Applications

Motion Deblurring and Super-resolution from an Image Sequence

B. Bascle, A. Blake, A. Zisserman

Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, England

Abstract. In many applications, like surveillance, image sequences are of poor quality. Motion blur in particular introduces significant image degradation. An interesting challenge is to merge these many images into one high-quality, estimated still. We propose a method to achieve this. Firstly, an object of interest is tracked through the sequence using region based matching. Secondly, degradation of images is modelled in terms of pixel sampling, defocus blur and motion blur. Motion blur direction and magnitude are estimated from tracked displacements. Finally, a high-resolution deblurred image is reconstructed. The approach is illustrated with video sequences of moving people and blurred script.

1 Introduction

Real images of a moving object can each be regarded as a degraded representation of the ideal image that would have been captured at a certain instant by an ideal camera. These degradations include: (i) optical blur (ii) image sampling by the CCD array (iii) motion blur. Often changing the cameras to improve the quality of the images is not an option, so post-processing is needed to restore the images. The aim of this paper is to recover from a sequence of images a higher resolution deblurred image that is as close as possible to an ideal image, by removing the blur due to the real image formation process. Firstly, the object is tracked using area-based deformable regions. Secondly, given an initial estimate of the ideal image, the physical image formation process is simulated. Finally the ideal image is estimated recursively by minimising the difference between the real images and the simulated ones. The originality of our approach is in: (i) avoiding explicitly calculating an inverse filter of the blurring process, which is ill-conditioned for a single image (ii) addressing the problem of removing motion blur for non-purely translational motions and using an image sequence rather than a single image, contrary to previous approaches (iii) studying the problem of removing a combination of motion blur and optical blur, which have only been studied separately before.

The paper is organised as follows: firstly a general image formation model is presented. Secondly, we review related approaches in the literature. Thirdly, the approach proposed by this paper is described. Lastly, experimental results on real images are shown.

2 Model of the image formation process

This section introduces the model of image formation used in this paper. It will be used to recover a high-resolution unblurred image from an image sequence.

Let I^{ideal} be the "ideal" image that a perfect pinhole camera would produce at time $t = 0$. This is the image that we'll try to estimate as closely as possible. Three types of distortions occurring during the image formation process and leading from the ideal image to the observed image sequence are considered: optical blur, motion blur and spatial sampling (see fig. 1). They can be described as follows:

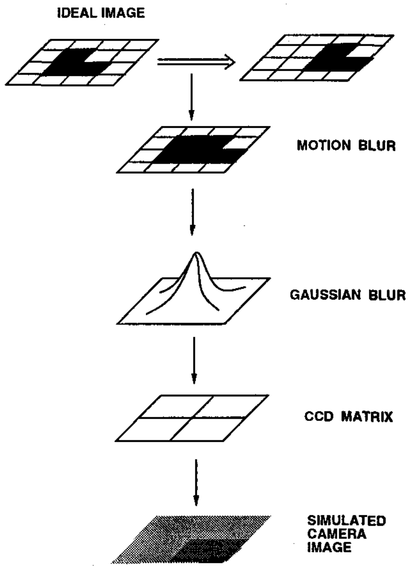


Fig. 1. Model of the image formation process used for deblurring. The ideal image has arbitrarily high resolution. First, since the object is moving, it is integrated over time by the camera, so that the resulting image is degraded by motion blur. This blur cannot be treated as purely translational and is modelled as affine in our approach. Second, given that the camera may be defocussed, and other image formation process defects, spatial blur is introduced. It is modelled to a first approximation as Gaussian. Third the CCD matrix limits the resolution of the image and performs averaging over pixel blocks of the ideal image. This model allows us to simulate the real images.

MOTION A 3D motion of the object induces a 2D motion in the image, so that the ideal image I_t^{ideal} that would be observed at time t is a 2D transform of the ideal image I^{ideal} at $t = 0$. In the general case, this image transform requires a 3D model [3]. In more restricted cases, this transform can be written as a pure 2D transform. Here we will consider a 2D affine motion model, though the approach can be easily applied to other pure 2D transforms such as homographic. Let $M(b)$ be the motion model with parameters b chosen to describe the image transform corresponding to the object 3D motion. Then:

$$I_t^{\text{ideal}}(x,y) = I_0^{\text{ideal}}(u_t,v_t) \quad \text{with:} \quad \begin{pmatrix} x \\ y \end{pmatrix} = M(b)_0^t \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

MOTION BLUR The transformed versions I_t^{ideal} of the ideal image I^{ideal} are integrated by the camera during its integration time T. If the object is moving significantly during this time, the resulting image I_t^{Mot} is smeared by motion blur. This effect is particularly noticeable when the camera is not equipped with an electronic shutter, as often in surveillance applications. The motion-blurred image I_t^{Mot} can be written as follows:

$$I_t^{Mot}(x,y) = \int_{\tau=t-T}^{\tau=t} I_{t-T}^{\text{ideal}}(u_\tau,v_\tau)\,d\tau \quad \text{with} \quad \begin{pmatrix} x \\ y \end{pmatrix} = M(b)_{t-T}^{\tau} \begin{bmatrix} u_\tau \\ v_\tau \end{bmatrix} \qquad (1)$$

where $M(b)_{t-T}^{\tau}$ is the estimated motion of the object between time $t-T$ and t - with parameter values b - using the chosen motion model.

OPTICAL BLUR In addition to motion blur, the physics of image formation by a camera create other blurring effects, that we will refer to as "optical blur", and that affect as much static objects as moving ones. It groups several effects: (i) the blur introduced by the optics and thick lenses of the camera. Radial distortions in the periphery of the image are not considered here (ii) the blur induced by the camera being out-of-focus. To a first approximation, the "optical blur" will be modelled by 2D Gaussian blur.

SPATIAL SAMPLING The image is then sampled by the camera CCD array. The characteristics of this array determine the image resolution and the amount of averaging occurring when the signal is integrated by each cell of the CCD array. This sampling effect is modelled by averaging of the image over blocks of pixels. The CCD cells' point spread function is supposed to be already modelled by Gaussian blur (see previous paragraph).

Given the motion-blurred image I_t^{Mot}, the image S_t resulting after optical blur and spatial sampling can be written as:

$$S_t = \uparrow [S_{CCD} * G_y(x,y) * G_x(x,y) * I_t^{Mot}(x,y)]$$

where G_x and G_y are 1D Gaussian kernels along respectively the x and y directions. These two kernels apply 2D Gaussian blur to the image (in a separable way). S_{CCD} represents the averaging over pixel blocks introduced by the CCD matrix. This image is further sampled by the operator $\uparrow$ so that only 1 pixel out of the block of averaged pixels is retained. S_t simulates the low-resolution and blurred image that should be observed by the camera at time t.

3 Review of related approaches

The aim of this paper is to retrieve the "ideal" image of an object, of which real cameras give only a degraded version, using a sequence of images. The previous section describes our model of the image formation process, including both optical blur and motion blur. This section presents related approaches in the literature. To our knowledge, the removal of the combined effects of optical and motion blur has never been done before. However, these effects have been studied separately.

Inverse filtering of optical blur and super-resolution: The approaches aiming at removing the effects of "optical blur" and subsampling are called "super-resolution methods". They do not consider motion blur. Tsai and Huang [6] solved the problem in the frequency domain, disregarding blurring and assuming inter-frame translations. It is difficult to generalise to non-translational motions. Gross [5] merged the low resolution images by interpolation and obtained the high-resolution image by deblurring this merged image. He also assumed translational motion. Peleg and Keren [9] simulate the imaging process, and optimise the high-resolution image so as to minimise the difference between the observed and simulated low-resolution images. However the minimisation method is relatively simple: each pixel is examined in turn, and its value is incremented by 1, kept constant, or decreased by 1, so that the global criterion is decreased. Irani and Peleg [7, 8] minimise the same difference between observed and simulated images, using a back-projection method similar to that used in Computer Aided Tomography. However, the use of a back-projection operator limits the method to blurring processes for which such an operator can be calculated or approximated. Zevin and Werman consider directly a 3D model of the world with perspective cameras. Berthod *et al* [2] improved this method by considering reflectance models and their method is partly based on height from shading.

Inverse filtering of motion blur: The inverse filtering of motion blur has also been addressed in the literature, separately from the super-resolution problems. It has mainly been studied for translational motions and for single images. The approaches can be decomposed into two main groups, depending whether filtering is done in the spatial domain (Sondhi's filter [10]) or frequency space (inverse Wiener filtering [4]). Both these types of approaches are difficult to generalise to more complicated motions than purely translational.

4 Motion deblurring and Super-resolution from an image sequence

In this section, we describe our work in detail. This extends the work of [8] who showed for the case of optical blur that, though restoring degraded images is an ill-conditioned problem, the use of a sequence of images to accumulate information about the object can help to partly overcome this indeterminacy. Here we also consider distortions introduced by motion blur. Motion blur is known to be a particularly ill-conditioned blur and has previously been studied for purely translational motions and single images only (see [4]).

Object tracking approach: First the object is tracked through the sequence of images using an approach combining area-based and contour-based deformable models [1]. The tracking approach can be described as follows: (i) First the region is tracked by a deformable region based on texture correlation and constrained by the use of an affine motion model. The use of texture correlation ensures the robustness of tracking for textured images, and is also more reliable than deformable contours for blurred images (ii) Then the region contour is refined by a deformable contour. Thus the detection of the region edges is more precise. It also helps to correct tracking errors made by the deformable region in the case of occlusions and specularities. This refinement of the region contour is very useful if the image texture is poor. But it must be turned off in case of major occlusions or if too much blur renders the detection of edges unreliable.

This tracking approach gives a reliable and sub-pixel segmentation of a moving object. Such precision is necessary to recover a higher resolution image from a sequence of images. It also gives an estimation of the region 2D apparent motion. This motion estimation is used to register the images of the sequence back to the first frame, and to determine motion blur. Let $M(b)_{t-T}^{T} = (A, B)$ be the affine motion measured between the two images taken at time $t - T$ and T. It is supposed that to a first approximation the motion $M(b)_{t-T}^{\tau}$ varies linearly between the two images. Then, according to eq. 1, the image blurred by the motion (A, B) can be written as:

$$I_t^{Mot}(x, y) = \int_{\tau=t-T}^{\tau=t} I_{t-T}^{\text{ideal}}(u_\tau, v_\tau)\, d\tau$$

with

$$\begin{pmatrix} x \\ y \end{pmatrix} = A(\tau) \begin{pmatrix} u_\tau \\ v_\tau \end{pmatrix} + B(\tau)$$

and

$$A(\tau) = (A - I)\ \tau + I \quad , \quad B(\tau) = B\ \tau$$

Note that, once the region 2D motion is estimated, the transform between an image and its motion-blurred version is linear (but not spatially invariant). Thus it can be written in matrix form as: $I^{Mot} = \mathcal{B}\ I^{\text{ideal}}$.

Deblurring approach: Once the moving object is tracked and its motion estimated, the information gathered from the multiple images can be merged to enhance the images of the object, as shown by Irani and Peleg [8]. To this end, we use the image formation process described earlier. A high-resolution deblurred image of the object is iteratively estimated by minimising the difference between the real observed images and the corresponding images that are predicted by applying the modelled image formation process to the current estimation of the high-resolution image (see fig. 2). Our optimisation approach differs from [7] in that it doesn't use an explicit inverse filter - which can be difficult to approximate - of the image degradations in the feedback loop.

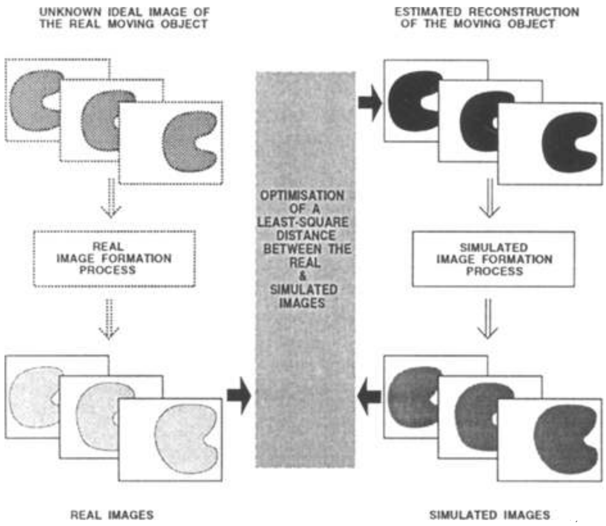


Fig. 2. Deblurring and super-resolution from image sequences. Given an initial estimate of the ideal image of the object, and having estimated the object's 2D motion by tracking, a prediction of the observed image sequence can be constructed by simulation of the image formation process. The optimisation of the reconstructed ideal image is then performed by minimising the difference between this predicted and the actual image sequence. This is done by minimisation of the corresponding least-square criterion by conjugate gradient descent. A regularisation term is also added, which ensures that the reconstructed image is not dominated by periodic noise, as can happen if the deblurring of motion blur is not done carefully. The advantage of this minimisation approach is that it does not require the construction of inverse filters for the different image formation and degradation processes.

Cost function The criterion to optimise measures the discrepancy between the real observed images R_i and the images S_i simulated by applying our image formation model (described in section 2 and eq. 4) to the estimated "ideal" image I:

$$\begin{aligned} E = & \textstyle\sum_{i=0}^{N} \sum_{(x,y)\in R_i} [R_i(x,y) - S_i(x,y)]^2 \\ & +\lambda . \textstyle\sum_{(x,y)\in I} [2 * I(x,y) - I(x+1,y) - I(x-1,y)]^2 \\ & +\lambda . \textstyle\sum_{(x,y)\in I} [2 * I(x,y) - I(x,y+1) - I(x,y-1)]^2 \\ & \text{where} \quad S_i = \uparrow [S_{CCD} * G_y(x,y) * G_x(x,y) * \mathcal{B}_i . I = \mathcal{D}_i I] \end{aligned}$$

where N is the number of images in the sequence. The second term adds second-order smoothness constraints on the reconstructed image I. These constraints improve the robustness of the method to noise. It also improves the conditioning and stability of the minimisation, which can be ill-conditioned, especially because of motion blur.

Optimisation Minimisation is performed by a multidimensional conjugate gradient method. Rewriting E, its gradient can be written as:

$$E = \sum_{i=0}^{N}[R_i - \mathcal{D}_i I]^2 + \lambda . IHI \quad \text{and} \quad \nabla = \sum_{i=0}^{N} 2[(\mathcal{D}_i^T . \mathcal{D}_i - \mathcal{D}_i R_i) . I] + 2\lambda H . I$$

where H is a matrix containing the smoothness constraints.

5 Experimental results

This section illustrates our approach to achieve super-resolution and deblurring of a moving object using a sequence of images, considering both optical blur and motion blur. An affine motion model is considered. Double resolution of the reconstructed image with respect to the real images is achieved. The deblurring algorithm has been applied only to the region of interest in the image, and the results are shown as blow-ups of these regions.

The first example is related to surveillance applications. A person is running in front of the camera, inducing motion blur in the image (see fig. 3). The person's face is tracked through four images and its 2D affine motion is estimated (see fig. 3). These multiple regions and motion estimations are then used to perform the inverse filtering of the motion blur on the face and to achieve double resolution for it. The results (see fig. 4) show a significant improvement of the level of details visible on the face (observe the ear and cheeks) and greatly reduce the smearing effect induced by motion blur. It can be noted that, when motion blur occurs, averaging the image sequence after registration (which is a standard method to increase resolution) gives an even more smeared image.

The second example (see fig. 5) shows the removal of blur from script images. The original images (a,b, detail A) are low-resolution and motion blurred. Our method (B) produces a deblurred and double resolution image of the label "pasta", using the ten images of the sequence. The label is visibly improved.

6 Conclusion

This paper has presented a new method for motion deblurring, focus deblurring and super-resolution from image sequences. Previous research has addressed copiously the problem of recovery from these individual degradations, but here it has proved crucial to address the three degradations above simultaneously. In particular, motion blur plays a key role in image degradation. An affine motion blur is considered, contrary to standard studies limited to purely translational motions. The motion blur is not assumed to be known; indeed it is actually derived from the estimated object motion. The methodology has been illustrated by experimental results. In the future, it is proposed to extend modelling to 3D motion, necessary for larger baselines with non-planar objects. Another area of future research concerns more ambitious applications of the reconstruction method. One that is particularly appealing is to reconstruct not merely a single frame but an entire sequence, by reintegrating the estimated "ideal" image over the recovered affine motion field. This would allow, for instance, the regeneration of slow-motion "action replay" sequences undegraded by excessive motion blur.

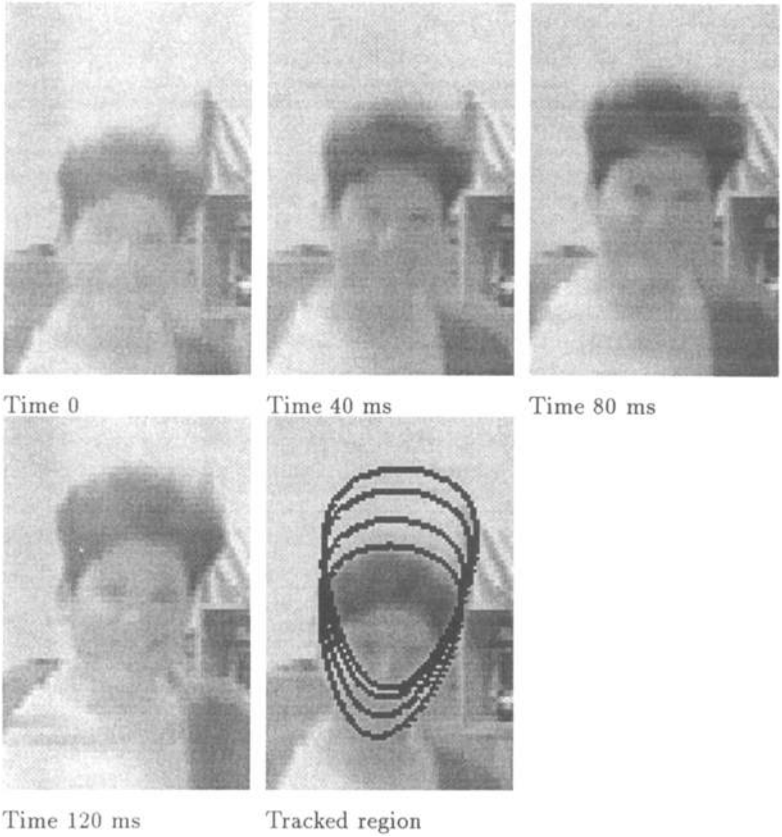


Fig. 3. Deblurring motion blur and super-resolution from an image sequence. Initial low-resolution image sequence, taken at video rate. Only even fields are retained. Because the camera is not equipped with an electronic shutter, motion blur in the image is significant. Here the face is tracked by an area-based deformable region with an affine motion model. As can be seen from the trajectories, here the motion is really affine and not purely translational.

References

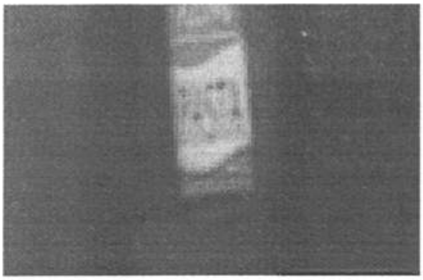
1. B. Bascle and R. Deriche. Region tracking through image sequences. *Proceedings of the 5th International Conference on Computer Vision (ICCV'95), Boston, USA*, June 1995.
2. M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia. Reconstruction of High Resolution 3D Visual Information. Rapport de Recherche 2142, INRIA, Dec 1993.
3. Faugeras, O. *Three-dimensional computer vision: a geometric viewpoint.* MIT Press, 1993.
4. R.C. Gonzalez and R.E. Woods. *Digital Image Processing.* Addison Wesley, 1993.

a. 4th image

b. Deblurred image

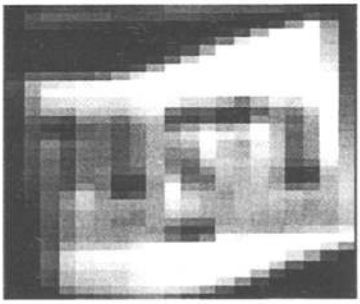
Fig. 4. Deblurring motion blur and super-resolution from an image sequence. The original images (a) are low-resolution and motion blurred. Our method (b) produces a deblurred and super-resolution (here double) image of the face, using the four images of the sequence. Details such as the nose and ear are retrieved.

5. D. Gross. Super-resolution from Sub-pixel Shifted Pictures. Master's thesis, Tel-Aviv University, Oct 1986.
6. T. Huang and R. Tsai. *Advances in Computer Vision and Image Processing*, volume 1, chapter Multiframe Image Restoration and Registration, pages 317–339. JAI Press Inc, 1984.
7. M. Irani and S. Peleg. Improving Resolution by Image Registration. *CVGIP:GMIP*, 53(3):231–239, May 1991.
8. M. Irani and S. Peleg. Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency. *Journal of Visual Communication and Image Representation*, 4(4):324–335, December 1993.
9. D. Keren, S. Peleg, and R. Brada. Image Sequence Enhancement Using Sub-pixel Displacement. In *Proc. of International Conference on Computer Vision and Pattern Recognition (CVPR'88)*, pages 742–746, Ann Arbor, Michigan, June 1988.
10. M.M. Sondhi. Image Restoration: The Removal of Spatially Invariant Degradations. *Proc. IEEE*, 60(7):842–853, 1972.



a. Image 1

b. Image 10



A. Detail of Image 1

B. Deblurred detail

Fig. 5. Deblurring motion blur and super-resolution from an image sequence. The original images (a,b,A) are very low-resolution and motion blurred, so that it is difficult to extract significant edges from them (I). Our method (B) produces a deblurred and double resolution image of the label "pasta", using the 10 images of the sequence. The label is now more easily readable than before