RESTORATION AND RESOLUTION ENHANCEMENT OF VIDEO SEQUENCES

Vadim Avrin and Its'hak Dinstein

Electrical and Computer Engineering Department
Ben Gurion University of The Negev, Beer-Sheva, 84105, Israel
vadim@newton.bgu.ac.il dinstein@bguee.bgu.ac.il

ABSTRACT

Given a sequence of blurred low resolution images, the aim of this work is to produce a sequence of higher resolution and restored images. It is assumed that the point spread function of the given imaging process is a combination of a known blurring function and an estimated local motion function. The local motion estimation is obtained by the respective group delays of local adaptive filters. Preliminary experimental results are presented.

1. INTRODUCTION.

Image restoration operations are intended to filter out degradation effects generated by the imaging process. The aim of image restoration is to compute a restored image that is as close as possible to the original scene. An extensive review of image restoration can be found in [1]. Resolution enhancement operations are intended to increase the resolution of an image by estimating the values of the image at samples between the given pixels [2]. This paper proposes an operation by which restoration and resolution enhancement of image sequences are simultaneously obtained.

Restoration of multi-frame data was first introduced in [3]. Restoration algorithms for multi-channel images [4] and its extension to image sequences [5] have been recently proposed. In both cases the correlation between channels or between frames contribute to the restoration results. Resolution enhancement based on image registration is proposed in [6]. It is assumed that there is no motion in the image sequence. The resolution enhancement approach proposed in [2] deals with independent object motion, and uses a Bayesian interpolation method.

The work presented here is intended for restoration and resolution enhancement of video sequences. The proposed algorithm yields a sequence of super-resolution images. Local small displacements between consecutive frames of the input sequence are compensated using adaptive local filters. Each super-resolution frame is generated from the previous super-resolution frame and the back projected filtered sub-sequence of input frames. The basic assumption is that the input images are decimated versions of a super-resolution blurred and noisy image sequence. The aim is to estimate the original super-resolution image sequence. The following is a description of the proposed algorithm using a notation similar to that of [6].

2. THE PROPOSED APPROACH.

Let the input monochrome video sequence be $\{g_k\}$. The imaging process is modeled by:

$$g_k(m,n) = \sigma_k[h(f_k(x,y)) + \eta_k(x,y)]$$

Where, g_k is the k-th input image, f_k is the k-th super-resolution image, h is the blurring operator, η_k is the additive noise, and σ_k is a decimation function with a decimation factor L.

Denote by $(m,n)_k$ the pixel (m,n) of the input image g_k. The gray level of this pixel, $g_k(m,n)$, is a weighted sum of the pixels belonging to the respective *receptive field* of f_k, as shown in Figure 1. The weights are the respective elements of the blurring function h. The receptive field is uniquely defined by its center (p,q) and

This work was supported in part by The Paul Ivanier Center for Robotics and Production Automation , Ben Gurion University of The Negev, Beer Sheva, 84105, Israel.

by the region of support of h. A pixel $(m,n)_k$ is said to be influenced by a pixel (x,y) belonging to f_k if (x,y) is within the receptive field of $(m,n)_k$.

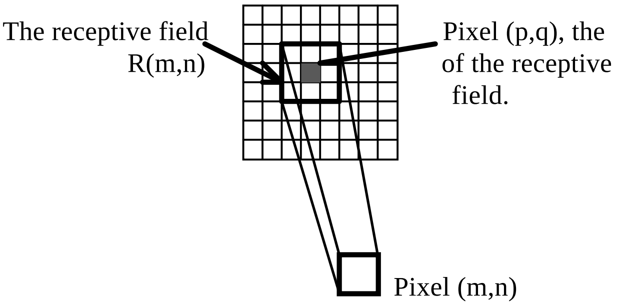

Figure 1. Receptive field illustration.

Let $\hat{g}_k$ be a decimated version of the blurred image $\hat{f}_k$, where $\hat{f}_k$ is an estimate of the super-resolution image f_k.

$$\hat{g}_k(m,n) = \sum_{(x,y)\in R_{m,n}} \hat{f}_k(x,y)\cdot h^{FP}_{k,m,n}(p-x,q-y)$$

where, $h^{FP}_{k,m,n}$ is a forward projection function explained later, $R_{m,n}$ is the receptive field of (m,n), and (p,q) is the center of $R_{m,n}$. The image $\hat{f}_k$ is an estimate of f_k. If the estimate is perfect, then $\hat{g}_k$ is equal to g_k. If this is not the case, The difference between $\hat{g}_k$ and g_k is used for the computation of $\hat{f}_{k+1}$ as follows:

$$\hat{f}_{k+1}(x,y) = \hat{f}_k(x,y) + \\ +\mu \sum_{(m,n)\in r_{x,y}} (\hat{g}_k(m,n) - g_k(m,n))\cdot h^{BP}_{k,m,n}(Lm-x,Ln-y)$$

where, μ is a normalizing factor, $r_{x,y}$ is the set of all pixels in $\hat{g}_k$ that are influenced by pixel (x,y), and $h^{BP}_{k,m,n}$ is a back projection function explained later.

Each one of the forward projection functions $h^{FP}_{k,m,n}$ is composed of two components. The first component is the blurring function h^{PSF} representing a known weighted sum averaging function of the sensing system. The second component, $h^{PRD}_{k,m,n}$, is a local function that minimizes the mean square difference between $g_{k+1}(m,n)$ and $h^{PRD}_{k,m,n} * \hat{g}_k(m,n)$, where * represents convolution.

We use the Wiener approach [7, page 20] to minimal square error prediction for the estimation of $h^{PRD}_{k,m,n}$.

Let the prediction error $e_k(m,n)$ be defined as:

$$e_k(m,n) = g_{k+1}(m,n) - \hat{g}_{k+1}(m,n)$$

The prediction $\hat{g}_{k+1}(m,n)$ in vector presentation is given by:

$$\hat{g}_{k+1}(m,n) = X^T_{k,m,n}\cdot W_{k,m,n}$$

Where $W_{k,m,n}$ is a column vector containing the lexicographic order of the two dimensional pxp local filter (m,n), and $X^T_{k,m,n} \in R^{p^2}$ is the transposed of a vector containing the lexicographic order of the pxp sub image of g_k centered at (m,n).

Let the local autocorrelation matrix at (m,n) be

$$R_{k,m,n} = E\{X_{k,m,n}\cdot X^T_{k,m,n}\}$$

The expectation $E\{\cdot\}$ is estimated within a small neighborhood containing $g_k(m,n)$. The correlation between the desired pixel value $g_{k+1}(m,n)$ and the respective input sub-image $X^T_{k,m,n}$ is given by

$$P_{k,m,n} = E\{g_k(m,n)\cdot X^T_{k,m,n}\}$$

The optimal prediction filter for the minimum mean square error $E\{e_k^2(m,n)\}$ is given by

$$W^{OPT}_{k,m,n} = R^{-1}_{k,m,n}\cdot P_{k,m,n}$$

The optimal two dimensional local prediction functions $h^{PRD}_{k,m,n}$ is obtained by reordering the elements of $W^{OPT}_{k,m,n}$.

An adaptive implementation can be used in order to compute the local prediction functions $h^{PRD}_{k,m,n}$ by the LMS algorithm [7] as follows:

The functions for k=0 are initialized in some manner. Then, for k=1,2,3,... the error $e_k(m,n)$ is computed by:

$$e_k(m,n) = g_{k+1}(m,n) - h^{PRD}_{k,m,n} * \hat{g}_k(m,n).$$

The error $e_k(m,n)$ is computed for each pixel $(m,n)_k$, using the prediction function $h^{PRD}_{k,m,n}$ that is particular to that pixel. Each prediction function is adapted according to

$$h_{k+1,m,n}^{PRD} = h_{k,m,n}^{PRD} - \gamma \cdot e_k(m,n) \cdot \hat{G}_k(m,n)$$

where γ is a convergence factor, and $\hat{G}_k(m,n)$ is a sub-image of $\hat{g}_k(m,n)$ respective to the region of support of $h_{k,m,n}^{PRD}$. The adaptation is performed within a small neighborhood containing $g_k(m,n)$. The group delay of each $h_{k,m,n}^{PRD}$ indicates the local motion vector.

The forward projection function $h_{k,m,n}^{FP}$ is the convolution between h^{PSF} and the interpolated (with factor L) version of $h_{k,m,n}^{PRD}$. The interpolation is required because the domain of the function $h_{k,m,n}^{PRD}$ is the input image sequence, and the domain of $h_{k,m,n}^{FP}$ is the super-resolution image sequence.

The function $h_{k,m,n}^{BP}$ is used for regulating the contributions of the respective pixels $(m,n)_k$ to the estimation of $\hat{f}_{k+1}(x,y)$. We follow [6] and set $h_{k,m,n}^{BP}$ to be equal to $h_{k,m,n}^{FP}$.

3. SIMULATION RESULTS.

We report here preliminary results as available at this time. The adaptive implementation has been applied to a sequence of shifted images. Figure 2 shows one of the images.

Figure 2. One of the images of the sequence.

One hundred and twenty images were used. The sequence was divided into three types of sub sequences, consisting of fifteen images each. The images of sub-sequence Type A where shifted one pixel diagonally towards the lower right corner. Type B consists of images with no shift, and the images of Type C were shifted diagonally one pixel towards the upper left corner.

We present here the results of the adaptation of one local 3x3 filter, adapted within a 10x10 window. In the ideal case, the shift in Type A sub-sequence should give rise to filter coefficients

$$h = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Similarly, the shift of Type C should yield

$$h = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and for the shift of Type B,

$$h = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The results of the adaptation process are summarized in Figure 3. The first three graphs present the values of the coefficients h(0,0) (the upper left), h(1,1), and h(2,2) of the filter matrix. The bottom graph depicts the mean square error between the prediction and the actual value.

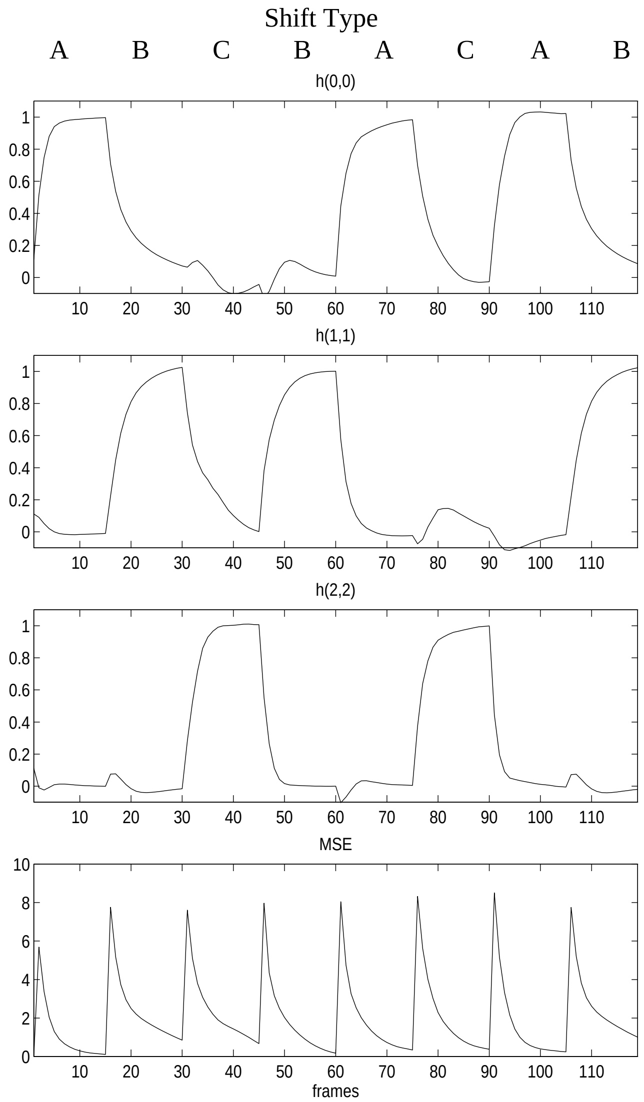

Figure 3. Convergence of Adaptive Filter Simulation.

We are currently applying the proposed restoration and super-resolution algorithm to image sequences taken under turbulence conditions. The input image sequences contain small local arbitrary motion of small neighborhoods between frames.

REFERENCES.

1. A. K. Katsaggelos, editor, Digital Image Restoration, Springer-Verlag, New-York, 1991.
2. R. R. Schultz, and R. L. Stevenson, “Extraction of High-Resolution Frames from Video Sequences”, IEEE Trans. on Image Processing, Vol. 5, No. 6, pp. 996-1011, June 1996.
3. R. Y. Tsai, and T. S. Huang, “Multiframe Image Restoration and Registration”, in Advances in Computer Vision and Image Processing, T. S. Huang editor, JAI Press, Greenwich CT., Vol. 1, pp. 317-339, 1984.
4. N. P. Galatsanos, et. al., “Least Squares Restoration of Multichannel Images”, IEEE Trans on Signal Processing, Vol. 39, No, 10, pp. 2222-2236, , 1991
5. M. K. Ozkan, et. al., “Efficient Multiframe Wiener Restoration of blurred and Noisy Image Sequences”, IEEE Trans on Image Processing, Vol. 1, No. 4, pp. 453-476, 1992.
6. M. Irani, and S. Peleg, “ Improving Resolution by Image Registration”, CVGIP: Graphical Models and Image Processing, Vol. 53, No. 3, May, pp. 231-239, 1991,
7. B. Widrow, and S. D. Stearns, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs, N.J., 1985.