

Robust Super-Resolution*

Assaf Zomet Alex Rav-Acha Shmuel Peleg
School of Computer Science and Engineering,
The Hebrew University of Jerusalem,
91904, Jerusalem, Israel
E-Mail: {zomet,alexis,peleg}@cs.huji.ac.il

Abstract

A robust approach for super resolution is presented, which is especially valuable in the presence of outliers. Such outliers may be due to motion errors, inaccurate blur models, noise, moving objects, motion blur etc. This robustness is needed since super-resolution methods are very sensitive to such errors.

A robust median estimator is combined in an iterative process to achieve a super resolution algorithm. This process can increase resolution even in regions with outliers, where other super resolution methods actually degrade the image.

1. Introduction

Super resolution is a method for reconstructing a high resolution image from several overlapping low-resolution images. Most super resolution techniques present the process in the following way [11, 12, 9, 17, 2, 14, 15]: The low resolution input images are the result of resampling a high resolution image. The goal is to find the high resolution image which, when resampled in the lattice of the input images according to the imaging model, predicts well the low resolution input images.

The success of super resolution algorithms is highly dependent on the accuracy of the model of the imaging process. If, for example, the motion computed for some of the images is not correct, the algorithm may degrade the image rather than enhance it.

One solution proposed to handle local model inaccuracies and noise is regularization [2, 14, 4, 5]. The super resolution image is presented as the MAP solution of a stochastic optimization, and prior smoothness assumptions are used to reduce the effects of inconsistent measurements. In most cases the enforced smoothness results in the suppression of high-frequency information, and the results are

*This research has been supported by a grant from the Israeli Ministry of Science and Technology.

blurred. Regularization may be successful when the scene is strongly restricted, e.g. a binary text image. [5].

Another approach [10] handles the case of moving objects by motion segmentation. For this approach an accurate motion segmentation must be available, a segmentation that is difficult to obtain in the presence of aliasing and noise.

In this paper, a robust median-based estimator is used to discard measurements which are inconsistent with the imaging model. This can handle local model inconsistencies such as highlights, moving objects, parallax, etc., at a low computational cost. A bias detection procedure reduces artifacts due to bias of the median estimator.

Using median in the context of super resolution has been proposed in the past [6], where the upsampled images were combined by median, and the result image was convolved by a high pass filter. Our method is inherently different, as it uses all the input data in the super resolution high pass mechanism.

A robust super resolution algorithm was proposed in the context of mosaicing [7]. This algorithm minimizes the error under a norm which is more robust than the commonly used ℓ_2 norm. In Section 2.1 we refer to this approach.

This paper describes a novel super resolution algorithm which is robust to outliers caused by model inaccuracies, moving objects etc. The resolution of the inlier regions (the background) is enhanced, provided it appears in at least half of the input images.

The algorithm requires a motion computation step which is both accurate, and robust to outliers. In our implementation we have used the algorithm described in Section 4.1.3. of [1] to compute homographies between the images. Other methods, e.g. the methods described in [16, 3], can be used as well.

2 Robust Super Resolution

We follow the notational framework where images are reordered in column vectors [9]. The basic operations in the image formation model, such as convolution, sampling and warping are linear in the image intensities, and thus can

be represented as matrices operating on these vector images. Given n input images g_1, \dots, g_n , the image formation process of g_k from the super resolved image f can be formulated in the following way [9]:

$$\vec{Y}_k = D_k C_k F_k \vec{X} + \vec{E}_k \quad (1)$$

where:

- \vec{X} is the high resolution image f reordered in a vector.
- \vec{Y}_k is the k -th input image g_k reordered in a vector.
- \vec{E}_k is the normally distributed additive noise reordered in a vector.
- F_k is the geometric warp matrix.
- C_k is the blurring matrix.
- D_k is the decimation matrix.

The total squared error of resampling the high resolution image f (represented by \vec{X}) is:

$$L(\vec{X}) = \frac{1}{2} \sum_{k=1}^n \|\vec{Y}_k - D_k C_k F_k \vec{X}\|_2^2 \quad (2)$$

taking the derivative of L with respect to \vec{X} , the gradient of L is the sum of gradients computed over the input images

$$\begin{aligned} \vec{B}_k &= F_k^T C_k^T D_k^T (D_k C_k F_k \vec{X} - \vec{Y}_k) \\ \nabla L(\vec{X}) &= \sum_{k=1}^n \vec{B}_k \end{aligned} \quad (3)$$

The simplest gradient-based iterative minimization method updates the solution estimate in each iteration by

$$\vec{X}^{n+1} = \vec{X}^n + \lambda \nabla L(\vec{X}) \quad (4)$$

where λ is a scale factor defining the step size in the direction of the gradient. In the image domain [17], it is a version of the Iterated Back Projection method [11]. In each iteration, the high resolution estimate is resampled in the lattices of the input images. The difference between this resampled image and the input image is projected back to the high-resolution lattice. Each term \vec{B}_k of the sum in Eqn. (3) corresponds to such a back-projected difference image.

In order to introduce robustness into the procedure, the sum of images in Eqn. (3) is replaced with a scaled pixel-wise median:

$$\nabla L(\vec{X})(x, y) \approx n \cdot \text{median}\{\vec{B}_k(x, y)\}_{k=1}^n \quad (5)$$

For a symmetric distribution, a median can approximate the mean quite accurately, given a sufficient set of samples. In case of distant outliers, the median is much more robust than the mean. The median estimate can be biased when

the outlier measurements are organized non-symmetrically with respect to the mean. This case is handled by the bias-detection procedure described next. In the Appendix, a proof of the symmetry of this distribution is given for the case of translational motion.

If some prior assumptions can be made about the solution, one can combine this robust scheme in a constrained optimization algorithm, for example by adding a prior term [17, 5], or by using the Projection-Onto-Convex-Sets method [12].

2.1 Why Pixel-wise median ?

A straight-forward approach to make the super resolution algorithm robust would be to minimize the error under a different norm which can handle outliers better than the ℓ_2 [7]. This means that large projection errors would have a small influence on the computed solution. Large projection errors may be due to outliers, but are also related to aliasing in regions containing high frequencies. Note that aliasing is the main source of information for resolution enhancement. As a result, such approach would suppress the influence of the most informative measurements. By treating each pixel in the estimated solution independently, we ensure the enhancement of pixels in regions containing high-frequencies. In addition, minimizing under robust norms is computationally expensive.

In addition to the theoretical analysis in the Appendix, we have also experimented several robust pixel-wise estimators, including median, Least-Median-of-Squares and Trimmed-mean. The median filter achieved the best results. Moreover, in case there were no outliers, the results achieved using the median estimator were indistinguishable with respect to the standard algorithm. This is illustrated in Fig. 3.

2.2 The Bias Detection Procedure

The aim of the bias-detection (BD) procedure is to detect most of the outlier measurements which may bias the median estimator. These measurements can be extracted from the differences between aligned images. Equally important, the bias detector should not detect image differences due to aliasing, the main source of information for super resolution. The result of the BD procedure for every input image can be a binary mask.

Given the current estimate of the high resolution image $f^{(j)}$, The BD mask for input image g_k in the j -th iteration is determined from the error images e_k^j :

$$e_k^j = g_k - P_j(f^{(j)}) \quad (6)$$

where $P_j(f^{(j)})$ is the resampling of $f^{(j)}$ on the lattice of image g_k , composed of blur, warp and decimation operators, as described in Eqn. (1).

A typical case in which the values of e_k^j are non-zero due to aliasing and zero-mean noise is presented in Fig. 1.

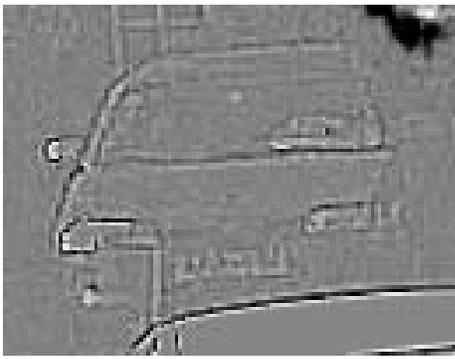


Figure 1. An example of an error image as described in Eqn. (6), where positive values are bright, and negative values are dark. The errors are mainly along the edges of the car, and their intensities in small regions are symmetric with respect to zero. An example for an outlier region appears on the top right corner of the image.

Most of the energy of e_k^j is concentrated near intensity edges, where every positive value has a neighboring negative value. A local average on e_k^j will thus be close to zero.

When e_k^j contains non-zero values due to small homogeneous outlier regions, then the image e_k^j tends to be non-symmetric with respect to zero. When e_k^j contains large values, but it is symmetric with respect to zero, then the median estimator is expected to be less effected by these outliers.

Thus detecting local homogeneous regions of g_k , with locally symmetric values of e_k^j helps to discard most of the outliers while minimizing the effect of aliasing and noise.

The BD mask can be computed by a convolution of e_k^j with a low pass filter, such as a local-neighborhood mean filter.

It can be further divided by a local measurement of the variability of g_k . This enables to determine a global threshold which is invariant of the local variability of the images.

3 Experiments

To test the quality and robustness of the super resolution algorithm, the results of four different algorithms were compared:

1. IBP (Iterated Back Projection) [11].
2. IBP, combined with the Bias Detection procedure, described in Sect. 2.2.
3. IBP with a median replacing the sum.
4. IBP with median and Bias Detection.

All the above algorithms used the same initial estimate for f ; The input images have been enlarged and warped to

a common coordinate system, and their median was computed. In all the experiments the zoom factor was 2, and the camera was shaking, inducing a planar-projective motion.

The results of the four algorithms on the “Junction” sequence are presented in Fig. 2. The results of the mean-based methods contain more leftovers than the median-based methods. The “dirty” region in the middle of the road is observed also in the initial estimate, the median of the aligned images. This implies that in this region the algorithm assumptions were not valid: The region was occluded in more than half of the input images. This region can be detected by a Quality Validation module which tests whether the reprojection error of this region on the input images satisfies the assumptions: Low error in at least half of the input images.

One can see in Fig. 2-g,h) that in addition to the moving-objects removal, the algorithm also enhances the resolution. Another example of the resolution enhancement of the algorithm is shown in Fig. 3.

In Fig. 4 the results of the sum-based IBP algorithm and the proposed algorithm are compared, in case the algorithm assumption is valid. Leftovers of the moving objects (pedestrians) are observed in the IBP result. By replacing the sum with a median, these leftovers are removed.

Experiments were conducted replacing the proposed pixel-wise median with two other robust estimators. Least-Median-Of-Squares (LMeds) [13], and Trimmed-Mean. In order to achieve up-to-50% robustness, the Trimmed-Mean was computed by sorting the error values by their magnitude, and discarding the top 50%. The LMeds estimator can also be computed very efficiently on 1-D data. [13]. The results of the different estimators are shown in Fig.3. While the result of using the median is almost identical to the result of using the mean, both the LMeds and Trimmed-Mean introduce artifacts due to bad estimation of the mean.

4 Summary

A robust approach for super resolution has been presented, in which a median estimator was used to robustly estimate a sum in an iterative framework. This resulted with an increased robustness at a low computational cost, and with no distinguishable loss of accuracy.

Future research of robust super resolution may expand this work to reconstructing a high resolution video [8], in the presence of model inaccuracies such as moving objects and highlights.

Acknowledgments

The authors would like to thank Moshe Ben-Ezra for providing “Solvex”, the package for solving linear equations under the ℓ_1 norm, and for some helpful advices.

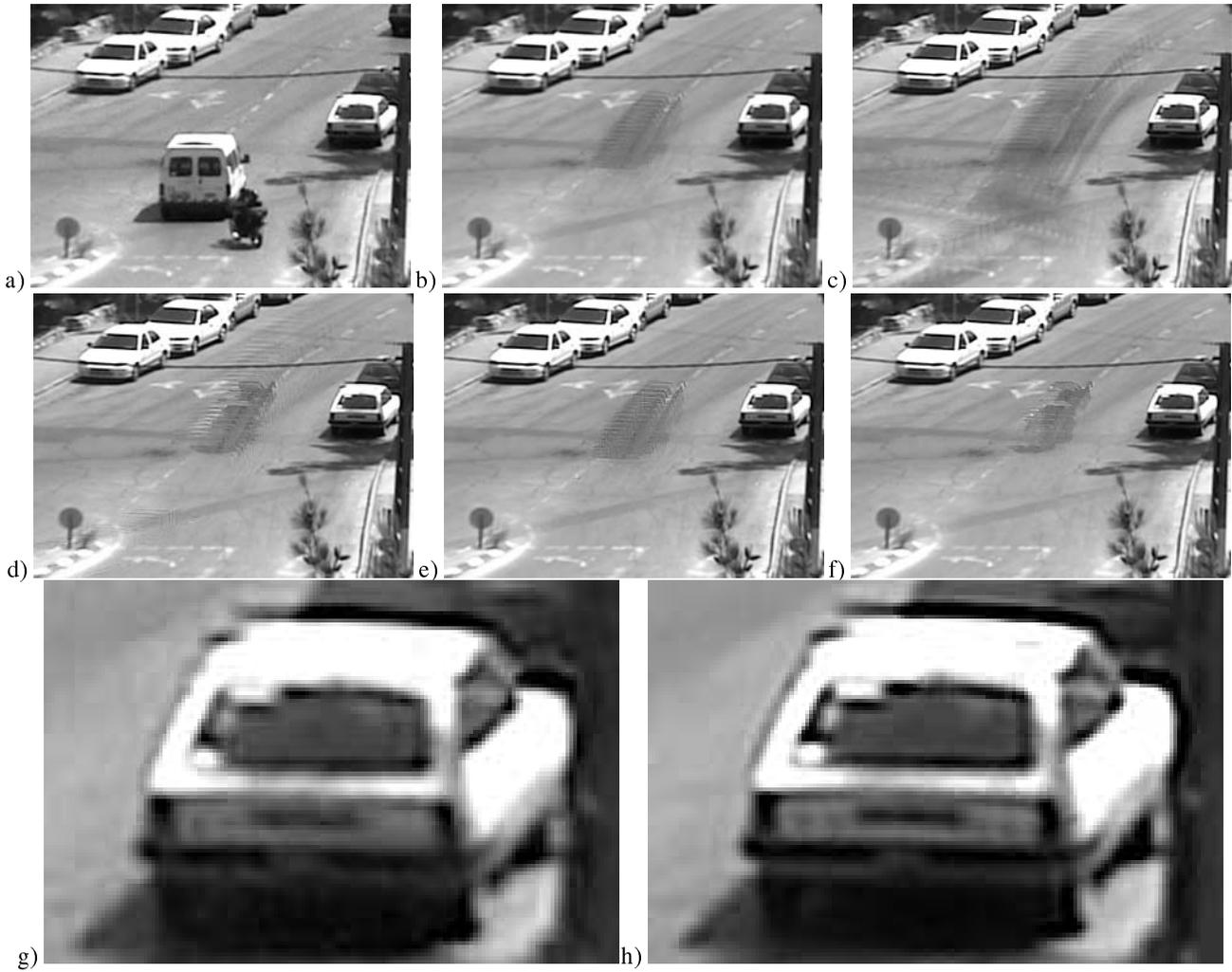


Figure 2. The results on the “Junction” sequence. a) One of the input images. b) The median of the aligned images, the initial estimate of the SR algorithms. c) IBP. d) IBP with Bias Detection. e) Median IBP. f) Median IBP with Bias Detection. g) Enlarged region from the input image. h) Enlarged region from (f). See Section 3 for further analysis.

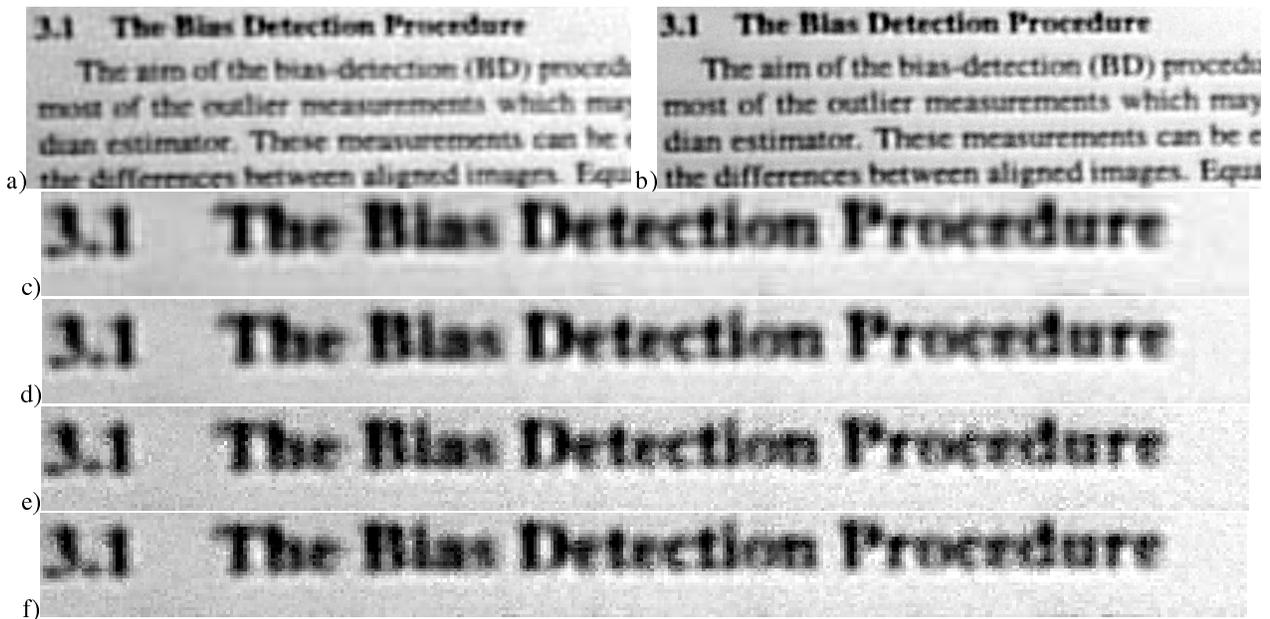


Figure 3. The results on the “Text” sequence. a) One of the input images. b) The result of the proposed algorithm, without bias detection. c-f) are enlarged images. c) is the result of using Mean, as in the original IBP algorithm. d) The proposed algorithm, replacing the mean by a median. e) Using LMeds. f) Using Trimmed Mean. It can be seen that both e) and f) are more noisy than d).



Figure 4. The results on the “University” sequence. a) A patch from one of the input images. b) The IBP algorithm. c) The proposed median IBP algorithm.

Appendix: The Symmetry of the Re-projected Error.

In the proposed algorithm, we estimate the average of a set of images with a scaled pixel-wise median. In order to statistically justify this step, we show that for every location in the high-resolution image the distribution of re-projected error is symmetric. Thus a median can be used to estimate the mean of this distribution. The proof is given for 1-D signals with a random translation.

Let $f \in \mathbb{R}^{2n}$ be the super resolution solution, and let $\hat{f} \in \mathbb{R}^{2n}$ be the current estimate of the solution. Let $g_u \in \mathbb{R}^n$ be an input image such that

$$g_u = f * k * \delta(u) \downarrow$$

where $*$ marks convolution, k is the camera non-perfect low-pass filter, $\delta(u)$ is a u -translation operator, and \downarrow marks decimation. The re-projected error from image g_u is:

$$d_u = (g_u - \hat{f} * k * \delta(u) \downarrow) \uparrow * \delta(-u) * \hat{k} = (f - \hat{f}) * k * \delta(u) \downarrow \uparrow * \delta(-u) * \hat{k} \quad (7)$$

Where \uparrow marks up-sampling by zero-padding, and \hat{k} is the flipped kernel of k [17] (for a symmetric kernel, $\hat{k} = k$).

Theorem 1 Let $p_x(u) = d_u(x)$, and assume u distributes uniformly in the range $[-1,1)$. Then $p_x(u)$ distributes symmetrically about its expectancy i.e. $\forall \alpha, P(p_x(u) = E_u(p_x) - \alpha) = P(p_x(u) = E_u(p_x) + \alpha)$

Proof: We mark the Fourier transform of a signal by the corresponding upper-case letter, e.g. $F = \mathcal{F}(f)$, $K = \mathcal{F}(k)$. Eqn (7) can be rewritten in the frequency domain:

$$D_u(w) = [(F(w) - \hat{F}(w))e^{\frac{2\pi i w u}{2n}} K(w)] \downarrow^F e^{-\frac{2\pi i w u}{2n}} \hat{K}(w) \quad (8)$$

Where \downarrow^F is the frequency folding occurring due to low-rate sampling followed by zero padding:

$$G(W) \downarrow^F = \begin{cases} G(w) + G(w - n) & 0 \leq w \leq n \\ G(w) + G(w + n) & -n \leq w < 0 \end{cases} \quad (9)$$

Let $s : \mathfrak{R} \rightarrow \{-1, 1\}$ be the sign function, and let $\bar{w} = w - s(w)n$. Plugging Eqn.(9) in Eqn. (8):

$$\begin{aligned} D_u(w) &= (F(w) - \hat{F}(w))e^{\frac{2\pi i w u}{2n}} K(w)e^{-\frac{2\pi i w u}{2n}} \hat{K}(w) + \\ &\quad (F(\bar{w}) - \hat{F}(\bar{w}))e^{\frac{2\pi i \bar{w} u}{2n}} K(\bar{w})e^{-\frac{2\pi i \bar{w} u}{2n}} \hat{K}(w) = \\ &\quad (F(w) - \hat{F}(w))K(w)\hat{K}(w) + \\ &\quad (F(\bar{w}) - \hat{F}(\bar{w}))K(\bar{w})e^{-\pi i u s(w)} \hat{K}(w) \end{aligned}$$

The first term does not depend on u . Let $T_u(w)$ be the second term. In order to complete the proof, it is sufficient to show that $t_u(x) = \mathcal{F}^{-1}(T_u(w))$ distributes symmetrically about 0.

Lemma 1 $\forall x, \forall 0 \leq u \leq 1, t_{u-1}(x) = -t_u(x)$. Since u distributes uniformly in $[-1, 1]$, it follows that $t_u(x)$ distributes symmetrically about 0.

Proof:

$$\begin{aligned} T_{u-1}(w) &= (F(\bar{w}) - \hat{F}(\bar{w}))K(\bar{w})e^{-\pi i(u-1)s(w)} \hat{K}(w) = \\ &= (F(\bar{w}) - \hat{F}(\bar{w}))K(\bar{w})e^{-\pi i u s(w)} \hat{K}(w)e^{\pi i s(w)} = -T_u(w) \end{aligned}$$

Applying the inverse Fourier transform, $t_{u-1}(x) = -t_u(x)$. ■ ■

References

[1] M. Ben-Ezra, S. Peleg, and M. Werman. Real-time motion analysis with linear programming. *CVIU*, 78(1):32–52, April 2000.

[2] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia. Reconstruction of high resolution 3D visual information using sub-pixel camera displacements. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 654–657, 1994.

[3] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields. *CVIU*, 63(1):75–104, January 1996.

[4] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 885–891, June 1998.

[5] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *Int. Conf. Pattern Recognition*, pages Vol I: 600–605, 2000.

[6] M. Chiang and T. Boulton. Efficient super-resolution via image warping. *IVC*, 18(10):761–771, July 2000.

[7] F. Dellaert, S. Thrun, and C. Thorpe. Mosaicing a large number of widely dispersed, noisy, and distorted images: A bayesian approach. Technical Report CMU-RI-TR-99-34, Robotics Institute, Carnegie Mellon University, March 1999.

[8] M. Elad and A. Feuer. Super-resolution reconstruction of continuous image sequence. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, September 1996.

[9] M. Elad and A. Feuer. Restoration of a single super-resolution image from several blurred, noisy, and undersampled measured images. *IP*, 6(12):1646–1658, December 1997.

[10] P. Eren, M. Sezan, and A. Tekalp. Robust, object based high resolution image reconstruction from low resolution video. *IP*, 6(10):1446–1451, October 1997.

[11] M. Irani and S. Peleg. Improving resolution by image registration. *Graphical Models and Image Processing*, 53:231–239, 1991.

[12] A. Patti, M. Sezan, and A. Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Trans. Image Processing*, 6(8):1064–1076, August 1997.

[13] P. Rousseeuw and A. Leroy. Robust regression and outlier detection, 1987.

[14] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Processing*, 5(6):996–1011, June 1996.

[15] H. Shekarforoush and R. Chellappa. Data-driven multi-channel superresolution with application to video sequences. *Journal of the Optical Society of America*, 16(3):481–492, March 1999.

[16] P. Torr. Motion segmentation and outlier detection. In *Ph. D.*, 1995.

[17] A. Zomet and S. Peleg. Efficient super-resolution and applications to mosaics. In *ICPR00*, pages Vol I: 579–583, 2000.