ICPR'04 Best Industry Related Paper Award

Super-Resolving Compressed Video with Large Artifacts

WenYi Zhao
Sarnoff Corporation
201 Washington Road, Princeton, NJ 08540
Email: wzhao@sarnoff.com

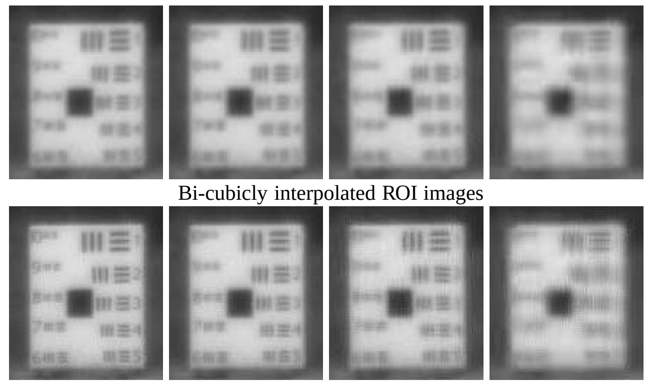
Abstract

We propose methods to super-resolve compressed video sequences that may consist of frames with missing blocks of pixels in addition to compression artifacts. Different from traditional resolution enhancement algorithms, our methods include two key components that are crucial to handle compressed video sequences. The first component is dynamic masking *that dynamically computes image masks used to reject outliers. The second component is* flow-based image repairing *that reconstructs missing blocks or a whole frame by exploring both temporal and spatial information. We demonstrate the proposed methods with real MPEG video sequences.*

1 Introduction

This paper presents methods to super-resolve compressed video sequences. In particular, we demonstrate their applications on super-resolving compressed video sequences that may consist of frames with missing blocks of pixels in addition to compression artifacts. Figure 1 shows a real example where super-resolution for compressed video is very effective until certain bit rate. However, it is not the focus of this paper to discuss the theoretical limitation of super-resolution for compressed video. Instead, we discuss how to robustly improve the resolution of compressed sequences with large artifacts such as blocky effects, missing blocks or even a whole frame. Such artifacts often occur in practice due to large compression ratio and/or transmission error.

Like traditional super-resolution methods, the basic idea is to explore temporal redundancy embedded in the video sequences to enhance image resolution. Unlike traditional methods, we need to handle artifacts in compressed sequences that violate the basic assumption that warped low-resolution images after alignment are sub-sampled from a single high-resolution image. One efficient method to disable these artifacts (outliers) is to apply image masks. To utilize pixel information from other images that is missing or corrupted in the *reference* image that needs to be super-resolved, we propose a *dynamic masking* method that computes image masks every time after the super-resolved reference image is updated.



(a) Original (b) 2M bps (c) 1M bps (d) 256K bps

Figure 1: Comparison of super-resolution results based on original images (a) and MPEG compressed ones at different bitrates (image size 640x480): (b) 2M bps, (c) 1M bps, and (d) 256K bps. Here we only plot a small ROI of size 106x120: the top row plots bi-cubicly interpolated ROI images and the bottom row plots super-resolved results based on consistent flow (Eq. 6) [11]. Notice that super-resolution results with nine images are effective at 2Mbps or higher.

To handle other significant errors such as missing blocks, we propose a *flow-based image repairing* method. Unlike existing image repairing methods (e.g., image impainting [2]) that is purely based on spatial information, the proposed method is based on the assumption that missing spatial information in one frame can often be obtained from other frames. To solve the problem of *computing flow without correspondence* (due to missing pixels), we propose an effective multi-frame flow methods modified from [11]. To handle the case of missing whole frames, the proposed method essentially performs frame interpolation (also called *image tweening*). In this paper, accurate flow estimation methods are used for super-resolution since successful super-resolution is based on the assumption of subpixel displacement among images. Motion vectors available in compressed video are simply not accurate enough.

The paper is organized as follows: After discussing related super-resolution work in Section 2, we propose the dynamic masking method to enhance compressed videos in Section 3. In Section 4, we propose the flow-based image repairing method to repair damaged images by exploring both temporal and spatial information. Experimental results are reported in Section 5 based on real MPEG sequences. Finally, we conclude our paper in Section 6.

2 Related work

The basic principle of super-resolution is to enhance the spatial image resolution by combining high-frequency spatial information spread temporarily across a video sequence. The majority of super-resolution algorithms formulate the problem as a signal reconstruction problem from multiple samples with (non-)uniform sampling theorems as the foundation. Over the past decade, many reconstruction algorithms have been proposed, including frequency-based methods, MAP methods [5], BP (back-projection) methods [7], POCS (projection onto convex set) methods, and hybrid methods [6].

Super-resolution of compressed video is a relatively new topic and is becoming increasingly important. The driving force is the wide use of compressed digital video for storage, transmission and processing. Unlike the traditional problem of super-resolution, there are new issues associated with super-resolving compressed video [10]. The first one is the intensity noise introduced by quantizing transform coefficients based on both statistical and perceptual significance. In order to achieve low bit rate, compression schemes intend to remove differences between frames that are critical for resolution enhancement. A complete model from high-resolution images to low-resolution compressed images is proposed [10] to include compression-induced errors. To reconstruct the high-resolution images, the authors present a Bayesian framework that also incorporates known motion vectors. To apply this framework, several probability distributions need to be defined and then estimated in practice from compressed bit streams. For example, quantization noise in the spatial domain can be modeled as a Gaussian distribution with an unknown covariance matrix [10].

In practice, it is not easy to estimate these density functions that may depend on contents. Alternatively, we propose the general dynamic masking method that adaptively reject outliers due to compression, bad motion estimation, etc. In addition, we propose the flow-based image repairing method to deal with scenarios that are difficult to handle by such theoretical framework, for example, scenarios where large missing blocks of pixel occur due to transmission error in motion vectors.

3 Super-Resolution with Dynamic Masking

3.1 Flow-based super-resolution

The basic assumption of reconstruction algorithms is the modeling of high-to-low resolution image formation process. In particularly, the following model (commonly in matrix notations) has been used [6],

$$Y_k = D_k C_k F_k X + N_k \quad (1)$$

where X is the original high-resolution image, Y_k is the kth low-resolution frame, D_k, C_k , F_k are *decimation*, *blurring* and *motion-warping* matrices, respectively, that embody the corresponding transformations, N_k is the noise at low-resolution. Assuming zero-mean Gaussian noise, the ML-estimator of X from Y_k is

$$\tilde{X} = \arg_X \min\{(Y - HX)^T W (Y - HX)\} \quad (2)$$

where W is the weight matrix determined by noise N_k, matrix H is defined as $[H]_k = D_k C_k F_k$. To regularize the solution to X, penalty (prior) terms can be added to the cost function, resulting in a MAP estimation problem [5]. For regular images, the noise is often assumed to be un-correlated, i.e., a diagonal matrix W. For compressed images, quantization error also contributes to noise N_k. Instead of modeling and estimating the probability distribution of these errors, we propose the dynamic masking method by treating errors as outliers.

To facilitate discussion, we choose the BP methods [7] as the base-line algorithm and use the estimated consistent flow [11] for image warping. However, we can apply any other super-resolution algorithms. Specifically, the iterative update of high-resolution image in the BP method [7] is expressed as:

$$I_h^{(n+1)} = I_h^{(n)} + \frac{1}{K} \sum_{k=1}^{K} \{[(g_k - g_k^{(n)}) \uparrow s]^{F_k} \cdot p \parallel \quad (3)$$

where $I_h^{(n)}$ is the recovered high-resolution image at the n-th iteration, p is a back-projection kernel, $\uparrow s$ denotes an up-sampling operator by a factor s, $[\cdot]^{F_k}$ denotes a forward-warping process. The low-resolution image $g_k^{(n)}$) is simulated from $I_h^{(n)}$ based on the following imaging model

$$\tilde{g}_k = \{[I_h]^{B_k} \cdot h\} \downarrow s \quad (4)$$

where $\downarrow s$ denotes a down-sampling operator by a factor s, $[\cdot]^{B_k}$ denotes a backward-warping process and h is a blurring kernel.

3.2 Dynamic masking

Even for regular video sequences, rejecting outliers due to object occlusions or large motions is necessary. *Static masking* has been used effectively to obtain high-quality images [11]. Static masking is a preprocessing approach where the cross-correlations between warped low-resolution images and the low-resolution reference image g are computed prior to the main iteration (Eq. 3). If the correlation scores are below a certain threshold, corresponding warped pixels are ignored. Alternatively, correlation scores can be used to weight pixels in the main iteration.

For compressed images, one issue exists for applying static masking: the reference image itself could be highly degraded due to compression error. Applying static masking would prevent useful information from other images passing to the reference image during the main iteration. To overcome this shortcoming, we propose dynamic masking where the cross-correlations between warped low-resolution images and the updated low-resolution reference image (subsampled from $I_h^{(n)}$) can be computed at each iteration.

Dynamic masking

1. Obtaining g_k^r warped from g_k towards reference g
2. Begining of the main super-resolution iteration
3. Handling pixels
   - Computing cross-correlation between g_k^r and subsampled $I_h^{(n)}$
   - Taking pixels if scores are $> T_1$
   - Rejecting pixels else if correlation scores are $< T_2$ ($T_1 > T_2$)
   - Weighting pixels else
4. End of the main super-resolution iteration

The proposed method can iteratively repair the degraded reference images during the main iteration while rejecting true outliers. Hence, it is appropriate for handling compressed images with or without large artifacts.

4 Flow-based Image Repairing

4.1 Computing consistent flow fields

Optical flow has been used in [1] for super-resolution, but it is in [11] the feasibility of flow-based super-resolution has been demonstrated. It is based on the following arguments: 1) it is the warping error, not the flow error, that has direct impact on the process of super-resolution, 2) image warping error is typically not as catastrophic as the flow error because large/small motion errors are associated with small/large image gradients (in gradient-based flow algorithms). Observing that flow consistency should be fully enforced to achieve accurate flow estimate, authors in [11] proposed algorithm to compute consistent flow. In the case of two frames, traditional algorithms [3] seek to minimize the following one-directional least square errors

$$Err_i = (I_i(\mathbf{p}_i) - I_j(\mathbf{p}_i + \mathbf{u}_i[\mathbf{p}_i]))^2, \quad (5)$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the coordinates of frame 1 and 2 respectively. The consistent flow algorithm [11] seeks to minimize the consistent least-square error

$$Err_{cons} = [I_1(\mathbf{p} - \alpha\mathbf{u}[\mathbf{p}]) - I_2(\mathbf{p} + (1-\alpha)\mathbf{u}[\mathbf{p}))]^2, \quad (6)$$

where p is the coordinate of a virtual frame and α is a control parameter that is in the range of [0,1]. The choice of the exact value for a depends on the noise statistics of the two frames. Typically, the statistics of the two frames are similar and the value 0.5 is chosen.

To enforce flow consistency across multiple frames, a multi-frame algorithm was proposed to minimize the consistent least square error [11]

$$\begin{aligned} Err_{cons} &= \sum Err_{f2r} + Err_{f2f} \\ &= \sum_{i \neq r} [(I_i(\mathbf{p} - \mathbf{u}_i[\mathbf{p}]) - I_r(\mathbf{p}))^2 \\ &+ \sum_{i \neq j} (I_i(\mathbf{p} - \mathbf{u}_i[\mathbf{p}]) - I_j(\mathbf{p} - \mathbf{u}_j[\mathbf{p}]))^2, \end{aligned} \quad (7)$$

where Err_{f2r} are the errors between each image and the reference image I_r and Err_{f2f} are the errors between a pairs of images (I_i, I_j) other than I_r. However, in this multi-frame formulation, all the images are assumed to be of good quality. To handle the case where some images have missing blocks is the topic of the next section.

4.2 Flow-based image repairing

Our problem can be stated as follows: *given a sequence of images and assuming some images are good and others are damaged, we want to repair the damaged images*. In an extreme case that one whole image is missing, we can compute the flow fields between two surrounding frames and interpolate the missing middle frame (image tweening). Next, we focus on a case where known blocks of pixels are missing from one frame. Different from existing image repairing methods [2], we propose a flow-based method that repairs damaged regions by first computing flow fields from other frames to the damaged frame and then performing image interpolation to repair the damaged frame.

Without loss of generality, we further assume that we have three frames (e.g. frames 1, 5 and 7) and only the middle frame is damaged with bad regions being marked (Fig. 4). Now the problem becomes computing flow without correspondence in the damaged regions. To estimate flow fields from good frames to the middle frame, we need to modify the existing multi-frame flow algorithms (Eq. 7). Recall that the computation of flow fields from other frames to the middle frame are based on iteratively solving the following linear equation [11]

$$\begin{bmatrix} 2\sum(\nabla I_1')(\nabla I_1')^T & -\sum(\nabla I_1')(\nabla I_3')^T \\ -\sum(\nabla I_3')(\nabla I_1')^T & 2\sum(\nabla I_3')(\nabla I_3')^T \end{bmatrix} \begin{bmatrix} \delta\mathbf{u}_1 \\ \delta\mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} I_{t1}\nabla I_1' + I_{t13}\nabla I_1' \\ I_{t3}\nabla I_3' + I_{t31}\nabla I_3' \end{bmatrix} \quad (8)$$

where I_i' are the warped version I_i^w of I_i using motion from previous iteration, $\delta\mathbf{u}_1$ and $\delta\mathbf{u}_3$ are the incremental flows computed at each iteration. $I_{tij} \stackrel{\text{def}}{=} I_i' - I_j'$ is equivalent to $-I_{tji}$ and $I_{tj} \stackrel{\text{def}}{=} I_j' - I_r'$.

Notice that in damaged regions, both temporal gradients I_{ti} and spatial gradients $\nabla I_i'$ can not be computed and we can set them to be zero. However, setting the spatial gradients to be zero yields a degenerated linear equation (Eq. 8). To overcome this, we propose the following modified formula to compute the spatial gradients:

$$\nabla I_i' = (m_i^w \nabla I_i^w + m_r \nabla I_r)/(2[m_i^w + m_r]), \quad (9)$$

where m_r and m_i^w are (warped) image masks indicating damaged pixels for the reference image and warped image I_i^w respectively. Recall that such image masks are readily available in compression algorithms. Now for damaged regions in the reference image, we have $m_r = 0$ but $m_i^w \neq 0$; hence, a non-zero $\nabla I_i'$ and a solvable linear equation (Eq. 8). For the temporal gradients I_{ti}, we can either set them to be zero or assign a value by "interpolation": $I_{ti} = I_{tij}/2$. To ensure the success of the proposed algorithm, we apply a multi-resolution scheme [4] to explore spatial information, i.e., areas surrounding the damaged regions.

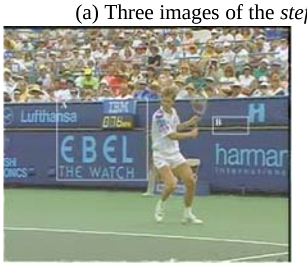
5 Experiments

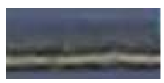
We have tested the proposed methods on two real MPEG video sequences. The first experiment demonstrates the efficacy of applying dynamic masking to super-resolve these two MPEG sequences (Figs. 2,3). For comparison, we have run robust super-resolution method based on median filter [9] (Fig. 2(d)). To see the difference of applying static masking and dynamic masking, we have tried the challenging *foreman* sequence (Fig 3) where video sequence has been downsampled to 10 fps and smooth facial skins further complicate motion computation.

In the second experiment we show how flow-based image repairing can effectively recover the missing pixels. We have simulated the block-missing errors (Fig. 2) by manually marking out four regions in the reference image (Fig. 2 vs Fig 4(a)). For the detailed results of image repairing please see Fig. 4(c). After the reference image is repaired, super-resolution is straight-forward and the result is similar to the case when good-quality reference image is available.

Reference - 1 Reference Reference + 1

(a) Three images of the *stefan* sequence of five frames





Bi-cubic

Dynamic masking

(b) Super-resolved image (c) Close-up region B

Bi-cubic Dynamic masking Robust Median[9] No masking

(d)Comparison of region A

Figure 2: Comparison of super-resolution results based on the CIF format *stefan* sequence of five compressed frames: (a) Three of the five images (notice the large motions); (b) Super-resolved reference image based on *dynamic masking* and consistent flow; (c) Effect of super-resolution (notice how blocky artifacts are smoothed out); (d) Comparison of different high-resolution results of region A (notice that dynamic masking produces the best overall resuslts).

6 Discussion and Conclusions

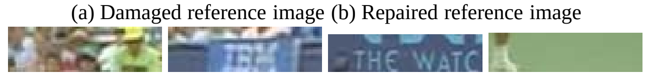
In this paper, we have presented the dynamic masking method to handle compression artifacts. To handle other significant errors such as missing blocks of pixels, we have presented the method of flow-based image repairing. The efficacy of these methods have been supported by experimental results with real MPEG sequences. The two methods can be combined, for example, by first repairing and then super-resolving the reference image. For future work, we would consider motion vectors already available in bit streams as the initial estimate of flow. For detection of unknown missing blocks, we would consider existing methods [8].

(a) Three images of the *forman* sequence of nine frames

(b) Bi-cubic Static masking Dynamic masking

Figure 3: Comparison of dynamic masking and static masking with the CIF format *foreman* sequence downsampled to 10fps. (a) Three frames from this sequence include the reference image (middle); (b) To save space, we only plot and compare super-resolved images in the nose region. Notice that dynamic masking produces sharper image than spatial interpolation and less artifacts than static masking.

(a) Damaged reference image (b) Repaired reference image



Repaired missing regions

Scaled-up difference from ground-truth

(c) Close-up damaged regions

Figure 4: Flow-based image repairing based on three frames: (a) Damaged reference image (compared to Fig. 2(a)); (b) Repaired reference image using other good frames (Fig. 2(a)); (c) Close-up damaged regions: repaired regions and comparison to ground-truth.

References

[1] S. Baker and T. Kanade, "Superresolution Optical Flow," Carnegie Mellon Univ., Robotics Institute, Tech. Report: CMU-RI-TR-99-36, 1999.

[2] C. Ballester, V. Caselles, J. Verdera, M. Bertalmio, and G. Sapiro, "A Variational Model for Filling-In Gray Level and Color Images," In *Proc. Int. Conf. Comp. Vision*, 2001.

[3] J. Bergen, P. Anadan, K. Hanna, and R. Hingorani, "Hieararchical Molde-Based Motion Estimation," In *Proc. European Conf. Comp. Vision*, pp. 237-252, 1992.

[4] P. Burt and E. Adelson, "The Laplacian Pyramid as A Compact Image Code," *IEEE Trans. on Communications*, Vol 31, pp.583-540, 1983.

[5] D. Capel and A. Zisserman, " Super-resolution Enhancement of Text Image Sequences," In *Proc. Conf. Comp. Vision and Patt. Recog.*, 2000.

[6] Elad, M. and Feuer, A. 1997. Restoration of a single superresolution image form several blurred, noisy and undersampled measured images. *IEEE Trans. on Image Processing*, pp. 1646–1658.

[7] M. Irani and S. Peleg, "Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transprareny," *Journal of Visual Comm. and Image Repre.*, Vol. 4, pp. 324-335, 1993.

[8] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner, "Detection of Missing data in image sequences," *IEEE Trans. Image Processing*, 11(4), pp. 1496-1508, 1995.

[9] A. Rav-Acha, A. Zomet, and S. Peleg, "Robust Super Resolution," In *Proc. Conf. Comp. Vision and Patt. Recog.* Vol I, pp. 645-650, 2001.

[10] C.A. Segall, R. Molina, and A.K. Katsaggelos, "High-Resolution Images from Low-Resolution Compressed Video," *IEEE Signal Processing Magazine*, pp. 37-48, May, 2003.

[11] W. Zhao and H. Sawhney, "Is super-resolution with optical flow possible?" In *Proc. European Conf. Comp. Vision*, 2002.