A Saliency-Driven LCD Power Management System

A B, A B, A B, A B, and A B

I. Introduction

SINCE the invention of the first television, the display system has manifested into a critical human-computer-interface capable of bringing vivid and abundant visual information to users. As technology evolved, flat panel LCD display systems were invented and are now widely used in various multimedia systems, ranging from large screen home theatre systems to personal laptops and mobile devices. As projected in [1], LCD TVs are becoming progressively bigger in size every year, with customers demanding powerful visual experiences as video technology evolves. As the smartphone industry explores newer dimensions of innovation, there is an increasing demand for larger screens in this multimedia category too [2].

The LCD panel in a common LCD display system will not emit light by itself. Therefore, a backlight panel and a light-diffusing component are used under the LCD panel as a source of lighting in the system. Currently, there are two kinds of lighting sources employed for the backlight panel. An older method uses cold cathode fluorescent lamps (CCFL); while a more recent one utilizes light-emitting diodes (LEDs) since LEDs can offer greater dynamic contrast, wider color gamut and less power consumption. Although this new backlight technology can provide better power efficiency, it needs to be pointed out that the backlight panel still remains the largest portion of the entire system power consumption [3]. With the increasing demand for larger screens, optimizing and enhancing the power efficiency of the LCD display system, especially televisions [4], is thus attracting continuing research efforts.

While approximate computing is becoming a powerful paradigm to save energy in the vision space [**?**], another promising way to solve the power problem specific to LCD panels is to dim the LED backlight. Many different methods have been proposed in the same vein. Active dimming approaches adjust the luminance level of the backlight based on pixel information such as contrast, color, or brightness. Passive dimming methods modify the luminance level by monitoring user attention with a camera or sensors. The backlight is dimmed when users are away and restored to full level when users are in front of the display.

In [5], an active dimming strategy is proposed, where visual saliency is used to adaptively change the luminance level of the backlight panel at a zone granularity. Most objects of interest have distinct features that stand out in contrast to their background. [5] used three feature channels - *intensity*, *color* and *orientation* - for locating salient or interesting objects/regions in the scene. These low-level features, when applied to images, perform extremely well, as was demonstrated in [5]. However, due to the static nature of these channels, the system, when operating on a video stream, is constrained in that it fails to behave like a person who can focus attention on new objects entering the frame, or moving objects across a series of frames. To overcome these handicaps, we extend [5] with the following contributions:

- LCD-based devices these days invariably have streaming video applications running on them and consume more power than when a static image is being observed. As highlighted in [6], two additional processing channels, *motion* and *flicker* are effective in finding salient regions in a temporal environment. We thus incorporate these two channels into [5] to account for dynamic occurrences across a stream of video frames.
- We then highlight the impact of these changes on user experience. Since the original system settings in [5] introduce a *shimmer* (discussed later) in the final compensated video, new dimming and compensation coefficients are used. A movement constraint for salient zones between two continuous frames is also adopted for preserving video quality. Using these new compensation coefficients, verification of our system is undertaken with 29 different videos as part of a user perception evaluation test.
- To address system constraints such as power, performance and resource tradeoffs, this paper adopts a generic field-programmable gate array (FPGA) channel architecture. On average, a **50%** power saving can be achieved with our extended system while maintaining real-time constraints.

The organization of the rest of this paper is as follows: Section **??** highlights related work on bio-inspired systems in general and various LED power saving schemes in specific; Section **??** introduces Saliency, the adopted biological attention model, in detail; the LED backlight system and its power model is outlined in Section **??**; Section **??** shows the design and implementation details of the proposed power management system; all the experimental results are provided in Section **??**; finally we conclude the paper in Section V.

II. Beyond Deep Learning

Convolutional Neural Networks (CNNs) have become extremely popular and being used to solve a variety of image recognition problems. More recent and advanced CNN architectures have become deeper and more complex haing 10 to 20 layers of Rectified Linear Units, hundreds of millions of weights, and billions of connections between units. The reader is pointed to [7] for insights on deep architectures in general and [8] for CNN-based learning and their recent advances.

While CNNs are an important thrust of research, they tend to be computationally expensive and deploying them on mobile

Authors are with the School of Electrical Engineering and Computer Science, Pennsylvania State University, PA, USA (email: {vijay}@cse.psu.edu)
Manuscript received September 15, 2016; revised Month Date, 2016; accepted Month Date, 2015.

platforms results in huge memory overheads. Another important insight recently unearthed by [9], is that the accuracies of CNNs can saturate after a few million images of training data. Also the overall efficacy of the image recognition pipeline is contingent upon having a good region proposal scheme that feeds regions of interest (RoIs) into the CNN. A variety of strategies can be used to augment the capabilities of neural networks and we discuss a few below.

A. Visual Attention

Humans process only certain streams of visual information depending upon the task at hand. From a systems perspective, visual attention can be used as an efficient mechanism to prioritize visual processing. Pixel-level saliency models such as Attention by Information Maximiztion (AIM) [10] can be used in automatic household pantry organization and maintenance, particularly as part of visual assist systems for the visually impaired. Computationally, AIM determines visual salience based on the amount of information present in local regions of the image within the context of its surrounding region. Suppose a product (say cookies) is wrongly placed in a shelf that stores products of another type (say shampoo), the segment of the shelf image containing cookies is "less likely" (higher self-information) to appear in the scene which mostly has image patches of shampoo, and therefore it is easily distinguishable or is considered "salient". This is shown in Figure 1. Once a salient region is detected, a second stage of object classification can be deployed to identify the wrongly placed object.

(a) Original image (b) Thresholded saliency map

Fig. 1. Saliency used for missplaced item detection.

B. Context

While deep learning models use learnt features to recognize objects in a scene, another contrasting approach is to use graphical models that build hierarchical representations of objects [11]. Compositional rules can be used to build context cues to recognize objects never seen before. For example, an object having four wheels can be classified as a vehicle even if it is a new model of a car. While CNNs have been widely used for image category recognition, when it comes to recognizing objects in video streams, spatial context can play a huge role in reducing the workload on these computationally intensive classifiers. As shown in Figure 2, visual scenes can be represented as a knowledge graph. For example, in [12], the authors proposed a Bayesian network called Visual Co-occurrence Network (ViCoNet) to not only improve the performance of their system, but also increase the recognition rates.

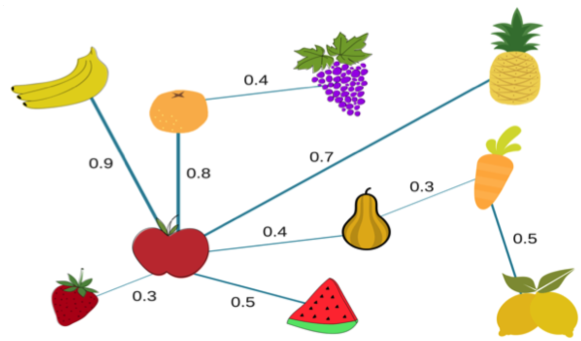


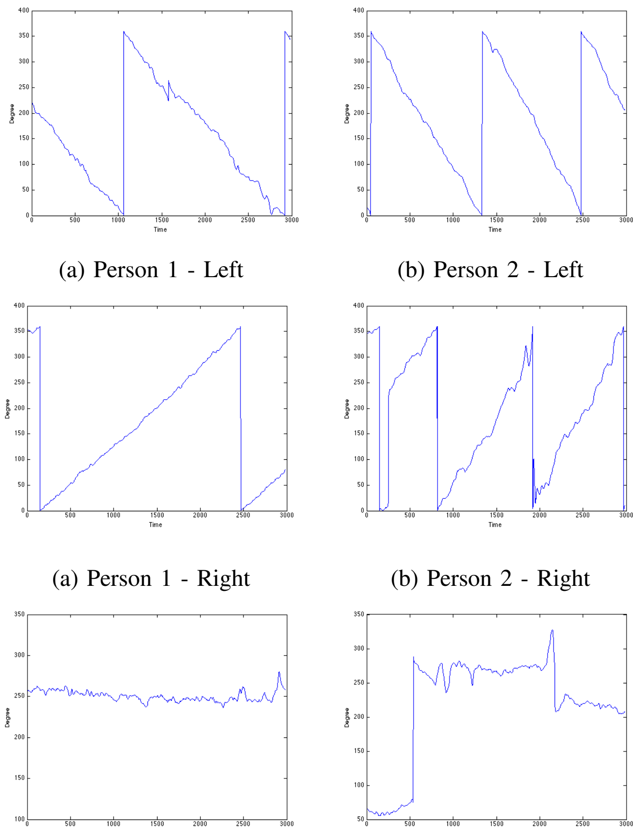
Fig. 2. Spatial relations exist between frequently co-occurring objects. These relationships can then be used as context cues to guide the classifiers.

C. Multimodal Fusion

Humans use multisensory information from different sensory systems and combine it to influence perception, decisions, and overt behavior [13]. Wearables can be used in a similar fashion to help users in different tasks. A rich topic of exploration is figuring out a way to fuse multi-sensor information, especially data from vision that is fundamentally two-dimensional with a temporal unidimensional stream of data from other sensors to make predictions of the current state of the user. Multi-sensor information coming in from different devices can be streamed to distributed networks that can then make real-time updates. In Figure 3, we illustrate data recorded from a wearable device while two users walk in three different directions. These sensors are sensitive enough to be used as localization cues.

III. Interfaces

Since the invention of the first television, the display system has manifested into a critical human-computer-interface capable of bringing vivid and abundant visual information to users. As technology evolved, flat panel LCD display systems were invented and are now widely used in various multimedia systems, ranging from large screen home theatre systems to personal laptops and mobile devices. As projected in [1], LCD TVs are becoming progressively bigger in size every year, with customers demanding powerful visual experiences as video technology evolves. As the smartphone industry explores newer dimensions of innovation, there is an increasing demand for larger screens in this multimedia category too [2].



(a) Person 1 - Left (b) Person 2 - Left

(a) Person 1 - Right (b) Person 2 - Right

(a) Person 1 - Straight (b) Person 2 - Straight

Fig. 3. Other sensor information used for localization.

The LCD panel in a common LCD display system will not emit light by itself. Therefore, a backlight panel and a light-diffusing component are used under the LCD panel as a source of lighting in the system. Currently, there are two kinds of lighting sources employed for the backlight panel. An older method uses cold cathode fluorescent lamps (CCFL); while a more recent one utilizes light-emitting diodes (LEDs) since LEDs can offer greater dynamic contrast, wider color gamut and less power consumption. Although this new backlight technology can provide better power efficiency, it needs to be pointed out that the backlight panel still remains the largest portion of the entire system power consumption [3]. With the increasing demand for larger screens, optimizing and enhancing the power efficiency of the LCD display system, especially televisions [4], is thus attracting continuing research efforts.

While approximate computing is becoming a powerful paradigm to save energy in the vision space [**?**], another promising way to solve the power problem specific to LCD panels is to dim the LED backlight. Many different methods have been proposed in the same vein. Active dimming approaches adjust the luminance level of the backlight based on pixel information such as contrast, color, or brightness. Passive dimming methods modify the luminance level by monitoring user attention with a camera or sensors. The backlight is dimmed when users are away and restored to full level when users are in front of the display.

In [5], an active dimming strategy is proposed, where visual saliency is used to adaptively change the luminance level of the backlight panel at a zone granularity. Most objects of interest have distinct features that stand out in contrast to their background. [5] used three feature channels - *intensity*, *color* and *orientation* - for locating salient or interesting objects/regions in the scene. These low-level features, when applied to images, perform extremely well, as was demonstrated in [5]. However, due to the static nature of these channels, the system, when operating on a video stream, is constrained in that it fails to behave like a person who can focus attention on new objects entering the frame, or moving objects across a series of frames. To overcome these handicaps, we extend [5] with the following contributions:

- LCD-based devices these days invariably have streaming video applications running on them and consume more power than when a static image is being observed. As highlighted in [6], two additional processing channels, *motion* and *flicker* are effective in finding salient regions in a temporal environment. We thus incorporate these two channels into [5] to account for dynamic occurrences across a stream of video frames.
- We then highlight the impact of these changes on user experience. Since the original system settings in [5] introduce a *shimmer* (discussed later) in the final compensated video, new dimming and compensation coefficients are used. A movement constraint for salient zones between two continuous frames is also adopted for preserving video quality. Using these new compensation coefficients, verification of our system is undertaken with 29 different videos as part of a user perception evaluation test.
- To address system constraints such as power, performance and resource tradeoffs, this paper adopts a generic field-programmable gate array (FPGA) channel architecture. On average, a **50%** power saving can be achieved with our extended system while maintaining real-time constraints.

The organization of the rest of this paper is as follows: Section **??** highlights related work on bio-inspired systems in general and various LED power saving schemes in specific; Section **??** introduces Saliency, the adopted biological attention model, in detail; the LED backlight system and its power model is outlined in Section **??**; Section **??** shows the design and implementation details of the proposed power management system; all the experimental results are provided in Section **??**; finally we conclude the paper in Section V.

IV. Hardware

Since the invention of the first television, the display system has manifested into a critical human-computer-interface capable of bringing vivid and abundant visual information to users. As technology evolved, flat panel LCD display systems were invented and are now widely used in various multimedia systems, ranging from large screen home theatre systems to personal laptops and mobile devices. As projected in [1], LCD TVs are becoming progressively bigger in size every year, with customers demanding powerful visual experiences as video technology evolves. As the smartphone industry explores newer dimensions of innovation, there is an increasing demand for larger screens in this multimedia category too [2].

The LCD panel in a common LCD display system will not emit light by itself. Therefore, a backlight panel and a light-diffusing component are used under the LCD panel as a source of lighting in the system. Currently, there are two kinds of lighting sources employed for the backlight panel. An older method uses cold cathode fluorescent lamps (CCFL); while a more recent one utilizes light-emitting diodes (LEDs) since LEDs can offer greater dynamic contrast, wider color gamut and less power consumption. Although this new backlight technology can provide better power efficiency, it needs to be pointed out that the backlight panel still remains the largest portion of the entire system power consumption [3]. With the increasing demand for larger screens, optimizing and enhancing the power efficiency of the LCD display system, especially televisions [4], is thus attracting continuing research efforts.

While approximate computing is becoming a powerful paradigm to save energy in the vision space [**?**], another promising way to solve the power problem specific to LCD panels is to dim the LED backlight. Many different methods have been proposed in the same vein. Active dimming approaches adjust the luminance level of the backlight based on pixel information such as contrast, color, or brightness. Passive dimming methods modify the luminance level by monitoring user attention with a camera or sensors. The backlight is dimmed when users are away and restored to full level when users are in front of the display.

In [5], an active dimming strategy is proposed, where visual saliency is used to adaptively change the luminance level of the backlight panel at a zone granularity. Most objects of interest have distinct features that stand out in contrast to their background. [5] used three feature channels - *intensity*, *color* and *orientation* - for locating salient or interesting objects/regions in the scene. These low-level features, when applied to images, perform extremely well, as was demonstrated in [5]. However, due to the static nature of these channels, the system, when operating on a video stream, is constrained in that it fails to behave like a person who can focus attention on new objects entering the frame, or moving objects across a series of frames. To overcome these handicaps, we extend [5] with the following contributions:

- LCD-based devices these days invariably have streaming video applications running on them and consume more power than when a static image is being observed. As highlighted in [6], two additional processing channels, *motion* and *flicker* are effective in finding salient regions in a temporal environment. We thus incorporate these two channels into [5] to account for dynamic occurrences across a stream of video frames.
- We then highlight the impact of these changes on user experience. Since the original system settings in [5] introduce a *shimmer* (discussed later) in the final compensated video, new dimming and compensation coefficients are used. A movement constraint for salient zones between two continuous frames is also adopted for preserving video quality. Using these new compensation coefficients, verification of our system is undertaken with 29 different videos as part of a user perception evaluation test.
- To address system constraints such as power, performance and resource tradeoffs, this paper adopts a generic field-programmable gate array (FPGA) channel architecture. On average, a **50%** power saving can be achieved with our extended system while maintaining real-time constraints.

The organization of the rest of this paper is as follows: Section **??** highlights related work on bio-inspired systems in general and various LED power saving schemes in specific; Section **??** introduces Saliency, the adopted biological attention model, in detail; the LED backlight system and its power model is outlined in Section **??**; Section **??** shows the design and implementation details of the proposed power management system; all the experimental results are provided in Section **??**; finally we conclude the paper in Section V.

V. Conclusion

This work highlighted the efficacy of personal visual assist systems in our day to day activities. As technological advances spur growth, more and more consumer products will become available.

VI.

Acknowledgment

This work is supported in part by NSF Expeditions: Visual Cortex on Silicon CCF 1317560.

References

[1] Y. Ishii, "The World of Liquid-Crystal Display TVs - Past, Present, and Future," *Journal of Display Technology*, vol. 3, no. 4, pp. 351–360, Dec 2007.
[2] iPhone 6s/6s Plus. (Date last accessed 29-Jan-2016). [Online]. Available: http://www.apple.com/iphone-6s/specs/
[3] S. Sinofsky, 2009 Windows 7 Energy Efficiency.
[4] Samsung. (Date last accessed 29-Jan-2016). [Online]. Available: http://www.samsung.com/us/video/tvs/
[5] Y. Xiao, K. Irick, V. Narayanan, D. Shin, and N. Chang, "Saliency Aware Display Power Management," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*, March 2013, pp. 1203–1208.
[6] R. Peters and L. Itti, "Applying computational tools to predict gaze direction in interactive visual environments," *ACM Trans. on Applied Perception*, 2007.
[7] Y. Bengio, "Learning Deep Architectures for AI." Now Publishers, 2009.
[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," pp. 436–444, May 2015.
[9] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CoRR*, vol. abs/1503.03832, 2015. [Online]. Available: http://arxiv.org/abs/1503.03832
[10] N. Bruce, "An Information Theoretic Model of Saliency and Visual Search."
[11] I. Kokkinos and A. Yuille, "HOP: Hierarchical Object Parsing," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 802–809.
[12] S. Advani, B. Smith, Y. Tanabe, K. Irick, M. Cotter, J. Sampson, and V. Narayanan, "Visual co-occurrence network: Using context for large-scale object recognition in retail," in *Embedded Systems For Real-time Multimedia (ESTIMedia), 2015 13th IEEE Symposium on*, Oct 2015, pp. 1–10.
[13] B. E. Stein, T. R. Stanford, and B. A. Rowland, "The Neural Basis of Multisensory Integration in the Midbrain: its Organization and Maturation," *Hearing Research*, vol. 258, no. 1, pp. 4–15, 2009.

Peter Zientara received the B.E. degree in Electrical Engineering from the Huozhong University, China in 2005, the M.S. degree in Electrical Engineering from the North Carolina State University in 2008 and the Ph.D. degree in Computer Science and Engineering from the Pennsylvania State University in 2014. He is currently a Senior R&D Engineer at Synopsys where he joined in 2014. He is currently involved in physical optimization related research.

Siddharth Advani received the B.S. degree in Electronics Engineering from the Pune University, India in 2005 and the M.S. degree in Electrical Engineering from the Pennsylvania State University in 2009. He is currently pursuing his Ph.D. in Computer Science and Engineering at the Pennsylvania State University. His research interests include embedded vision design for smart wearables and reconfigurable computing for domain-specific applications.

John (Jack) Sampson received the B.S. degree in Electrical Engineering and Computer Science from the University of California, Berkeley in 2002 and the Ph.D. in Computer Engineering from the University of California, San Diego in 2010. He is an Assistant Professor in the Department of Computer Science and Engineering at the Pennsylvania State University. His research interests include energy-efficient computing, architectural adaptations to exploit emerging technologies, and mitigating the impact of Dark Silicon.

Vijaykrishnan Narayanan received the B.S. degree in Computer Science and Engineering from the University of Madras, India, in 1993 and the Ph.D. in Computer Science and Engineering from the University of South Florida, Tampa, USA, in 1998. He is a Professor of Computer Science & Engineering and Electrical Engineering at the Pennsylvania State University. His research interests include power-aware and reliable systems, embedded systems, nanoscale devices and interactions with system architectures, reconfigurable systems, computer architectures, network-on-chips, domain-specific computing.

Dr. Narayanan has received several awards including the Penn State Engineering Society Outstanding Research Award in 2006, IEEE CAS VLSI Transactions Best Paper Award in 2002, the Penn State CSE Faculty Teaching Award in 2002, the ACM SIGDA outstanding new faculty award in 2000, Upsilon Pi Epsilon award for academic excellence in 1997, the IEEE Computer Society Richard E. Merwin Award in 1996 and the University of Madras first rank in Computer Science and Engineering in 1993. He is currently the editor-in-chief of IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems. He has received several certificates of appreciation for outstanding service from ACM and IEEE Computer Society. He is a Fellow member of IEEE and ACM.