A SENSOR-BASED APPROACH TO LINEAR BLUR IDENTIFICATION FOR REAL-TIME VIDEO ENHANCEMENT

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

Super-resolution (SR) methods are largely affected by the accurate evaluation of the Point Spread Function (PSF) that is related to the input frames. When the frames are degraded by heavy motion blur, the PSFs are highly non-isotropic, which further complicates their estimation. The ill-posed nature of blur identification is usually addressed using the assumption of linear and uniform motion. However, in real-life systems, this may deviate significantly from the actual motion blur. To resolve the above, this work proposes combining a scheme that validates the initial motion assumption with the real-time reconfiguration property of an adaptive image sensor. If the linearity and uniformity assumption is invalid for a given motion region, the sensor is locally reconfigured to larger pixels that produce higher frame-rate samples with reduced blur. Once the appropriate configuration that gives rise to a valid motion assumption is applied, highly accurate PSFs are estimated, resulting to an improved SR reconstruction quality.

***Index Terms*—** motion blur, super-resolution, blur identification, adaptive image sensor

1. INTRODUCTION

For super-resolution (SR) [1, 2] to be effective, the PSFs of the input frames should be accurately estimated, since these describe the contribution of each frame in the final reconstruction. In real-time video capturing, fast moving objects produce locally heavy blur, and are thus related to significantly non-isotropic PSFs, whose identification is a highly ill-posed problem. In the literature, motion blur identification is simplified by employing a linear and uniform motion assumption [3, 4], where 'linear' implies a motion trajectory that is accurately approximated with a first order polynomial, and 'uniform' indicates identical PSF weights. This assumption reduces the problem of PSF estimation to the identification of two parameters: the motion extent and direction. However, in real-life systems, the actual blur may deviate considerably from an ideal linear and uniform PSF. This renders the initial assumption unrealistic, directly affecting the reconstruction.

In [5], a video enhancement system based on an adaptive image sensor [6, 7] is proposed. Possible sensor configurations that maximize the captured raw data are explored and are combined with processing methods that increase the spatio-temporal resolution of the output. Two approaches are presented, a deconvolution-based and a SR-based approach. For highly dynamic regions, the SR-based approach is proven to be more appropriate. In this approach, each motion region is locally configured to a uniform grid of large pixels, rendering high frame-rate samples with reduced motion blur, whereas areas with slow motion or no motion are configured to the elementary pixels of the sensor. The spatial resolution of the motion areas is increased by fusing the high frame-rate samples with SR techniques. However, for very fast motion, these samples are inevitably blurred themselves, and the fidelity of the SR output is bounded by the blur effect.

This work addresses the above by proposing an interaction of joint blur identification and validation (BIV) [8] with an adaptive image sensor. BIV not only estimates linear blur, but also validates the initial assumption. This keeps the blur identification process simple, since linear and uniform motion is assumed, while identifying cases where the assumption is invalid and thus the PSF estimation is inaccurate. To resolve such cases, the adaptive sensor is configured to larger pixels that produce samples with reduced blur. Once the appropriate pixel size is employed for which the linear motion assumption is valid, accurate PSFs are estimated, increasing the SR reconstruction quality. In [8], a real-time BIV architecture is proposed and implemented on reconfigurable hardware.

The contributions of the paper are: (i) A methodology is proposed that includes the interaction of an adaptive image sensor with a joint blur identification and validation scheme. The proposed methodology increases the accuracy in the estimation of the PSFs related to the SR inputs, thus improving the SR reconstruction quality. (ii) The complete algorithm for real-time video enhancement is presented, including both the sensor configuration framework and the processing of the captured data. The system's performance is evaluated under linear and non-linear motions and under various noise levels.

2. JOINT IDENTIFICATION AND VALIDATION

The joint blur identification and validation (BIV) scheme of [8] employs the autocorrelation-based blur identification

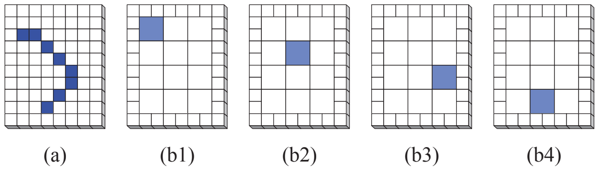


Fig. 1. Outputs for two sensor configurations during HR integration.

framework of [3]. The motion direction is the one that gives the minimum total intensity (TI) for the directional image derivative [3]. On that derivative, a mean autocorrelation function (ACF) is calculated. As explained in [8], if the linear and uniform motion assumption is valid, this ACF includes three dominant lobes: a positive lobe at lag 0 and two symmetrical negative lobes at a lag equal to the motion extent. A different number of lobes or a minimum ACF coefficient at a lag that could not possibly correspond to the motion extent, indicate an invalid assumption. Invalid assumption is also indicated, when there is not a clear minimum for the normalized total intensities of the directional derivatives.

3. ACCOUNTING FOR INTRA-FRAME MOTION IN A SYSTEM WITH AN ADAPTIVE SENSOR

This section explains how the reconfigurability of the sensor is combined with the BIV scheme to improve the reconstructed output. For the rest of the paper, ‘LR’ (‘HR’) refers to low (high) spatial and, thus, high (low) temporal resolution.

The adaptive image sensor can be locally configured to form LR areas that produce high frame-rate samples, due to the space-time trade-off [5], thus fragmenting the motion trajectory. Fig. 1 shows the raw outputs for two sensor configurations during the HR integration interval, *i.e.* the time required by the elementary pixels to achieve a certain signal to noise ratio (SNR). A grid of elementary pixels would render a single, motion-blurred frame (Fig. 1(a)), while a 2×2 configuration would give 4 time samples (Fig. 1(b1-b4)), each containing a fragment of the trajectory of Fig. 1(a). To increase the spatial resolution of the output, the LR samples are fused with SR techniques [2]. For SR to be effective, a PSF should be estimated within a certain accuracy for each sample.

The proposed system utilizes a simple blur detection block, which is based on the comparison of the strongest edges of the object with those of the background, to identify cases where the current configuration produces samples with negligible motion blur, as in Fig. 1(b1-b4). In such cases, an isotropic Gaussian PSF can be employed to associate each LR pixel to the HR pixels of the underlying HR grid [2]. In all other cases, the Gaussian assumption is inadequate, and blur identification is required to estimate the motion parameters. Thus the BIV scheme discussed in Sect. 2 is executed. If BIV finds the linear and uniform motion assumption invalid, the pixel size increases in the next sensor reconfiguration, to

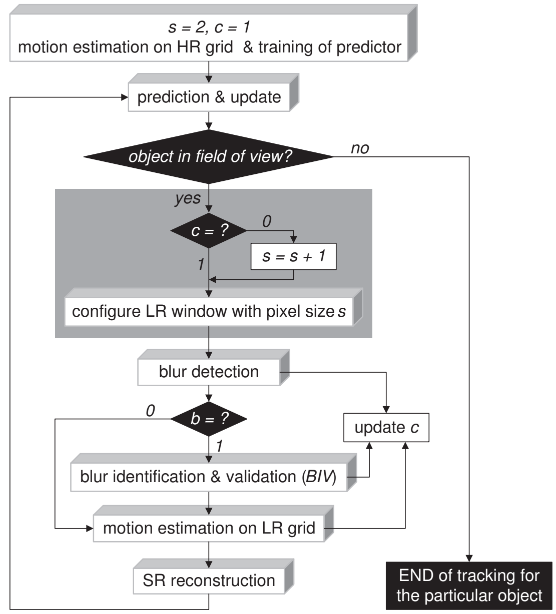


Fig. 2. The algorithm of the system operation for a dynamic region.

produce samples with reduced, more linear motion blur. In this manner, the initial non-linear and/or non-uniform motion trajectory is fragmented into shorter, more linear parts. If BIV finds the initial motion assumption valid, accurate linear PSFs can be estimated, and the pixel size thus remains constant.

The pixel size of the adaptive sensor depends on the outputs of *blur detection* and *BIV* blocks, as described above. These blocks comprise a classifier, whose binary output c determines the pixel size in the next sensor reconfiguration: If $c = 1$, the pixel size remains constant, as it allows an accurate estimation of the PSF, employing either the Gaussian or the linear and uniform motion assumption. If $c = 0$, the pixel size increases in the next reconfiguration. A mechanism that reduces the pixel size every N frames can be accommodated.

For linear motion, the non-isotropic PSFs should be consistent with the inter-frame motion vectors. In case of inconsistency, invalidity of the initial linearity assumption is indicated. The above can be used as an additional validity check, and is thus incorporated into the classifier that is described above, giving $c = 0$ in the case of inconsistency.

Ideally, the adaptive sensor would be reconfigured at every new HR integration. In reality, reconfiguration is sparser, depending on the technology of the given sensor. The proposed video enhancement system operates as follows. Moving objects are detected with a rough motion estimation on the HR grid, as indicated in Fig. 2. For every moving object, the control of Fig. 2 is employed, where s denotes the LR pixel size, with $s = 2$ corresponding to the smallest LR pixel size as demonstrated in Fig. 1(b1-b4), and c denotes the binary output of the classifier that determines the change in the pixel size in the next reconfiguration. Moreover, b denotes

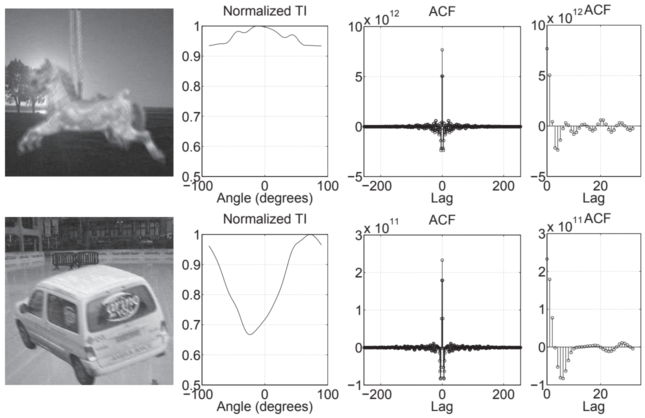


Fig. 3. TI and ACF for *carousel* (top) and *ambulance* (bottom row). On the right, a detail of the ACF is presented, for lags 0 to 33.

the binary output of the *blur detection* block; $b = 0$ indicates negligible motion blur, for which the Gaussian assumption is adequate and BIV is skipped, while $b = 1$ indicates that BIV is required. The part of Fig. 2 included in the gray rectangle is executed only in those HR integration intervals when the sensor is reconfigured. A Kalman filter predictor is employed to determine the position of the object in the next HR integration. When sensor reconfiguration occurs, that position determines the location of the LR area, while its pixel size depends on the validity of the linearity assumption in the last HR integration, indicated by the value of c. For $c = 0$ the pixel size increases, whereas for $c = 1$ it remains the same. For each LR area, an LR sequence with reduced blur is produced, and the PSFs are estimated, based on the outputs of the *blur detection* and *BIV* blocks, as described in the previous paragraphs. The LR samples are registered using a *motion estimation* block, and the corresponding PSFs are used by the *SR* block that executes the reconstruction on the HR grid [2]. This produces for each LR area an output with high resolution both in space and time, and thus motion deblurring is locally executed on the dynamic regions of the scene. At every new HR integration, the control starts at the second block of Fig. 2. The loop ends when the particular object exits the field of view.

4. PERFORMANCE EVALUATION

The performance of the video enhancement system of Fig. 2 is evaluated on the basis of the quality of the final output. Different parameters are employed, including the type of PSF assumption, the LR pixel size, and the noise level. The evaluation uses semi-synthetic data, *i.e.* real images are shifted, blurred, downsampled, and contaminated with noise, to produce LR sequences. Thus the 'ground-truth' frame is known and is used to evaluate the reconstruction quality. White Gaussian noise is applied, with SNRs ranging from 10 to 50 dB. The iterative SR approach of [2] is used, and 30 iterations

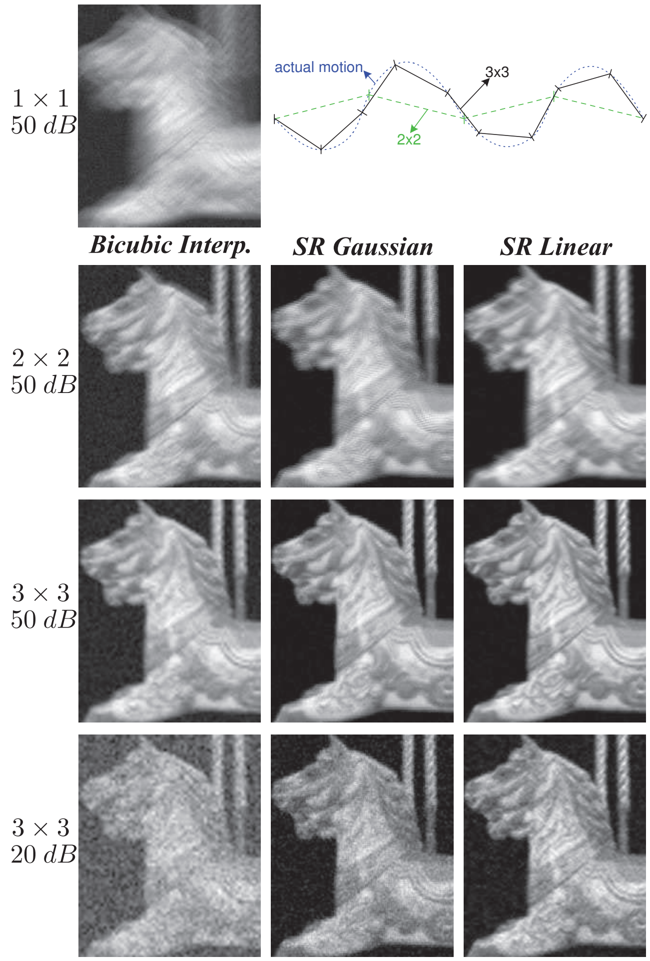
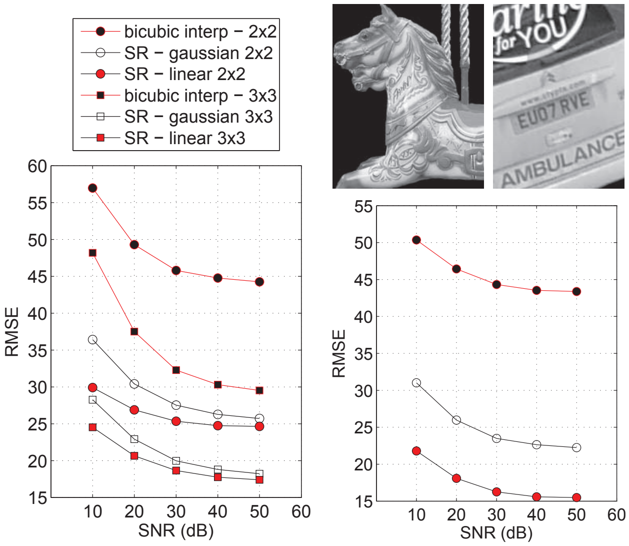


Fig. 4. *Row 1:* The raw output for configuration 1×1 (elementary pixel grid), and the intra-frame motion of the time samples produced during HR integration for each configuration. *Rows 2-4:* SR reconstructed outputs when employing the indicated PSF approximations.

are executed for each estimation. To exclude any evaluation errors due to the blending of the object with the background, SR is applied on the isolated foreground objects. The number of frames produced during HR integration for each pixel size is subject to the space-time trade-off [5]. If this number is k for the current configuration, k additional neighboring frames are used in SR, for increased robustness [2].

Fig. 3 shows the TI and ACF outputs of BIV (Sect. 2) for two moving objects with non-negligible motion blur ($b = 1$). Both cases employ 2×2 configuration and 256×256 LR resolution. The frame on the left is one of the 4 samples generated during HR integration at 50 dB SNR. According to the validation criteria described in Sect. 2 and more extensively discussed in [8], the TI and ACF outputs of Fig. 3 indicate that for the 2×2 configuration the linearity and uniformity assumption is valid for *ambulance*, and invalid for *carousel*. Indeed, the actual motion of *carousel*, is not sufficiently linear



(a) (b)

Fig. 5. Errors for the two sets of experiments, for various SNRs. The legend applies to both graphs, with (b) containing only the 2×2 configuration values. The ground-truth frames are shown at the top.

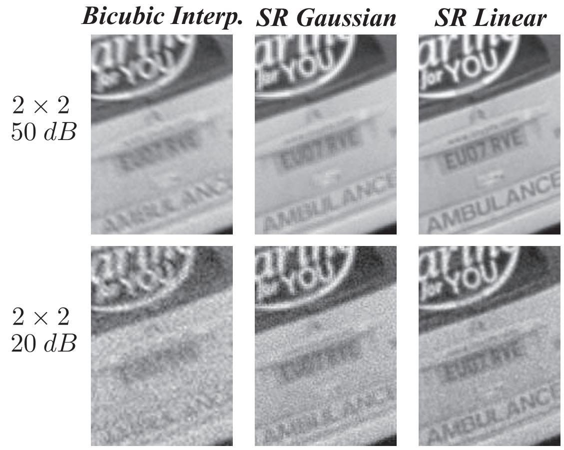


Fig. 6. Reconstructed outputs for *ambulance*.

for a 2×2 configuration, as the trajectory shows in Fig. 4 (top right). Thus a 3×3 configuration is employed next.

The detailed images of Fig. 4 show the SR reconstructed output for *carousel*, for 2×2 and 3×3 configurations. For reference purposes, bicubic interpolation is applied on a single LR frame, with magnification factors 2 and 3, respectively. The SR output is given both for Gaussian PSF approximations, whose support corresponds to the LR pixel size [1, 2], and linear PSFs estimated by BIV. The system output is that of the 3×3 configuration with SR that uses linear PSFs (Fig. 4).

The last row of Fig. 4 demonstrates the system robustness, presenting the reconstructed outputs for significantly noisy LR samples (SNR = 20 dB). Fig. 5(a) quantifies the evaluation giving the RMSE values with respect to the ground-truth, for the above scenarios and SNR from 10 to 50 dB.

Contrary to *carousel*, *ambulance* passes the validity check for 2×2 configuration (Fig. 3); thus the pixel size remains at 2×2. Fig. 6 presents the indicated outputs for 50 and 20 dB. The system output is that of SR with the linear PSF approximation. The associated errors are given in Fig. 5(b). It can be concluded that when the linear PSF is estimated, the SR output improves compared to the use of a Gaussian PSF.

5. CONCLUSION AND FUTURE WORK

This paper proposes a methodology for combining the reconfiguration property of an adaptive image sensor with a blur identification and validation scheme. The aim is to increase the reconstruction quality of the final output. The sensor grid is adapted to the local motions depending on the validity of the linear and uniform motion assumption. In this manner, the appropriate configuration is applied, and thus the optimal PSF approximation is employed in SR, improving the reconstruction of the output. In this paper, it has been demonstrated through experiments that the proposed system surpasses limits of traditional SR, proving the significance of adapting the sensor grid to the given motion. Future work includes extending the reconstruction to nonrigid motions, and investigating the approximations of PSFs with second order polynomials.

6. REFERENCES

[1] 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.

[2] M. E. Angelopoulou, C.-S. Bouganis, P. Y. K. Cheung, and G. A. Constantinides, "Robust Real-Time Super-Resolution on FPGA and an Application to Video Enhancement," *ACM Transactions on Reconfigurable Technology and Systems, to appear*, 2009.

[3] Y. Yitzhaky and N. S. Kopeika, "Identification of Blur Parameters from Motion Blurred Images," *Graphical Models and Image Processing*, vol. 59, no. 5, pp. 310–320, Sept. 1997.

[4] M. M. Chang, A. M. Tekalp, and A. T. Erdem, "Blur Identification Using the Bispectrum," *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2323–2325, Oct. 1991.

[5] M. E. Angelopoulou, C.-S. Bouganis, and P. Y. K. Cheung, "Video Enhancement on an Adaptive Image Sensor," in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2008, pp. 681–684.

[6] Foveon, Inc., "Foveon X3 Image Sensor," *www.foveon.com*.

[7] 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.

[8] M. E. Angelopoulou, C.-S. Bouganis, and P. Y. K. Cheung, "Blur Identification with Assumption Validation for Sensor-Based Video Reconstruction and its Implementation on FPGA," *IET Computers & Digital Techniques, submitted*, 2009.