

# VIDEO ENHANCEMENT ON AN ADAPTIVE IMAGE SENSOR

Maria E. Angelopoulou, Christos-Savvas Bouganis, Peter Y. K. Cheung

Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BT, UK  
email: {m.angelopoulou, christos-savvas.bouganis, p.cheung}@imperial.ac.uk

## ABSTRACT

The high density pixel sensors of the latest imaging systems provide images with high resolution, but require long exposure times, which limit their applicability due to the motion blur effect. Recent technological advances have led to image sensors that can combine in real-time several pixels together to form a larger pixel. Larger pixels require shorter exposure times and produce high-frame-rate samples with reduced motion blur. This work proposes ways of configuring such a sensor to maximize the raw information collected from the environment, and methods to process that information and enhance the final output. In particular, a super-resolution and a deconvolution-based approach, for motion deblurring on an adaptive image sensor, are proposed, compared and evaluated.

**Index Terms**— motion deblurring, resolution, adaptive image sensor

## 1. INTRODUCTION

State-of-the-art imaging systems employ high density pixel sensors. Due to the small pixel size of such sensors, long exposure times are required for the photodiodes to achieve an adequate signal to noise ratio [1]. If the camera is shaking or/and objects are moving in the scene during the integration of light on the photodiodes, the result is motion blur, *i.e.* low temporal resolution. This trade-off between spatial and temporal resolution limits the applicability of such imaging systems. For the rest of the paper, LR (HR) refers to low (high) spatial and, thus, high (low) temporal resolution.

Recently, researchers have focused on enhancing resolution in both time and space. In [2], multiple cameras are used to capture a fast moving scene with different subpixel and subframe shifts. The algorithm in [2] treats motion blur independently of the cause of the temporal change, but requires a large number of cameras. The large number of cameras introduces additional registration problems, not only due to the large number of captured images, but also due to the increased distances between the camera axes, which limits the applicability of the system. In [3], an LR imaging device deblurs the image captured by the HR device, by obtaining motion information for the estimation of the point spread function (PSF). This approach focuses on capturing a single image and solv-

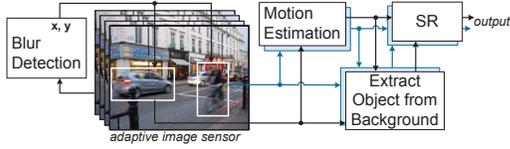
ing the blur caused by the undesired global motion due to camera shaking. The proposed system uses either two separate image sensors or a sensor with an LR periphery. The first technique can be extended to deal with the motion of objects, as motion trajectories can be detected anywhere in the frame. However, the use of two sensors results in registration-related problems and an increased size of the device.

Recent advances in imaging technology have produced sensors no longer subject to the constraint of time and space invariant pixel size [4, 5]. Elementary pixels can be grouped together to form a larger pixel where and when necessary.

Inspired by these advances, this work explores how an adaptive image sensor can be configured, so as to maximize the raw information collected from the environment. Appropriate methods are proposed to further process that information and reconstruct a final output of both high temporal and high spatial resolution. In particular, an adaptive sensor is used for motion deblurring, based on the configuration of certain pixels of the sensor to larger sizes to produce high-frame-rate samples. In summary, the contributions of this paper are: (i) A deconvolution-based motion-deblurring system is proposed, which employs spatially multiplexed pixels of different sizes on the sensor. (ii) A super resolution (SR) motion-deblurring system is proposed and the effect of motion magnitude and corresponding motion blur on the desired pixel size is explored. (iii) A detailed comparison of the proposed SR-based and deconvolution-based systems is presented.

## 2. SR-BASED MOTION DEBLURRING

Let  $S_h$  denote the size of the elementary pixel of resolution HR (*i.e.* the highest spatial and lowest temporal resolution), and let  $m$  and  $n$  be the height and width of an area of the sensor measured in  $S_h$  units. That area may include pixels larger than  $S_h$  and, thus, produce multiple time samples during the HR integration. If all pixels, *regardless of their size*, are considered as points in the 3-D space, then during the HR integration  $m \times n$  such points are produced for an  $m \times n$  area. The distribution of these points between time and space is determined by the pixel size. Increasing the pixel size of an area, decreases the density of these points on the 2-D plane and increases their density along the time axis, as the total number of points should remain  $m \times n$  for the given area. Thus, if the



**Fig. 1.** SR-based motion-deblurring on an adaptive image sensor.

pixel size of an area  $Q$  equals  $k \times k S_h$ , then  $k \times k$  LR time samples are produced for  $Q$  during the HR integration.

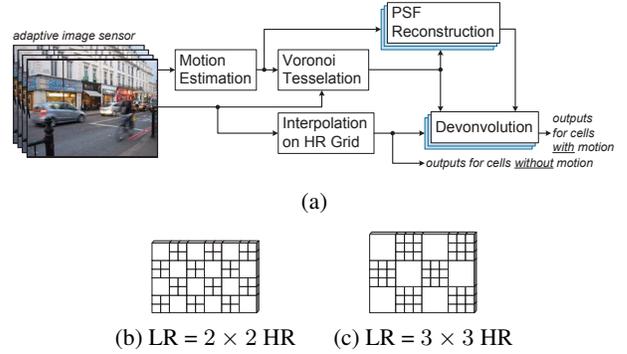
The proposed SR-based system is shown in Fig. 1. Motion areas are located on the frame and are configured to larger pixel sizes forming LR areas, whereas areas with slow motion or no motion are configured to HR pixels. During HR integration, a sequence of LR frames with reduced blur is produced at every LR area. Each LR area is spatially enhanced using SR to estimate a high resolution frame based on the LR inputs [6]. In [7], we have presented a real-time hardware implementation of SR on a field-programmable gate array (FPGA).

The *Blur Detection* block locates areas for LR configuration. If the blur is due to camera shaking, a single LR area spans the entire sensor. The *Motion Estimation* block reads the sequence of LR frames produced at each LR area and returns the motion vectors, *i.e.* the displacements between each LR frame and the reference LR frame. These are used by the SR unit to enhance the spatial resolution. Error values rendered by *Motion Estimation* [8] are used to weight the contribution of different LR samples. Also, to increase the spatial information available to the SR block, LR samples before and after the integration interval of interest contribute in SR with adjustable weights. Before executing SR on the group of LR frames, the static background is removed based on the information from the motion vectors. The spatial resolution and the frame-rate of the output are those of the HR sequence, where the motion blur has been removed in the LR regions.

### 3. DECONVOLUTION-BASED MOTION DEBLURRING

An overview of the deconvolution-based system is shown in Fig. 2(a). Under this framework, a motion PSF needs to be constructed for every moving object. Then, a deconvolution is applied separately on each object to produce a deblurred output for the part of the frame occupied by that object.

In order to achieve deblurring via deconvolution using a single sensor, a hybrid grid consisting of spatially multiplexed HR and LR pixels is used. Different sizes of LR pixels can be configured as shown in Fig. 2(b, c), so that larger LR pixels will be formed in areas with larger blur. The time samples captured by the LR pixels are the inputs to the *Motion Estimation* block. This block renders a set of 2-D points (one point per LR sample) belonging to the continuous motion trajectories of individual features during the HR integration. We refer



**Fig. 2.** (a) Deconvolution-based motion-deblurring on an adaptive image sensor. (b, c) Multiplexed LR and HR pixels of a hybrid grid.

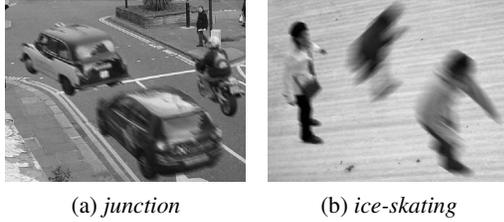
to this discrete subset of the actual motion trajectory by *discrete motion set*. Neighboring features with similar discrete motion sets are grouped together and a center of mass is determined on the frame for each group. Voronoi tessellation is applied on the set of these centers of mass. For each Voronoi cell a motion PSF is constructed on the HR grid by applying spline interpolation on the points of the associated discrete motion set and calculating the “energy” along the PSF as in [3]. Once the PSFs are formed, deconvolution is applied separately on the HR grid of each cell. The input of each deconvolution block consists of the raw HR pixels of the hybrid grid and the interpolated HR information formed at the LR pixels of the grid. The interpolated HR pixels are computed by solving a linear system formed by weighting the surrounding raw HR pixels and the underlying LR information. For the cells where no motion exists, the reconstructed output is the output of this interpolation. The final reconstructed frame consists of the individual outputs produced for every cell.

As different degrees of blur may exist on the same frame, at every configuration of the sensor, LR areas of certain pixel size may be formed on independent motion areas, as in the SR approach. This is achieved by including in the system of Fig. 2(a) a *Blur Detection* block to locate LR areas, as in Fig. 1. Then, at every HR integration, each LR area is tessellated, and different subregions are related to different PSFs.

### 4. COMPARING THE TWO APPROACHES

Depending on the technical specifications of the sensor itself, we distinguish two cases: (a) the sensor can be configured in every HR integration, and (b) a sparser configuration can be performed, *i.e.* every  $N$  HR frames. The difference between (a) and (b) lies in the accuracy in locating motion regions that will be configured to larger pixel sizes to form LR areas.

Ideally, all pixels of the scene belonging to the static background would be configured on the sensor with HR pixels and would, therefore, *not* belong to any LR area. Due to the high accuracy in locating motion regions, case (a) is closer to that

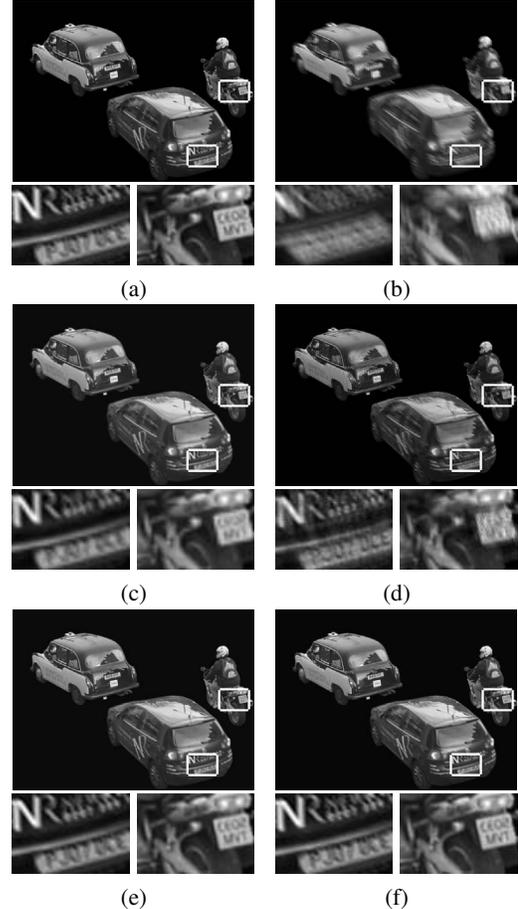


**Fig. 3.** Motion blur decreases the temporal resolution of frames.

ideal situation than case (b). Thus, in (a), LR areas include only few background pixels and mainly consist of dynamic parts. For the reconstruction of dynamic parts, the SR approach is the most effective of the two, due to: (i) In this approach, the LR samples at the input of the SR block experience reduced motion blur. Therefore, the blending, between the moving objects and the static background (Fig. 3(b)), is reduced. In the deconvolution approach, the input of the deconvolution block comprises an HR frame of low temporal resolution where that blending is large for large motions, complicating background extraction compared to the SR approach. (ii) As blur increases, the size of LR pixels increases to produce an output with sufficiently reduced blur. Larger LR pixels require more LR samples for SR reconstruction [6]. These can be obtained, as the number of produced samples increases with the pixel size (Sect. 2), and the SR approach is effective. In the deconvolution approach, the quality of the HR pixels, that are interpolated on each LR pixel of the hybrid grid to produce the input of deconvolution (Sect. 3), degrades as the LR pixel gets larger. This affects the quality of the output. Thus, the SR approach is more effective for extensive blur.

In (b), LR areas are not formed as often on motion regions, as the sensor is configured sparsely. Thus, moving objects are not located as often on the frame and wider LR areas are created that include moving objects, but also significant parts of the background. For very sparse configuration, a single LR area spans the entire frame. The configuration of LR areas should allow high-quality reconstruction of both moving and static parts. If large uniform LR pixels are formed, as in the SR approach, the quality of the reconstructed static regions cannot surpass that of interpolation applied on a uniform LR grid. If a hybrid grid is formed, 50% of the sensor is configured to HR pixels. Thus, interpolation is limited on the LR pixels of the grid and is optimized using the raw HR information of the surrounding HR pixels. As configuration gets sparser, LR areas contain more background pixels, and the hybrid grid of the deconvolution approach is more suitable.

To conclude, the SR approach is the most suitable for the reconstruction of moving parts and prevails when, due to frequent configuration, the LR areas mainly consist of non-static parts. The hybrid grid of the deconvolution method allows effective reconstruction of static parts and is suitable for sparse configuration, where LR areas also include large static parts.



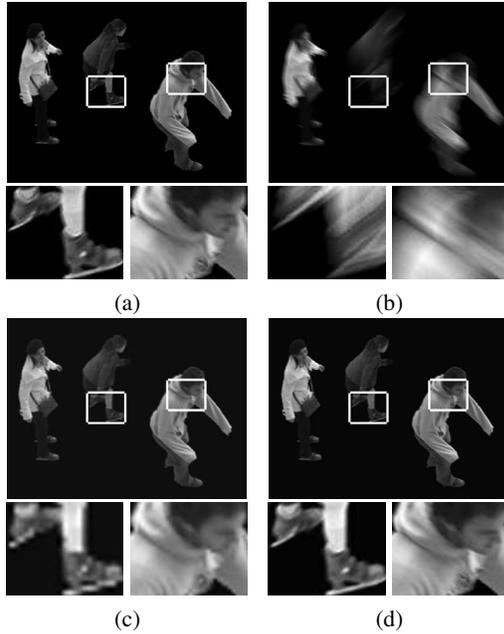
**Fig. 4.** (a) Ground-truth frame. (b) Blurred output for HR spatial resolution. (c) Bicubic interpolation of reference LR frame. (d) Output of deconvolution-based motion deblurring. (e, f) Output of SR-based deblurring after 10 (e) and 30 (f) iterations of the algorithm.

## 5. PERFORMANCE EVALUATION

In both approaches, *Motion Estimation* is implemented using Lucas-Kanade optical flow [8] and Shi-Tomasi good feature extraction [9]. Iterative Back Projection [6] is employed for SR, and Lucy-Richardson deconvolution is implemented.

The methods are evaluated on *junction* and *ice-skating* (Fig. 3), with HR frame size  $480 \times 640$ . The moving objects are isolated from the background, to exclude errors associated with the background extraction method from the evaluation process. The LR sequences were synthetically produced by applying temporal and spatial blur on a dense sequence  $S_d$  which acts as the real-world scene. The time blur of a sensor with HR pixels spans, in both experiments, 144 frames of  $S_d$  (Fig. 4(b), 5(b)). In the first experiment, the two methods are compared. In the second experiment, the SR approach, which is the most suitable for extensive blur (Sect. 4), is used to deblur a frame containing different motion magnitudes.

Due to space limitation, for the reconstruction of dy-



**Fig. 5.** (a) Ground-truth frame. (b) Blurred output for HR spatial resolution. (c) Bicubic interpolation of reference LR frame. (d) Output of SR-based deblurring after 30 iterations of the SR algorithm.

dynamic parts we only display qualitative results, which clearly demonstrate the sharpness achieved after motion deblurring.

The *junction* sequence includes three moving objects with similar blur. The LR sequence was generated by applying a temporal blur of 36 frames and a spatial blur of  $2 \times 2$  on  $S_d$ .

To apply the two approaches, all motion areas are configured to LR pixels of size  $2 \times 2$  HR for the SR approach, and to the grid of Fig. 2(a) for the deconvolution approach. In the SR approach, two LR frames at each side of the integration interval are used for additional robustness. Figures 4(d-f) show the reconstruction of foreground objects by the two approaches, and the detail images demonstrate the higher sharpness of the reconstructed outputs compared to Figures 4(b, c).

If the configurations described in the previous paragraph are applied on the entire sensor, the static background also needs to be reconstructed on the HR grid. Considering the background of Fig. 3(a) without the presence of foreground objects, the root mean square error between the reconstructed and the ground-truth background is: (i) 4.6 after spatially enhancing the hybrid grid of the deconvolution approach and (ii) 6.9 after applying bicubic interpolation on the uniform large pixels of the SR approach.

In *ice-skating*, moving objects with different blur are acquainted. Clock-wise from left (Fig. 3(b)), these are: *Susan*, *Amy*, *Jack*. The corresponding LR sequences were created independently, by applying on  $S_d$  a temporal blur of 36, 9 and 16 and a spatial blur of  $2 \times 2$ ,  $4 \times 4$  and  $3 \times 3$ , respectively.

Three motion areas are formed in the SR approach, with

LR pixel size  $2 \times 2$ ,  $4 \times 4$  and  $3 \times 3$  HR, producing 4, 16 and 9 time samples during HR integration. As the LR pixel size increases with the local motion magnitude, SR reconstruction (Fig. 5(d)) outperforms bicubic interpolation (Fig. 5(c)) and successfully handles large degrees of blur.

## 6. CONCLUSION

In this paper, we propose two methods of configuring an adaptive image sensor to maximize the information collected from the environment. For each configuration, an appropriate system is developed to process that raw information and enhance the output. Specifically, an SR and a deconvolution-based approach, that execute motion deblurring on an adaptive sensor, are proposed, compared and evaluated. Results demonstrate that the SR method performs better in dynamic regions and is, thus, preferable when the sensor can be frequently configured and motion areas are accurately located. Moreover, this method effectively manages large blur by increasing the LR pixel size. Experiments show that the deconvolution approach achieves better reconstruction of static regions and is suitable for sparse configuration, as large static parts are included in LR areas. Future work includes evaluating errors associated with the extraction of foreground objects from the static background and extending the methods to non-rigid objects.

## 7. REFERENCES

- [1] A.E. Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuits & Devices Magazine*, vol. 21, no. 3, pp. 6–20, May–June 2005.
- [2] E. Shechtman, Y. Caspi, and M. Irani, "Space-Time Super-Resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 531–545, April 2005.
- [3] M. Ben-Ezra and S.K. Nayar, "Motion-based motion deblurring," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 689–698, June 2004.
- [4] Foveon, Inc., "Foveon X3 Image Sensor," [www.foveon.com](http://www.foveon.com).
- [5] T.G. Constandinou, P. Degenaar, and C. Toumazou, "An Adaptable Foveating Vision Chip," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2006, pp. 3566–3569.
- [6] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *CVGIP: Graphical Models and Image Proc.*, vol. 53, no. 3, pp. 231–239, May 1991.
- [7] M.E. Angelopoulou, C.-S. Bouganis, P.Y.K. Cheung, and G.A. Constantinides, "FPGA-based Real-time Super-Resolution on an Adaptive Image Sensor," in *International Workshop on Applied Reconfigurable Computing (ARC)*, 2008, to appear.
- [8] J.-Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the algorithm," Microprocessor Research Labs, Intel Corporation, 2002.
- [9] J. Shi and C. Tomasi, "Good Features to Track," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 1994, pp. 593–600.