VI. Discussion

In this correspondence, we have proposed a new approach to solving nonlinear estimation problems, exemplified by the problem of filtering gamma-corrected noisy video signals. An adaptive OS estimator was formulated of which the weights depend only on the HOOS of the noise, and of which the data window used depends on the RT. Essential in the derivation is that the nonlinear signal model is approximated by a Taylor expansion of order M. This order can be chosen freely and does not influence the computational complexity of the resulting estimator once the weights of the estimator have been determined. The filtering results that we obtained are nearly independent on the value of γ, which means that the proposed filtering approach is quite successful in handling the signal-dependent noise resulting from the nonlinear model.

References

[1] G. Cortelazzo, G. A. Mian, and R. Parolari, “Statistical characterization of granular camera noise,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 536–543, 1994.

[2] R. P. Kleihorst, R. L. Lagendijk, and J. Biemond, “An efficient spatiotemporal filter for gamma-corrected video signals,” in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, 1994.

[3] R. P. Kleihorst, “Noise-filtering of image sequences,” Ph.D. dissertation, Inform. Theory Group, Delft. Univ. Technol., Delft, The Netherlands, 1994.

[4] A. C. Bovik, T. S. Huang, and D. C. Munson, “A generalization of median filtering using linear combinations of order statistics,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1326–1337, 1983.

[5] I. Pitas and A. N. Venetsanopoulos, “Non-linear mean filters in image processing,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 573–584, 1986.

[6] G. R. Arce and R. E. Foster, “Detail-preserving ranked-order based filters for image processing,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 83–98, 1989.

[7] R. P. Kleihorst, R. L. Lagendijk, and J. Biemond, “Higher-order order-statistic filters for signal dependent noise,” in *Proc. 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, Halkidiki, Greece, 1995, pp. 98–101.

[8] A. van der Ziel, *Noise*. Englewood Cliffs, NJ: Prentice-Hall, 1954.

[9] H. A. David, *Order Statistics*, 2nd ed. New York: Wiley, 1981.

[10] G. R. Arce, “Multistage order statistics filters for image sequence processing,” *IEEE Trans. Signal Processing*, vol. 39, pp. 1147–1163, 1991.

[11] M. K. Özkan, M. I. Sezan, and A. M. Tekalp, “Adaptive motion-compensated filtering of noisy image sequences,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 277–290, 1993.

[12] D. B. Duncan, “Multiple ranges and multiple f-tests,” *Biometrics*, vol. 11, pp. 1–42, 1952.

[13] A. E. Sarhan and B. G. Greenberg, Eds., *Contributions to Order Statistics*. New York: Wiley, 1962.

[14] G. de Haan, P. W. A. C Biezen, H. Huijgen, and A. O. Ojo, “True-motion estimation with 3-D recursive search block-matcher,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 368–379, 1993.

Robust, Object-Based High-Resolution Image Reconstruction from Low-Resolution Video

P. Erhan Eren, M. Ibrahim Sezan, and A. Murat Tekalp

***Abstract*— We propose a robust, object-based approach to high-resolution image reconstruction from video using the *projections onto convex sets* (POCS) framework. The proposed method employs a validity map and/or a segmentation map. The validity map disables projections based on observations with inaccurate motion information for robust reconstruction in the presence of motion estimation errors; while the segmentation map enables object-based processing where more accurate motion models can be utilized to improve the quality of the reconstructed image. Procedures for the computation of the validity map and segmentation map are presented. Experimental results demonstrate the improvement in image quality that can be achieved by the proposed methods.**

***Index Terms*—High-resolution reconstruction, object-based reconstruction, POCS, segmentation, tracking.**

I. Introduction

Recently, there has been much interest in reconstruction of high-resolution and/or expanded-view images from a low-resolution video source, with applications to printing high-quality stills from video and high-resolution standards conversion [1]–[8]. Here, we focus on the *projections onto convex sets* (POCS) method proposed by Patti *et al.* [5], reviewed in Section II, because it

1) is an image domain method that offers the flexibility of space-varying (pixel-by-pixel) processing, hence, allows use of validity and segmentation maps;
2) simultaneously accounts for blurring due to relative sensor-object motion and sensor integration;
3) is capable of handling low-resolution sampling lattices with arbitrary periodic geometry (e.g., an interlaced lattice).

However, it has only been applied to scenes containing a single moving object where motion vector field can be reliably estimated on the basis of a single motion model [5]. In case of scenes containing multiple objects, the estimated motion vector field is likely to be inaccurate around motion boundaries and occlusion regions, which results in objectionable artifacts. A previous work on object-based resolution improvement in the presence of multiple motion, by Irani and Peleg [2], proposed tracking of an object with dominant motion over several frames to improve its resolution using a backprojection algorithm, where occluded pixels are filled in by temporal averaging. However, they do not address resolution improvement over pixels

Manuscript received August 6, 1996; revised December 3, 1996. This work was supported in part by a National Science Foundation IUCRC grant and a New York State Science and Technology Foundation grant to the Center for Electronic Imaging Systems, University of Rochester, and by a grant from Eastman Kodak Company. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark R. Banham.

P. E. Eren and A. M. Tekalp are with the Department of Electrical Engineering and Center for Electronic Imaging Systems, University of Rochester, Rochester, NY 14627 USA (e-mail: tekalp@ee.rochester.edu).

M. I. Sezan is with the Digital Video Department, Sharp Laboratories of America, Camas, WA 98607 USA.

Publisher Item Identifier S 1057-7149(97)07121-2.

that are occluded in some of the frames, and they do not account for motion blurring.

We propose two extensions of the Patti *et al.* method [5] to provide improved reconstructions in the case of scenes with multiple moving objects. These extensions address artifacts resulting from inaccuracies in multiple motion estimation, which are related to the well-known problems of how to deal with covered/uncovered pixels effectively, and estimating reliable motion vectors in the vicinity of such occlusion regions. The first proposed extension allows robust reconstruction in the presence of motion estimation errors, where pixels are classified as those with reliable versus unreliable motion vectors specified by a validity map. The computation of the validity map is discussed in Section III. The second approach, presented in Section IV, is an object-based extension, where objects with distinct motions are specified by a segmentation map. In this paper, we assumed that objects of interest are marked on a reference low-resolution frame interactively, which are then automatically tracked on all available frames using a mesh-based object-tracking method [9]. Object-based processing allows use of different parametric/nonparametric motion models within each object, which should lead to more accurate motion estimation and improved overall reconstruction quality. In case utilities for interactive segmentation and/or tracking are not available, the method in Section III provides robust reconstruction by excluding observations with erroneous motion vectors; otherwise, methods in Sections III and IV can be combined for robust, object-based reconstruction.

Hence, the main contribution of this work is to exploit the space-varying (pixel-by-pixel processing) nature of the POCS framework to perform selective projections using only relevant observations as determined by the validity and segmentation maps. We also employ the same principle for high-resolution reconstruction of pixels that are covered/uncovered in some of the frames. Results showing the benefits of the proposed extensions are furnished in Section V.

II. Overview of the POCS Reconstruction

We now summarize the POCS-based high-resolution reconstruction discussed in [5]. We denote the low resolution image sequence by $g(m_1, m_2, k)$, and assume that an estimate of the high resolution image at time $k = t_r$ is sought. We define a family of closed, convex constraint sets, one for each pixel within the low-resolution image sequence [5]

$$C_{t_r}(m_1, m_2, k) = \{y(n_1, n_2, t_r)\colon |r^{(y)}(m_1, m_2, k)| \le \delta_0\}$$

where

$$r^{(y)}(m_1, m_2, k) \doteq g(m_1, m_2, k) - \sum_{(n_1, n_2)} y(n_1, n_2, t_r) h_{t_r}(n_1, n_2; m_1, m_2, k)$$

is the residual associated with an arbitrary member, y, of the constraint set. The quantity h_{t_r} denotes the combined blur point spread function (psf) accounting for sensor integration and relative sensor-object motion. It is computed using the estimated relative motion vector field between the reference image and the kth low-resolution image. The quantity δ_0 is an *a priori* bound reflecting the statistical confidence with which the actual image, f, is a member of the set $C_{t_r}(m_1, m_2, k)$. This family of constraints is referred to as data consistency constraints. An estimate of the high-resolution version of the reference image is determined iteratively starting from an arbitrary initialization. Each iteration is obtained by successive projections of the previous estimate onto the consistency sets as well as the amplitude constraint set that restricts the gray levels of the estimate to [0, 255].

III. Robust High-Resolution Reconstruction Using a Validity Map

The POCS framework has been successfully applied in [5] to the case where the sequence contains a single object whose motion can be accurately modeled and estimated using a single parametric model. In case of multiple objects, one can either use a nonparametric, dense optical flow estimation [10], or estimate multiple parametric models [11]. In either case, however, it is highly likely to have inaccurate motion vectors at moving object boundaries and occlusion areas. For instance, if the observed low-resolution pixel belongs to an uncovered region, absent in the reference image, or if the motion vector estimated for a pixel is inaccurate, the spatial relationship established between that low-resolution pixel and the current estimate of the high-resolution image will, indeed, be in error. In fact, the blur psf itself will be erroneous, since it is based on incorrect motion information. Use of such inaccurate information degrades the quality of the POCS reconstruction, as demonstrated in Section V. It is therefore rather important to discard inaccurate motion estimates during the reconstruction process. This is easily achieved within the POCS framework by defining the data consistency constraint sets and performing associated projections only for those pixel locations, (m_1, m_2, k), for which the motion vectors are accurate. Toward this end, we determine a validity map for each low-resolution image, specifying pixel locations with accurate motion estimates.

The major components of determining a validity map on the basis of motion estimation accuracy are 1) computing a two-dimensional (2-D) accuracy detection signal $d(m_1, m_2, k)$; 2) generating a space-varying threshold $\theta(m_1, m_2, k)$; and 3) thresholding the detection array to generate a binary map, which is then morphologically filtered to form the validity map $V(m_1, m_2, k)$. The accuracy detection signal is formed by the summed absolute differences (SAD), computed over a 3 × 3 window of pixels, between motion-compensated reference image, $G^{MCi}(\ell_1, \ell_2, t_r)$, and other low-resolution images, $G(\ell_1, \ell_2, k)$, for $k = t_r + i$ where i is an integer, as follows:

$$d(m_1, m_2, t_r + i) = \sum_{p=-1}^{1} \sum_{q=-1}^{1} |G(m_1 + p, m_2 + q, t_r + i) - G^{MCi}(m_1, m_2, t_r)|.$$

The accuracy detection signal is then compared with the space-varying threshold $\theta(m_1, m_2, k)$, determined on the basis of following observations. First, errors in motion estimation accuracy detection result in objectionable degradations in image regions with low spatial frequency content. It is therefore desirable to have a low threshold value for such image regions. On the other hand, it is in high-detail regions that high-resolution image reconstruction will produce the greatest increase in resolution. Therefore, the threshold should be high for those regions. This method of generating a spatially varying threshold is similar to that used in [12] in the context of dominant-motion-compensated deinterlacing.

IV. Object-Based High-Resolution Reconstruction

The selective projection formalism can be extended to object-based resolution improvement by including a region (segmentation) map, which corresponds to the support of the object(s) of interest at that image. The object-based high-resolution may correspond to processing of a selected image object, or object-by-object (layer-by-layer) processing of the entire image. This framework allows parametric/nonparametric modeling of the motion field within each

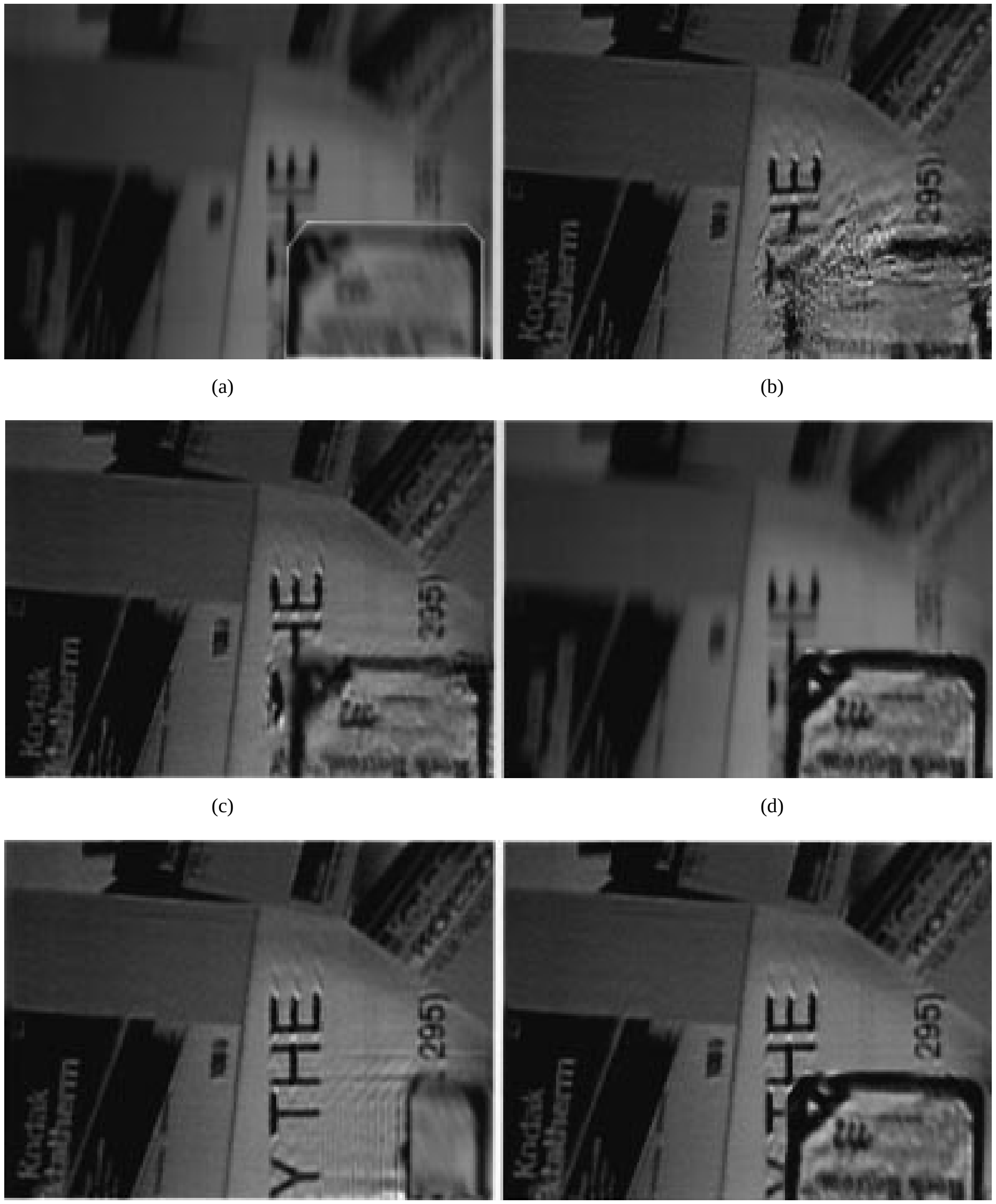


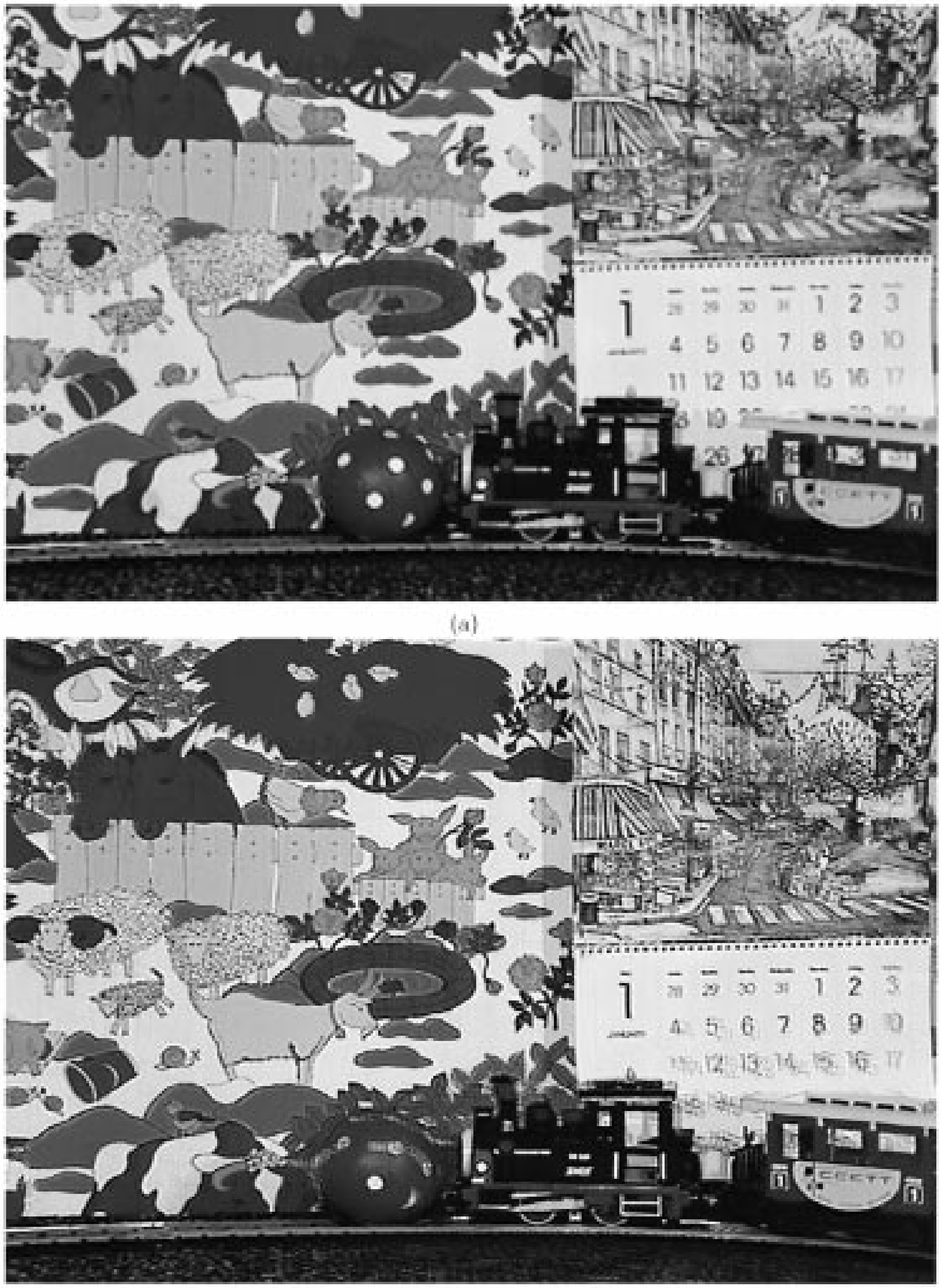
Fig. 1. (a) Reference frame vertically interpolated from field 1. (b) Reference frame reconstructed without using a validity map. (c) Reference frame reconstructed using a validity map. (d) Reconstructed foreground object in the reference frame. (e) Reconstructed background object and the uncovered background regions. (f) Reference frame reconstructed using different motion models for the background and foreground objects.

object for improved motion estimation accuracy. It also enables us to better deal with pixels that get covered/uncovered within the available frames.

A. Object-Based Motion Estimation

The object support information may be available in two forms: 1) segmentation maps of each frame are given or can be computed automatically [13], and 2) the segmentation map of a reference (key) frame is given, and then objects of interest are automatically tracked over the available sequence of images using a tracking algorithm. In the former case, pixel-by-pixel correspondence between objects in successive frames may be computed by dense motion (optic flow) or parametric motion estimation techniques. Object-based dense motion estimation may be carried out as described in the Motion Pictures Expert Group-4 (MPEG-4) verification model (VM) [14]. In the latter case, a two-dimensional (2-D) mesh-based object-tracking algorithm, which implies piecewise parametric motion estimation within each object, may be used [9].

In this paper, we marked the object of interest in a key frame interactively. This object was then tracked over the rest of the frames automatically [9]. The motion of the background layer was estimated between each low-resolution frame and the reference frame using a global affine motion model as in [15]. Furthermore, a validity map is also employed within each object to account for errors in motion tracking. POCS are engaged only for those pixels within the object of interest with accurate motion estimates.



(a)

(b)

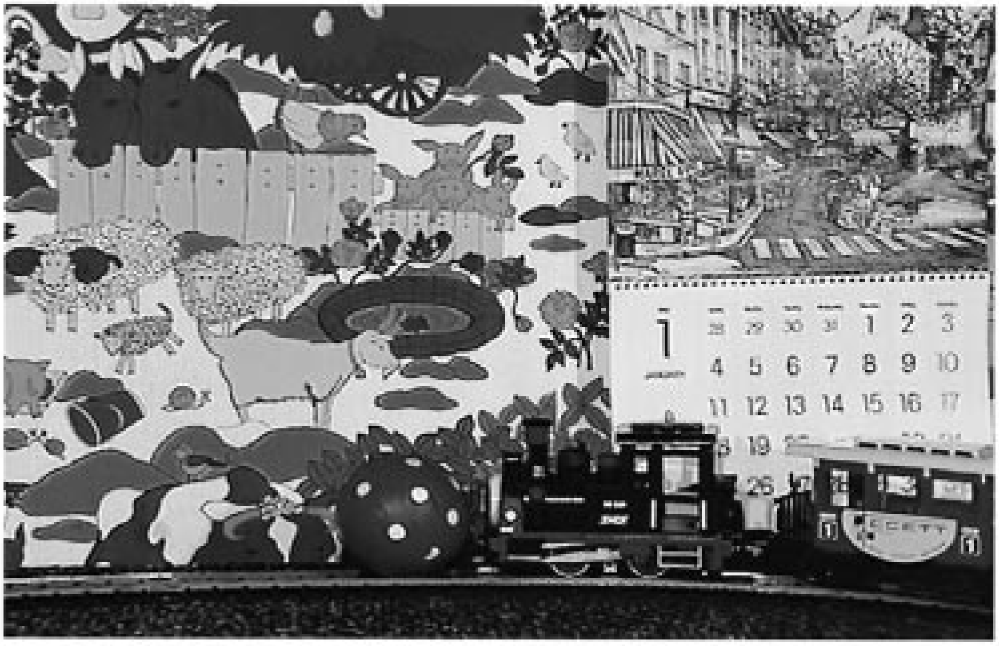
Fig. 2. (a) Reference frame bilinearly interpolated. (b) Reference frame reconstructed without using a validity map.

B. Reconstruction of Partially Covered Pixels

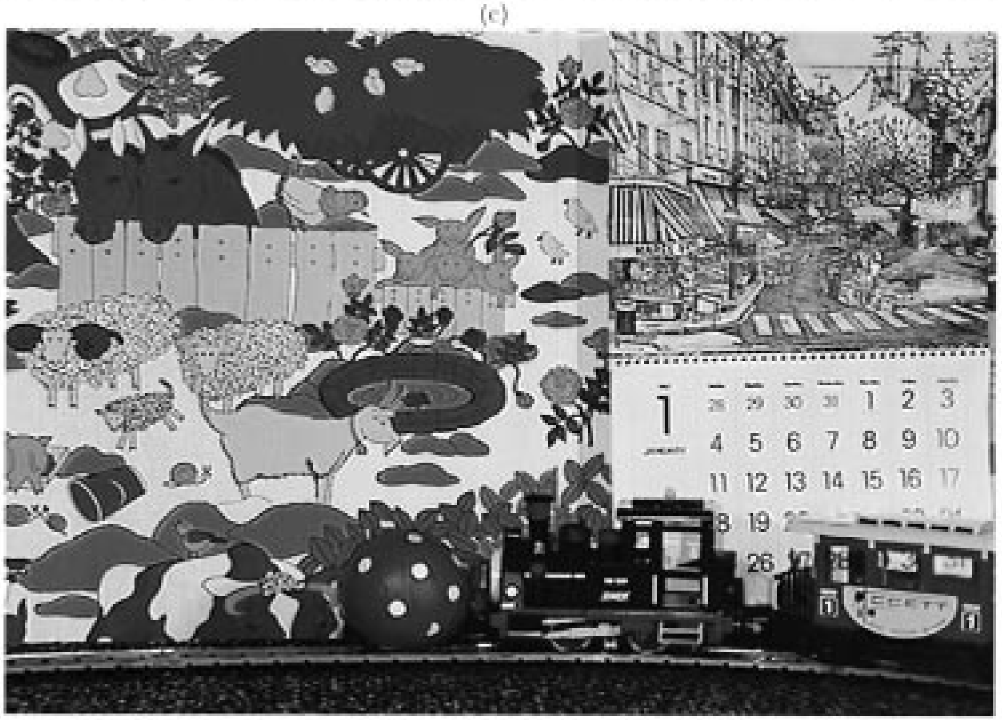
When the object of interest admits a single parametric motion model, parts of the object that may be covered in the reference frame can also be reconstructed by projections of their uncovered observations in subsequent frames. For example, consider an object moving against a background that exhibits a global transformation due to camera motion. Those background pixels that are behind the object in the reference frame can also be reconstructed by projections of their observations in other frames where they are visible. This process, which is demonstrated in Fig. 1(a) and (e), may be considered as high-resolution object mosaicking.

V. Results

The proposed methods are tested on two sequences: 1) nine-fields of an interlaced sequence that we acquired by a Sony Hi-8 Camcorder, where each field is 640 pixels $\times$ 240 lines, and 2) eight frames (frames 136–143) of the MPEG test sequence "mobile and calendar," which is 360 pixels $\times$ 240 lines and progressive. The first scene contains



(c)



(d)

Fig. 2. *(Continued.)* (c) Reference frame reconstructed using a validity map. (d) Object-based reconstruction of the reference frame using the validity map.

significant camera motion (panning) as well as independent motion of a foreground object. The background is a flat poster mounted on the wall. The foreground object is a rectangular flat sheet with some text on it [marked in Fig. 1(a)]. It is moved by hand horizontally rightward, in front of the background while the camera is panning in the vertical direction. As the foreground object moves, it uncovers a region of the background. The mobile and calendar sequence contains a translating toy train, a rolling ball, a calendar that moves up and down, and a moving background due to camera pan. Frame 136 is shown in Fig. 2.

The first experiment addresses high-resolution deinterlacing of the first field of the first sequence, accounting for motion, sensor, and optical blurs, as outlined in [5]. Fig. 1(a) shows the low-resolution reference (first) field, after vertical (linear) interpolation to frame format. The visible degradation is mainly due to motion blur caused by the camera and foreground object motions. This image is used to initialize the POCS algorithm in each of the following cases. In the first two cases, we obtain a high-resolution frame without using object segmentation information. All observed fields are first linearly interpolated (in the vertical direction) to frame format, and a dense motion field is estimated between each low-resolution frame and the reference frame by using the Fogel algorithm [10]. Fig. 1(b) shows the reconstructed image obtained without using a validity map. The background is significantly enhanced, and the letters are readable. However, there are severe artifacts that are due to erroneous projections in the vicinity of occlusion regions (i.e., pixels with inaccurate motion vectors). Fig. 1(c) shows the result obtained using a validity map for each low-resolution field. Note the improvement in the quality as a result of detecting inaccurate motion estimates and not using them in the reconstruction process. The next set of results demonstrates object-based deinterlacing. First, the foreground object is selected as the object of interest. The foreground object is tracked throughout the sequence using the method discussed in [9]. The reconstructed reference frame where only the foreground object is enhanced is shown in Fig. 1(d). In the next case, the background motion is estimated between each observed frame and the reference frame by an affine-model-based iterative method discussed in [15]. These motion estimates combined with the foreground tracking information are then used to reconstruct the background object in the reference frame together with the part of the background that is covered in the reference frame but gets uncovered in frames 2–9. This result is shown in Fig. 1(e). In these experiments, the high-resolution reconstruction process is performed over those pixels with "valid" motion estimates within the object-of-interest. Fig. 1(f) shows the high-resolution reconstruction of the reference frame that is obtained by using a different motion model separately for background and foreground objects. Note the improvement in quality due to more accurate motion modeling and estimation compared to the result shown in Fig. 1(c).

The second experiment demonstrates high-resolution interpolation of frame 136 by a factor of two in both directions. Fig. 2(a) shows the result of bilinear interpolation, which is used to initialize the POCS reconstruction in the following cases. The result of POCS reconstruction using a dense motion field estimated by the Fogel algorithm is shown in Fig. 2(b). Fig. 2(c) shows the result of POCS reconstruction using the validity map, which suppresses projections on errenous motion vectors. Finally, Fig. 2(d) shows the result of object-based high-resolution algorithm, where the object boundaries are manually marked on frame 136. The motion estimation is then performed on each object separately using the mesh-based tracking algorithm [9]. Inspection of the results indicate that Fig. 2(b) suffers from artifacts at the object boundaries, which is especially visible around the numbers on the calendar. Reconstruction using the validity map removes most of these artifacts at the expense of some blurring in these regions, where the object-based method provides excellent sharpness free of visible artifacts.

VI. Conclusion

This work provides two approaches for high-resolution image reconstruction in the presence of multiple motion. In the first approach, which does not require object boundary information, a dense motion field has been estimated; observations with unrealiable motion have been identified; and projections based on such observations have been turned off. The second approach, which requires segmentation of each frame into objects of interest (or segmentation of a key frame and then tracking of the object boundaries), estimates parametric (or dense) motion within each object individually for improved motion estimation accuracy. Projections based on observations with inaccurate motion vectors can still be turned off by using a combination of a segmentation and a validity map. Experimental results demonstrate the feasibility of the proposed region-based methods.

Acknowledgment

The authors thank Dr. A. Patti of Hewlett-Packard Laboratoris, Dr. J. Riek and Dr. S. Fogel of Eastman Kodak Company, and Mr. C. Toklu for their helpful discussions and contributions to the software used in this work.

References

[1] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing*, vol. 2, pp. 534–539, Oct. 1993.

[2] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. Visual Commun. Image Represent.*, vol. 4, pp. 324–335, Dec. 1993.

[3] S. Mann and R. W. Picard, "Virtual bellows: constructing high quality stills from video," in *Proc. IEEE Int. Conf. on Image Processing*, Austin, TX, Nov. 13–16 1994.

[4] Y. Nakazawa, T. Komatsu, and T. Saito, "High-resolution image acquisition based on temporal integration with hierarchical estimation of image warping," in *Proc. IEEE Int. Conf. on Image Processing*, Washington, DC, 1995, vol 3, pp. 244–247.

[5] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing,* vol. 6, pp. 1064–1076, Aug. 1997

[6] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996–1011, June 1996.

[7] N. R. Shah and A. Zakhor, "Multiframe spatial resolution enhancement of color video," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, pp. 985–988, Lausanne, Switzerland, 1996.

[8] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of video sequences using motion compensation," in *Proc. IEEE Int. Conf. on Image Processing*, Lausanne, Switzerland, 1996, vol. 1, pp. 713–716.

[9] C. Toklu, A. T. Erdem, M. I. Sezan, and A. M. Tekalp, "Tracking motion and intensity variations using hierarchical 2-D mesh modeling for synthetic object transfiguration," *Graph. Models Image Process.*, vol. 58, pp. 553–573, Nov. 1996.

[10] S. Fogel, "The estimation of velocity vector fields from time-varying image sequences," *CVGIP: Image Understand.*, vol. 53, pp. 253–287, 1991.

[11] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Processing*, vol. 3, pp. 625–638, Sept. 1994.

[12] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Robust methods for high quality stills from interlaced video in the presence of dominant motion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 328–342, Apr. 1997.

[13] M. M. Chang, M. I. Sezan, and A. M. Tekalp, "Simultaneous motion estimation and segmentation," *IEEE Trans. Image Processing,* vol. 6, pp. 1326–1332, Sept. 1997.

[14] ISO/IEC SC29 WG11 N1469, "MPEG-4 Video Verification Model 5.0," Nov. 1996.

[15] J. Bergen, P. Burt, R. Hingorani, and S. Peleg, "A three-frame algorithm for estimating two component image motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 886–896, Sept. 1992.