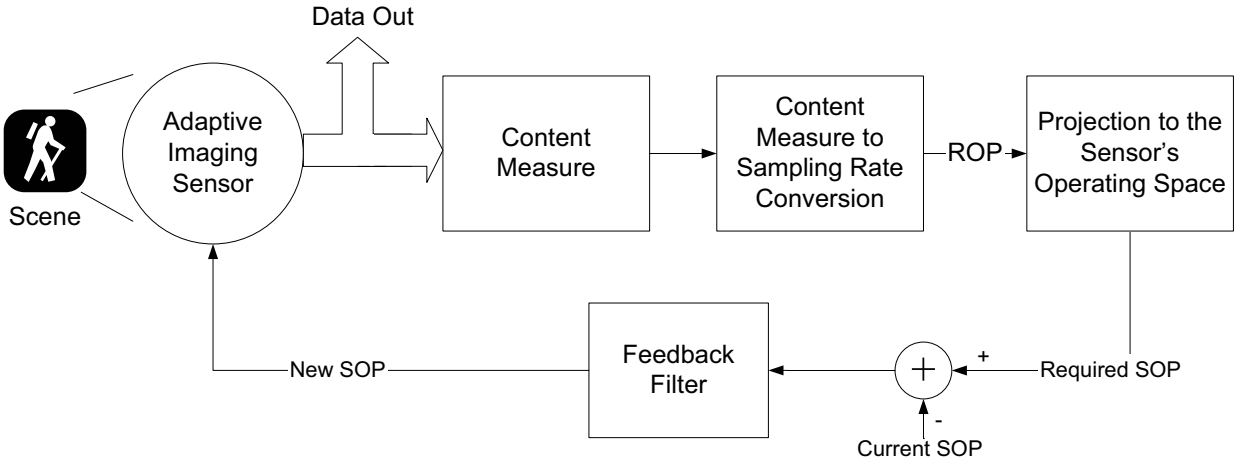
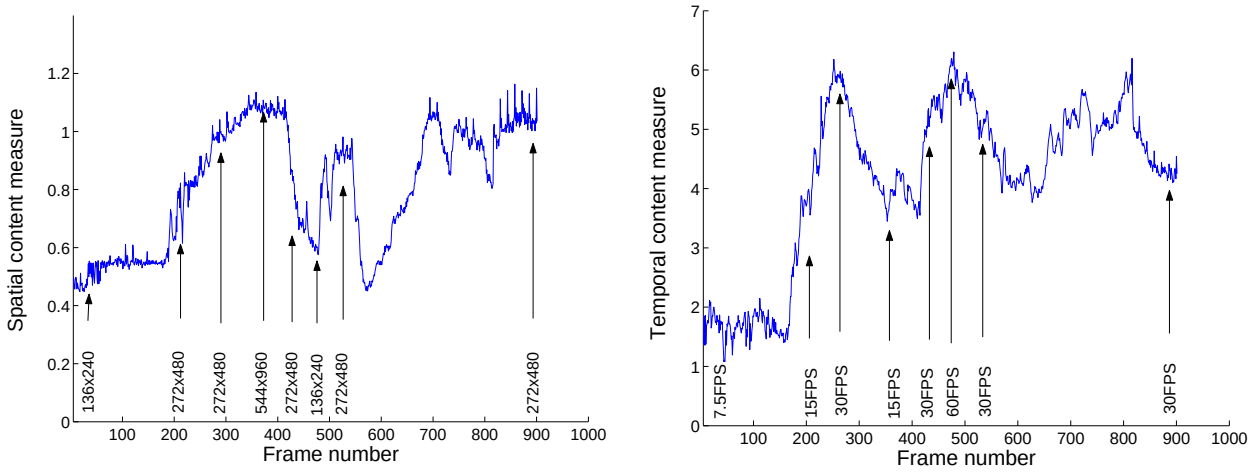


Figure 10. Closed loop operation

The complete closed loop system was simulated using real sequences from a high definition video source. The high definition video was composed of 1920×1088 pixels images at 60 frames per second frame rate. By down-sampling this sequence spatially and temporally, we created several discrete operating points. The conversion from ROP to sampling rate has been tabulated using thresholds as described in Section 3. The system output was evaluated through a simple display mechanism where images were spatially scaled and temporally repeated to create a high spatio-temporal sampling rate sequence. The simulation output results in a new sequence that has significantly reduced data bandwidth at certain points. The bandwidth reduction can be as significant as 50% of the original sequence for static scenes or scenes with low spatial bandwidth. The operating point dynamics show high correspondence to the image bandwidth as measured in the frequency domain. Figures 11(a) and (b) are the spatial and temporal content measures along 900 frames of a high definition video sequence. We marked the major operating point transitions along the curves and show the corresponding images in Figure 12. The sequence is a football match that starts from almost a static scene with relatively low content. The operating point converges at that point to the lowest spatio-temporal sampling rate (first image). The next operating point transition to a higher sampling rate happens when the players

move to their position (second and third images). Once the players are in place, the scene is less active and the camera zooms out getting more spatial details to the scene. At that point, the operating point has higher spatial and lower temporal sampling rates (forth image). When the play starts, the temporal sampling rate increases rapidly (fifth and sixth images) and settles back down for the rest of the sequence (seventh and eighth images).



(a) Spatial content measure

(b) Temporal content measure

Figure 11. Spatial and temporal content measures of the football sequence

Figure 12. Input images from the football sequence at the operating point transitions. The numbers below each picture indicate the computed operating points of the closed loop operation.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a novel approach for image and video sensing. Imaging is done with adaptation to the spatial and temporal content of the scene, optimizing the sensor's sampling rate and the camera transmission bandwidth. We developed a spatial and temporal content measure based on an ℓ_1-norm and characterized it with respect to noise, image bandwidth, and aliasing. A complete closed-loop system has been

simulated using natural scenes and the results show high correspondence to the scene dynamics and significant reduction in the camera output bit-rate.

The output of an adaptive sensor is a sequence of images with varying spatial and temporal sampling rates. This data stream captures the scene more efficiently and with fewer artifacts such that in a post-processing step an enhanced resolution sequence can be composed or lower bandwidth can be used. The non-standard stream requires a non-traditional mechanism to address the change in sampling rate. The post processing step can be part of future work in this framework for adaptive imaging. Well established video processing methods such as super-resolution[13] and motion compensated interpolation[14] are very appropriate for restoring a spatio-temporal high resolution sequence from adaptively captured date.

Acknowlegement: This work was supported in part by NSF CAREER Grant CCR-9984246 and AFOSR grant F49620-03-1-0387.

REFERENCES

1. E. Shechtman, Y. Caspi, and M. Irani, Increasing Space-Time Resolution in Video, Proc. Seventh European Conf. Computer Vision, vol. 1, p. 753, 2002.
2. M. Ben-Ezra, S. K. Nayar, Motion-Based Motion Deblurring, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 6, June 2004
3. S. H. Lim, Video Processing Applications of High Speed CMOS Sensors, PhD dissertation Stanford University, EE Department, March 2003
4. T. Chen, Digital Camera System Simulator and Applications, PhD dissertation Stanford University, EE Department, June 2003
5. H. You, Q. Zhu, A. Alwan, Entropy-Based Variable Frame Rate Analysis Of Speech Signals And Its Application To ASR, ICASSP, Montreal, Canada, May. 2004
6. Q. Zhu, A. Alwan, On the use of variable frame rate analysis in speech recognition, ICASSP, pp. 3264-3267, 2000.
7. C.E.Shannon, A Mathematical Theory of Communication, Bell Syst. Tech. J., 27, 379-423, 623-656.,1948
8. S. Kullback, Information Theory and Statistics, Dover Publications, Inc, 1968.
9. M. Green, Statistics of Images, the TV Algorithm of Rudin-Osher-Fatemi for Image Denoising and an Improved Denoising Algorithm, UCLA CAM Report 02-55, Oct. 2002.
10. A. Papoulis, S. Unnikrishna Pillai, Probability, Random Variables and Stochastic Processes, McGraw-Hill, 2001
11. J. Yang-Peláez, W. C. Flowers, Information Content Measures of Visual Displays, Proceedings of the IEEE Symposium on Information Vizualization 2000
12. H.R. Sheikh, A.C. Bovik, Image Information And Visual Quality, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3 pp. 709-12, May 2004
13. S. Farsiu, D. Robinson, M. Elad, P. Milanfar, Fast and Robust Multi-Frame Super-Resolution, IEEE Trans. Image Processing vol. 13 pp. 1327-1344, Oct. 2004
14. A. M. Tekalp, Digital Video Processing. Prentice-Hall, 1995.
15. M. Elad, On the bilateral filter and ways to improve it, IEEE Trans. Image Processing, vol. 11, no. 10, pp. 1141-1151, Oct. 2002.