INCREMENTAL SALIENT POINT DETECTION

Ioannis Patras and Yiannis Andreopoulos

Queen Mary, University of London, UK

ABSTRACT

In this paper, we investigate an approach that computes salient points, i.e. areas of natural images that contain corners or edges, *incrementally*. We focus on the popular Harris corner detector and demonstrate how such an approach can operate when the image samples are refined in a bitwise manner, i.e. the image bitplanes are received one-by-one from the image sensor. This has the advantage that the image sensing and the salient point detection can be terminated at any input image precision (e.g. at a bound set by the sensory equipment or by computation, or by the salient point accuracy required by the application) and the obtained salient points under this precision are readily available. We estimate the required energy for image sensing as well as the computation required for the salient point detection and compare them against the conventional salient point detector realization that operates directly on each source precision and cannot refine the result. Our experiments demonstrate the feasibility of incremental approaches for salient point detection in various classes of natural images. In addition, a first comparison between the results obtained by the intermediate detectors is presented.

Index Terms— *salient point detection, incremental refinement of computation, low-level feature detection,*

1. INTRODUCTION

Low-level feature detectors attempt to isolate image areas that contain visually important data, such as edges or corners. Many such approaches have been proposed within the last 30 years [1]-[3] and they are generally termed as detectors of points of interest or salient point detectors (SPDs). Some of the first works in the area (e.g. Moravec's work) were motivated by the fact that low-level real-time image analysis is extremely useful for various applications in robotics [1], real-time surveillance and monitoring [9], etc. Computational and sensing energy requirements are very important for these applications. Today, even though computing systems have evolved to the extend where derivative-based approaches can be executed fast, the image processing needs have also increased dramatically since image resolutions are higher and there is a significant demand for processing high frame-rate videos or images and views derived from multiple cameras in systems with limited computational and energy resources, e.g. in video sensor networks. Hence, the required computational and energy resources remain an important concern. Finally, most modern realization platforms utilize dynamic task scheduling, and low-power task scheduling, e.g. via the use of dynamic voltage scaling [4]. Hence, multimedia applications need to be able to produce the best possible result under rapidly-changing system resources [4] [7].

Besides the computational aspects, new trends have also emerged in the image sensor arena. Approaches for compressed sensing [5] are assuming the use of limited sensory equipment (even up to single-pixel sensors [5]) to derive a resolution and quality-refinable approximation of the input visual data. A more straightforward and already fully-functional approach is based on CMOS image sensors capturing the input source incrementally from the most-significant bits (MSBs) to the least-significant bits (LSBs) using successive analog-to-digital conversion [6].

In this paper, we consider the detector of Harris and Stevens and extend it to support incremental derivation of salient image points with increased input image precision. This complies with the incremental CMOS-based image sensing approach proposed by Yang et al [6] that produces image bitplanes hierarchically, from the MSBs to the LSBs, (Figure 1). We reformulate the detector for incremental refinement and thereby create a hierarchical computation framework for the derivation of image salient points for each new input image bitplane. We exploit the fact that the salient points of the previous input are known when new input bitplanes are processed and localize the computation and image sensing within a window surrounding the position of each previously-found salient point. This reduces the energy and computational requirements in comparison to the straightforward approach that senses the entire image and then computes the salient points. Finally, we investigate the computational requirements for a variety of images in comparison to the conventional approach.

The remainder of the paper is as follows. Section 2 presents the proposed formulation enabling incremental refinement for the computation of image salient points. Section 3 presents the analysis of the computational and energy requirements. Section 4 presents experiments deriving the performance of the incremental versus the conventional approach and Section 5 draws conclusions.

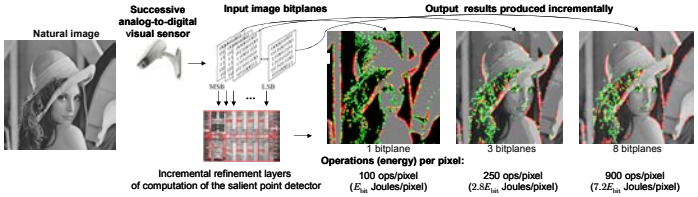

Figure 1. Overall framework for successive bitplane-based image sensing and incremental salient point detection.

2. INCREMENTAL SALIENT POINT DETECTION

The basic algorithm of Harris and Stephens [2] for image salient-point detection of an image $\mathbf{I}$ consisting of $R \times C$ pixels operates on the autocorrelation matrix of the image derivatives (X, Y):

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \qquad (1)$$

In order to suppress the noise and spurious salient points the elements of M are filtered with a Gaussian. Then, following the original paper of Harris and Stephens [2], we use the "trace matrix" and the "determinant matrix"of $\mathbf{M}$, and obtain the saliency matrix as:

$$\mathbf{R} = \det(\mathbf{M}) - k \cdot \mathrm{Tr}(\mathbf{M}) \circ \mathrm{Tr}(\mathbf{M}), \quad (2)$$

where k is a scaling factor. The second term of (2) is used to eliminate contour points with one strong eigenvalue in $\mathbf{M}$ [3]. Positive values of $\mathbf{R}$ occur in corner regions, negative values in edge regions, and small values in flat regions. Consequently, thresholding $\mathbf{R}$ followed by a non-maxima (minima) suppression results in the set of corners (edges), that is our set of salient points. Typically, the threshold Θ is set to a certain percentage θ of the maximum observed value of $\mathbf{R}$ [3].

2.1 Incremental Derivation of the Salient Point Matrix under Increased Image Sensing Accuracy

We now consider the computation of the algorithm of the previous section under the assumption that the image is sensed incrementally, i.e. from the most-significant to the least-significant bitplane. More specifically, the image is sensed in a bitplane-by-bitplane fashion, that is, each pixel of image $\mathbf{I}$ is represented by:

$$i^0_{\mathrm{sensed}}[r,c] = (-1)^{i^N_{\mathrm{sensed}}[r,c]} \sum_{n=0}^{N-1} i^n_{\mathrm{bit}}[r,c] \times 2^n \quad (3)$$

where $i^n_{\mathrm{bit}}[r,c]$ is the n th bit of sensed pixel $i^0_{\mathrm{sensed}}[r,c]$, with $i^n_{\mathrm{bit}}[r,c]$ being the LSB, $i^{N-1}_{\mathrm{bit}}[r,c]$ the MSB, and $i^N_{\mathrm{bit}}[r,c]$ reserved for the sign bit (zero for positive values, one for negative values). In the remainder of this section, we describe the steps for the incremental computation of the Harris salient point detector assuming that we are operating on any bitplane n, from the most significant bitplane $n = N-1$ to the least significant bitplane $n = 0$. Here, and for any quantity or matrix a used in the SPD algorithm, let us denote with a^n_{sensed} „ the computed value of a when the input image $\mathbf{I}$ is sensed from bitplane $N-1$ up to (and including) bitplane n. □Similarly let us denote with a^n_{bit}, the computed value of a when using only bitplane n of image $\mathbf{I}$.

Starting with the computations for the image derivatives, we notice that they can be calculated separately for each new bitplane $\mathbf{I}^n_{\mathrm{bit}}$ since the multiply-accumulate process performed during convolution is a linear operation that can be broken into separate multiply-accumulate processes for each bit that are summed afterwards. Hence, for input image bitplane $\mathbf{I}^n_{\mathrm{bit}}$ we obtain the horizontal component of the image gradient as $\mathbf{X}^n_{\mathrm{sensed}} = \left(\mathbf{I}^{n+1}_{\mathrm{sensed}} + \mathbf{I}^n_{\mathrm{bit}}\right) * \mathbf{D}$, where D is the derivative filter. This can be expressed as $\mathbf{X}^n_{\mathrm{sensed}} = \mathbf{X}^{n+1}_{\mathrm{sensed}} + \mathbf{X}^n_{\mathrm{bit}}$ with:

$$\mathbf{X}^n_{\mathrm{bit}} = \mathbf{I}^n_{\mathrm{bit}} * \mathbf{D}. \quad (4)$$

In (4) $\mathbf{X}^n_{\mathrm{bit}}$, (similarly $\mathbf{Y}^n_{\mathrm{bit}}$) contains the output of bitplane n. We remark that these matrices have slightly increased dynamic range in comparison to $\mathbf{I}^n_{\mathrm{bit}}$, i.e. they will not contain only binary values, due to the accumulation performed in the convolutions.

The next step involves filtering with the Gaussian filter. Due to the non-linearity introduced by the Hadamard products prior to the convolution operations, we break the computation as:

$$\mathbf{A}^n_{\mathrm{sensed}} = \left[\left(\mathbf{X}^{n+1}_{\mathrm{sensed}} + \mathbf{X}^n_{\mathrm{bit}}\right) \circ \left(\mathbf{X}^{n+1}_{\mathrm{sensed}} + \mathbf{X}^n_{\mathrm{bit}}\right)\right] * \mathbf{G} \quad (5)$$

which, by expanding the Hadamard product, can be written as:

$$\begin{aligned}\mathbf{A}^n_{\mathrm{sensed}} &= \left(\mathbf{X}^{n+1}_{\mathrm{sensed}} \circ \mathbf{X}^{n+1}_{\mathrm{sensed}}\right) * \mathbf{G} + \left(\mathbf{X}^n_{\mathrm{bit}} \circ \mathbf{X}^n_{\mathrm{bit}}\right) * \mathbf{G} + 2\left(\mathbf{X}^{n+1}_{\mathrm{sensed}} \circ \mathbf{X}^n_{\mathrm{bit}}\right) * \mathbf{G} \\ &= \mathbf{A}^{n+1}_{\mathrm{sensed}} + \mathbf{A}^n_{\mathrm{bit}}\end{aligned}$$

with: $\mathbf{A}^n_{\mathrm{bit}} = \left(\mathbf{X}^n_{\mathrm{bit}} \circ \mathbf{X}^n_{\mathrm{bit}}\right) * \mathbf{G} + 2\left(\mathbf{X}^{n+1}_{\mathrm{sensed}} \circ \mathbf{X}^n_{\mathrm{bit}}\right) * \mathbf{G}$.

The last equation shows that the non-linearity introduces the additional term $2\left(\mathbf{X}^{n+1}_{\mathrm{sensed}} \circ \mathbf{X}^n_{\mathrm{bit}}\right) * \mathbf{G}$ in the increment of the computation for $\mathbf{A}^n_{\mathrm{sensed}}$.(Similarly for $\mathbf{B}^n_{\mathrm{sensed}}$, and $\mathbf{C}^n_{\mathrm{sensed}}$).Then we derive the increment of the trace and the determinant as:

$$[\mathrm{Tr}(\mathbf{M})]^n_{\mathrm{bit}} = \mathbf{A}^n_{\mathrm{bit}} + \mathbf{B}^n_{\mathrm{bit}} \quad (6)$$

$$\left[\det(M)\right]^n_{bit} = \left(A^n_{bit} \circ B^n_{bit}\right) + \left(A^{n+1}_{sensed} \circ B^n_{bit}\right) + \left(B^{n+1}_{sensed} \circ A^n_{bit}\right) - \left[\left(C^n_{bit} \circ C^n_{bit}\right) + \left(C^{n+1}_{sensed} \circ C^n_{bit}\right)\right]$$

. Finally, the derivation of $\mathbf{R}^n_{\mathrm{bit}}$ [increment for (2)] is given by:

$$R^n_{bit} = \left[\det(M)\right]^n_{bit} - k\left(\left[Tr(M)\right]^n_{bit} \circ \left[Tr(M)\right]^n_{bit} + 2\left[Tr(M)\right]^{n+1}_{sensedt} \circ \left[Tr(M)\right]^n_{bit}\right) \quad (7)$$

Notice that (4)-(7) provide an incremental computation framework for the results needed in order to derive the increment $\mathbf{R}^n_{\mathrm{bit}}$ of the output salient point matrix $\mathbf{R}^n_{\mathrm{sensed}}$. The remaining step in order to complete the derivation of the salient points for the input image $\mathbf{I}^n_{\mathrm{sensed}}$ (i.e. the results from the image sensed up to, and including, bitplane n) is the addition of the increment of the results to their previous counterparts, i.e.:

$$\mathbf{Q}^n_{\mathrm{sensed}} = \mathbf{Q}^{n+1}_{\mathrm{sensed}} + \mathbf{Q}^n_{\mathrm{bit}} \quad (8)$$

with $\mathbf{Q} \in \{\mathbf{X}, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathrm{Tr}(\mathbf{M}), \det(\mathbf{M}), \mathbf{R}\}$. This process can then proceed to the final salient point selection to derive the salient points when the image is sensed up to bitplane n.

1.	*for each bitplane* n, $n = N-1, N-2, \ldots, 0$
2.	*for each point* (r,c), $0 \le r < R$, $0 \le c < C$
3.	*if* $n = N-1$ *// highest bitplane: sense all bits*
4.	*sense* $i^N_{\mathrm{bit}}[r,c]$ *(sign bit) and* $i^{N-1}_{\mathrm{bit}}[r,c]$
5.	*else if* $w_n[r,c] = 1$ *// sense only within the mask*
6.	*sense* $i^n_{\mathrm{bit}}[r,c]$
7.	*else*
8.	*set* $i^n_{\mathrm{bit}}[r,c] = 0$
9.	*calculate* $\mathbf{X}^n_{\mathrm{bit}}, \mathbf{Y}^n_{\mathrm{bit}}, \mathbf{A}^n_{\mathrm{bit}}, \mathbf{B}^n_{\mathrm{bit}}, \mathbf{C}^n_{\mathrm{bit}}, [\mathrm{Tr}(\mathbf{M})]^n_{\mathrm{bit}}, [\det(\mathbf{M})]^n_{\mathrm{bit}}, \mathbf{R}^n_{\mathrm{bit}}$
10.	*if* $n < N-1$ *//, create the sensed results*
11.	*apply the update process of eq. (8)*
12.	*derive the salient poinst* $\mathfrak{s}^n$ *for image* $\mathbf{I}^n_{\mathrm{sensed}}$
13.	*if* $n > 0$ *// create the mask for the next bitplane*
14.	*for each point* (r,c), $0 \le r < R$, $0 \le c < C$
15.	*derive* $w_{n-1}[r,c]$

Table 1. Pseudocode for the proposed approach

The process can be continued in the same way for each subse-

quent bitplane $n-1, n-2, \ldots, 0$ captured by the sensor. If the computation/energy resources are exhausted, or the derived salient points are considered sufficient for the particular application, the salient point detection can be terminated. For each bitplane n the incremental approach derives the same results as the conventional approach that processes all $N-n$ bitplanes simultaneously.

By observing the indicative results of Figure 1, one notices that the derived salient points for each bitplane tend to be clustered around the areas of previously-derived salient points from the higher bitplanes. This is to be expected as strong salient points are likely to be detected early on during the sensing process and then remain localized within a certain region as additional bitplanes are added to the input image. For this reason, we introduce an adaptive image scanning mask $\mathbf{W}_{n-1}$, which focuses the image sensing and the processing around the areas of previously-derived salient points. Formally, $w_{n-1}[r,c]$ is 1 if at least one a salient point was detected at distance Z from the point r,c in the immediately most significant bitplane n. and 0 otherwise The steps of the proposed approach are summarized in Table 1

3. COMPUTATIONS AND SENSING

In this section we present the metrics used for quantification of the required computation and sensing energy.

3.1 Complexity of Variable Bit-width Arithmetic Operations

We are quantifying the differences of the conventional computation of the salient point detection algorithm versus the proposed approach. These differences will be studied in terms of the computational effort required to complete the detection task, whether it is for a single bitplane or for the entire set of bitplanes. In this respect, a common metric is the required number of additions and multiplications. However, arithmetic operations in the proposed incremental refinement approaches deal with data with significantly-reduced bitwidth in comparison to the conventional computation of the salient point detection algorithm that processes all bitplanes at once. Thus, we follow the approach proposed in [7] and utilize the following metrics inspired by classic work in the area-time complexity of binary multiplication and addition [8].

Definition 1 (from [7]): Addition of two numbers represented with N_1 and N_2 bits, each having an additional bit as the sign bit as in (3), requires the following number of operations:

$$\mathrm{Cost}_{\mathrm{add}} = \begin{cases} \max\{N_1, N_2\} + 1 & \text{, if both numbers are nonzero} \\ 0 & \text{, otherwise} \end{cases}$$

Definition 2 (from [7]): Multiplication of two numbers represented with N_1 and N_2 bits, each having an additional bit as the sign bit as in (3), requires the following number of operations:

$$\mathrm{Cost}_{\mathrm{mult}} = \begin{cases} (\max\{N_1, N_2\} + 1) \cdot (\min\{N_1, N_2\})^{1+\xi} \\ 0 \end{cases}$$

with $\xi \geq 0$ a system parameter indicating how "hard" is binary multiplication in comparison to binary addition. □

As discussed in [7], they can be intuitively viewed as follows: Assume a virtual processing element (PE) able to perform signed addition between two bits and the carry information. Starting from the LSB, addition is (maximally) requiring $\max\{N_1, N_2\} + 1$ activations of the PE for two numbers with N_1 and N_2 bits. Similarly, by viewing multiplication as cumulative additions, the number of activations of the PE is given by with the parameter ξ indicating the cost of accumulating the intermediate results. If any of the two operands is zero, no operations are required (apart from a minimal "zero detection" effort), since the result is trivial.

3.2 Energy Requirements for Adaptive Image Sensing

If we assume that sensing an individual bit using successive analog-to-digital sensors (e.g. the CMOS-based image sensor of Yang et al [6]) consumes E_{bit} Joules, we can derive the overall energy consumption of the conventional approach for the computation of the detector up to (and including) bitplane n as:

$$\mathrm{Energy}_{\mathrm{conventional}}(n) = R \cdot C \cdot (N-1-n) \cdot E_{\mathrm{bit}} \text{ Joules} \quad (9)$$

This equation shows a linear increase with the number of bitplanes. On the other hand, the proposed approach involves adaptive sensing for each individual bitplane n based on the binary mask $\mathbf{W}_n$. The sensing requirements in this case are:

$$\mathrm{Energy}_{\mathrm{incremental}}(n) = R \times C + \sum_{i=n}^{N-2} \sum_{r=1}^{R} \sum_{c=1}^{C} w_i[r,c] \times \mathrm{E}_{\mathrm{bit}} \text{ Joules} \quad (10)$$

since we sense all bits of bitplane $N-1$ and then sense bitplanes $N-2, \ldots, n$ according to binary masks $\mathbf{W}_{N-2}, \ldots, \mathbf{W}_n$. As the binary mask $\mathbf{W}_i$ does not cover the entire input bitplane of the image, we expect the energy requirements of the proposed approach to be less than the that of the conventional approach.

4. EXPERIMENTAL RESULTS

We experimented with a variety of natural images including human faces, pictures of scenery and objects as well as surveillance pictures. Extensive experiments are presented in [10].

The experimental settings for the utilized SPD were set as in [3], i.e. $k = 0.06$, $\sigma^2 = 2$, $\Theta = 0.01 \cdot \max\{\mathbf{R}^n_{\mathrm{sensed}}\}$. Concerning parameters specific for the proposed incremental approach, we set the size of the window parameter for the sensing mask around each salient point to: $Z = 80$ pixels for $n = 7$, $Z = 60$ pixels for $n = 6$, $Z = 50$ pixels for $n = 5$, and $Z = 30$ for $0 \leq n \leq 4$. These settings were chosen such that identical results between the proposed incremental approach and the conventional algorithm were obtained for all our experiments.

First we present the derived results with the intermediate detectors when subsets of the input source are used. Figure 2 presents typical results for bitplanes n=7,...,3. We notice that most of the obtained salient points tend to remain in similar positions but certain points shift according to the new source information obtained from each input bitplane. In some cases, new salient points may appear from one bitplane to the next due to the appearance of new shades of illumination. Therefore an appropriately large window must be set for the sensing mask around each salient point.

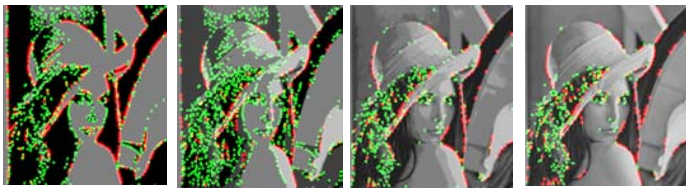

Figure 2. Incremental refinement of salient point detection at $n = 7, 6, 5, 3$. Green dots represent corners while red dots edges.

The quality and quantity of the obtained points depends on the image content. Overall, we observe that the number of salient points tends to decrease as edges and corners of the image are refined by sensing additional bitplanes of the image. Here, we summarize two important aspects of the presented results:

- Both the original SPD that processes the input source bitplanes simultaneously and the proposed incremental SPD that processes the source incrementally obtain the same points for corners and edges for all utilized source accuracies.
- The incremental approach derives these points progressively for each new input image bitplane by refining the results of the previous bitplanes. This means that the proposed incremental algorithm can terminate at any input image accuracy and provide the best results obtained for this accuracy level. In fact, all the results presented for each image were computed by executing the incremental algorithm once and extracting the output salient points after each input bitplane is processed. To the contrary, the conventional approach needs to be executed multiple times in order to produce the same results for each bitplane.

Apart from the visual inspection of the results, in order to quantitatively test the accuracy of each intermediate SPD result we use the Chamfer distance between the set of points derived for the lowest bitplane ($n = 3$) and the set of points of each individual bitplane. The derived distances for all the images of the utilized test-set are presented in Figure 3. Overall, we conclude that incremental refinement of salient points tends to provide results which are relevant to the final salient points selected by the detector. In addition, this relevance tends to improve with increased sensing.

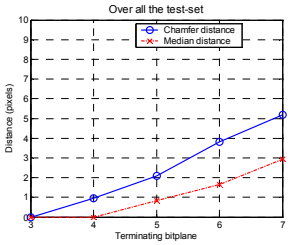

Figure 3. Chamfer and median distance (pixels) for all images.

Finally, we present the results for the computation and energy estimates of the proposed approach versus the conventional approach that performs the salient point detection utilizing the entire set of image bitplanes and is not refinable, utilizing the metrics of Definition 1 and Definition 2. We also report the required sensing energy based on (9)-(10). Typical results are presented in Figure 4 where with dashed lines we present theoretical results based on source modeling (for details see [10]).

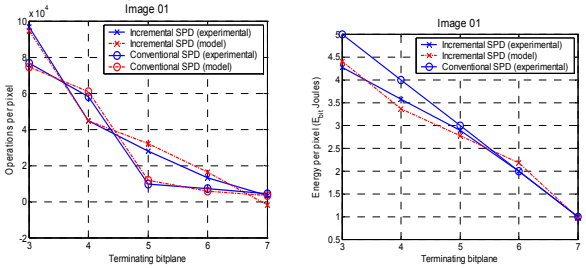

Figure 4. Computational and energy requirements for"Image 01". Left: Operations per pixel for terminating the processing at any bitplane $n = 7,6,5,4,3$. Right: Estimated energy requirements.

For terminating bitplanes $n = 5,6$, the results of Figure 4 demonstrate that the proposed incremental approach requires increased computational resources in comparison to the conventional realization of the SPD. However, for the cases of low terminating bitplanes ($n = 3,4$), comparable computation to the conventional approach is required. This is especially true for smooth images with low texture characteristics, where the incremental SPD also appears to provide very relevant results to the final detector for $n = 3$. In addition, we remark that the conventional approach would require significantly higher computational resources if it were to derive all the intermediate results that the incremental approach derives, since the conventional detector has to process the entire source for each newly-obtained bitplane. Concerning the sensing energy, the experimental results demonstrate that the proposed approach offers significant reductions for images with low texture characteristics (in the order of $20 \sim 50\%$), especially for the low terminating bitplanes.

5.CONCLUSIONS

We investigated an approach for deriving salient points incrementally by increasing the number of bits sensed from the input image. The inherent advantage of the proposed approach is that intermediate salient point detectors can be derived by increasing the source precision. We have analyzed the results of these detectors experimentally for a variety of images and presented initial evidence that incremental salient point detection can successively refine the quality of the derived output.

6.REFERENCES

[1] H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," *Tech. Report CMU-RI-TR-80-03*, Robotics Inst. Carnegie Mellon University, Sept. 1980.

[2] C. Harris and M. Stephens, "A combined corner and edge detector," *Alvey Vision Conf.*, vol. 1, pp. 147-151, 1988.

[3] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Internat. J. of Comp. Vision*, vol. 37, no. 2, pp. 151-172, Feb. 2000.

[4] W. Yuan and K. Nahrstedt, "Energy-efficient CPU scheduling for multimedia applications," *ACM Trans. on Computer Syst.*, vol. 24 , no 3, pp. 292 - 331, Aug. 2006.

[5] M. B. Wakin, J. N. Laska, M. F. Duartem D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," *Proc. IEEE Internat. Conf. Image Proc.*, ICIP, 2006.

[6] D. X. D. Yang, B. Fowler, and A. El Gamal, "A Nyquist-rate pixel-level ADC for CMOS image sensors," IEEE *J. of Solid State Circ.*, vol. 34, no. 3, pp. 348-356, March 1999.

[7] Y. Andreopoulos and M. van der Schaar "Incremental refinement of computation for the discrete wavelet transform," *IEEE Trans. on Signal Process.*, to appear.

[8] R. P. Brent and H. T. Kung, "The area-time complexity of binary multiplication," *J. of the Assoc. for Comp. Machin.*, vol. 28, no. 3, pp. 512-534, Jul. 1981.

[9] A. Oikonomopoulos, I. Patras and M. Pantic, "Spatiotemporal salient points for visual recognition of human actions," *IEEE Trans. Syst., Man, and Cybern.- Part B*, vol. 36, no. 3, pp. 710-719, Jun. 2006.

[10] Y. Andreopoulos and I.Patras "Incremental refinement of computation for the discrete wavelet transform," *IEEE Trans. on Image Process.*, submitted