Multi-sensor Super-Resolution

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

Abstract

Image sensing is usually done with multiple sensors, like the RGB sensors in color imaging, the IR and EO sensors in surveillance and satellite imaging, etc. The resolution of each sensor can be increased by considering the images of the other sensors, and using the statistical redundancy among the sensors. Particularly, we use the fact that most discontinuities in the image of one sensor correspond to discontinuities in the other sensors.

Two applications are presented: Increasing the resolution of a single color image by using the correlation among the three color channels, and enhancing noisy IR images.

Keywords: *Super-Resolution, Demosaicing, Color, Multi-Sensor, Restoration.*

1 Introduction

Image sensing is usually done with multiple sensors. A color image, for example, is a combination of three sensors: red, green, and blue. In visual surveillance and satellite imaging, sensors that are even more different are often used, e.g. some sensors in the visible domain and other sensors in the infra red domain. Combined depth-color cameras (e.g. [1]) are also becoming available.

In most multi-sensor applications, it is assumed that the images are aligned. Otherwise the motion between the images is computed [9, 20, 5], and the images are aligned by warping. The resampling done for warping degrades the quality of the combined image.

This paper presents a new way to combine the information from different non-registered sensors. Given a set of images of possibly different sensors viewing the same scene, the resolution of one image is improved by using the other images. In color RGB sensors, for example, the red channel is enhanced using the green and blue channels. Similarly the green channels are enhanced using the red and blue channels. The result of combining the enhanced resolution channels is a higher resolution color image.

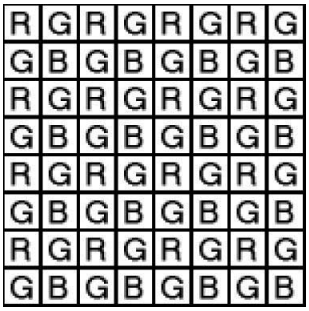

Figure 1. The Bayer pattern, a common way to organize the Red, Green and Blue sensors in a grid.

We adapt the super resolution engine [10, 15, 6, 18], developed for same-sensor images, to combine information from different sensors. The multi-sensor extension is enabled by exploiting statistical redundancy among the sensors. Particularly, we use the fact that most discontinuities in the image of one sensor correspond to discontinuities in the other sensors, and find a local affine mapping between the intensities of different sensors along the edges. We address the validity of this model both analytically and experimentally. Clearly, for very different sensors such as medical modalities [20] this model is less useful.

One application of the multi-sensor super-resolution is the improvement of resolution in a *single* multi-sensor image whose channels are not registered. This is the typical case in 1-CCD color cameras (see Fig. 1), where each pixel location has a sensor of a single color. It may also occur in 3-CCD cameras, where each color has a full CCD, and there is no perfect registration of the three sensors.

1.1 Previous Work

There have been much work on the combination of different sensors, and particularly recovery of color image values from noisy samples. One approach is color image restoration ([2, 12, 19, 17]), where the combination is usu-

Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)
0-7695-1858-3/02 $17.00 © 2002 IEEE

ally achieved by a joint-channel regularization term, aiming mainly to reduce noise. Since super resolution is a generalization of image restoration [6], the adaption of these algorithms to super resolution algorithms is straightforward. The approach of this paper is inherently different, since it minimizes a projection error rather than forcing spatial inter-channel smoothness constraints on the solution. When prior knowledge of the scene is available [4, 14] regularization terms [2, 12] can be combined in the algorithm. A comparison of the results of multi-Sensor super resolution and multi-sensor restoration is shown in the experiments.

Another bulk of related work are demosaicing algorithms [13, 11, 16, 7, 14], aiming in recovery of the missing samples in a 1-CCD color filter array (Fig. 1). The proposed approach is more general, allowing an arbitrary transformation between the different sensors.

It is interesting to compare the presented method to multi-sensor fusion [3], where the visual information captured by various sensors is combined into a single image. This combined image includes the information (i.e. edges) from all sensors. In the presented method, one of the input images is set as photometric reference image, and the resulting image contains only the edges which appeared in this reference image. An edge that is included in another sensor, but not in the reference image, will not appear in the enhanced image.

2 Multi-Sensor Super Resolution

Existing super resolution techniques present the process in the following way [10, 15, 6, 21, 18]: The low resolution input images $\{g_i\}_{i=1}^n$, all captured by the same sensor, are the result of imaging some high resolution (unknown) image f. The imaging model usually includes geometric warping, camera blur, decimation, and additive noise. The goal is to find the high resolution image f which, when imaged into the lattice of the input images according to the respective imaging model, predicts well the low resolution input images. Let $\hat{f}$ be an estimate of the unknown image f. Then the prediction error for $\hat{f}$ is:

$$d_i = g_i - B_i(W_i(\hat{f})) \downarrow \tag{1}$$

where W_i, B_i are the geometric warping and blurring operators respectively, and $\downarrow$ is decimation. B_i is the combination of optical blur, sensor blur, and motion blur. W_i is based on on the estimated displacement between the images.

Most super resolution algorithms [10, 15, 6, 21, 18] minimize the ℓ_2 norm of the error (possibly with regularization), by iteratively simulating the imaging process, and using d_i to update the solution estimate $\hat{f}$.

The proposed multi-sensor super-resolution algorithm uses the same framework. $\hat{f}$ is assumed to be a high resolution version of g_0 (w.l.o.g). Since $g_i, i > 0$, were possibly captured by different sensors than g_0, no imaging process has created g_i from f. Still, aligned images of different sensors are statistically related. Using a model M relating the projected image $p_i = B_i(W_i(\hat{f})) \downarrow$ with the input image, a virtual input image $\hat{g}_i$ and a virtual prediction error image v_i can be defined:

$$\hat{g}_i = M(p_i, g_i), \quad v_i = \hat{g}_i - p_i$$

The virtual prediction error v_i replaces the prediction error d_i in the super resolution algorithm, e.g. [10, 15, 18].

We use an affine relation between the intensities of different sensors in a local neighborhood. We select corresponding neighborhoods in two images p_i and g_i, possibly taken by different sensors. For this neighborhood we estimate an affine transformation relating the intensity values of p_i to the intensity values of g_i, $p_i(x,y) = a(x,y)g_i(x,y) + b(x,y)$. This mapping is used to compute the virtual input image $\hat{g}_i(x,y) = a(x,y)g_i(x,y) + b(x,y)$, and the virtual prediction error $v_i(x,y) = \hat{g}_i(x,y) - p_i(x,y)$. To simplify notations in the following equations, we will use a, b to represent $a(x,y), b(x,y)$.

Assuming the image is contaminated with a zero-mean white noise, the optimal estimator for the affine relation between a region in the projected image p_i and a region in the input image g_i minimizes the following squared error: $min_{\{a,b\}} \sum_{u,v} (ag_i(u,v) + b - p_i(u,v))^2$.

In Section 2.1 we discuss the validity of the local affine model for relating the intensity value of different sensors, and explain why it can be useful in the context of super resolution. Still, this model fails in some cases, and thus it is important to check its validity before using it. The absolute average-centralized normalized cross correlation is a good measure for affine similarity. The maximal absolute value of "1" indicates two signals that are related by an affine transformation.

In summary, the virtual input image $\hat{g}_i = M(p_i, g_i)$ can be computed for pixel (x,y) as follows:

- Estimate the absolute of centralized-normalized cross correlation in the (weighted) window of size $2k+1$ around (x,y):

$$s(x,y) = \frac{\overline{p_i g_i}(x,y) - \overline{p}_i(x,y)\overline{g}_i(x,y)}{\sqrt{(\overline{p_i^2}(x,y) - \overline{p}_i(x,y)^2)(\overline{g_i^2}(x,y) - \overline{g}_i(x,y)^2)}} \tag{2}$$

  Where $\overline{p_i}(x,y)$ is a weighted average of p_i in a neighborhood around (x,y):

$$\overline{p}_i(x,y) = \sum_{-k \le u,v \le k} w(u,v) p_i(x+u, y+v)$$

  and similarly for $\overline{g_i}, \overline{p_i g_i}, \overline{p_i^2}, \overline{g_i^2}$. Weighting the window by distance-decreasing kernel w reduces spatial discontinuities in the affine parameters.

Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)
0-7695-1858-3/02 $17.00 © 2002 IEEE

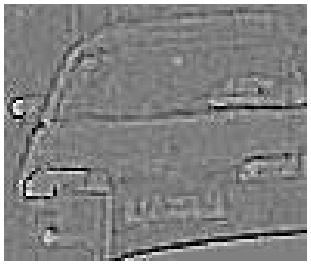

Figure 2. An example of a projection error image d_i, as defined in Eq. 1. The non-zero values concentrate along the edges of the car.

- If $\|s(x,y)\| < T$, the correlation is bellow some threshold, skip the affine estimation and assign $\hat{g}_i(x,y) = p_i(x,y)$.
- Otherwise, compute the least-squares estimate for the affine similarity in the (weighted) window:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{p_i^2}(x,y) & \overline{p}_i(x,y) \\ \overline{p}_i(x,y) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \overline{p_i g_i}(x,y) \\ \overline{g}_i(x,y) \end{bmatrix} \quad (3)$$

- Use the affine estimate $\hat{g}_i(x,y) = a g_i(x,y) + b$

$\hat{g}_i$ is then used for estimating the virtual prediction error v_i which is reprojected to the solution estimate.

The threshold T expresses the validity of the affine approximation. It was experimentally checked that modifying T in ± 0.05 induces negligible influence on the results.

2.1 Discussion: Local Affine Model for Inter-sensor Prediction.

The use of local affine model for estimating v_i from the input image g_i is based on two simple observations: First, most of the information used by the super resolution algorithms is concentrated along the edges. Second, in many practical cases, images of different sensors have edges at corresponding locations.

Fig. 2 shows a typical prediction error image d_i of a single-sensor super resolution algorithm. Most of the details for resolution enhancement are concentrated along the edges. Therefore, the most important regions to estimate in the virtual prediction image v_i, are along the edges.

Local affine relation between the intensity values of different sensors has been used by Irani and Anandan as a distance measure for image alignment [9]. While this assumption does not hold for general sensors, it is useful in several practical cases, e.g. for different color sensors (neglecting minor misalignments due to color aberration [8]).

To test the validity of the model between different color sensors, we have computed statistics over various images of different types (natural, urban, faces etc.). The results, presented in Fig. 4, show that typically more than 90% of the edges follow the affine model (correlation ≥ 0.8). This is translated to a low model prediction error (Fig. 4-b).

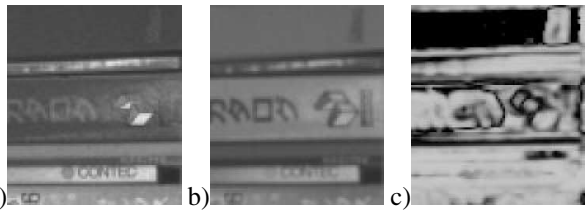

Figure 3. The correlation values between different sensors. a) The Red channel. b) The Green channel. c) The absolute of local correlation between the two colors. Most regions have large correlation, except regions containing more than two colors (corners) and uniform regions were most intensity variations are due to noise.

The model validity can be also shown analytically. Looking on small neighborhoods, most edges can be modeled as abrupt transitions between two relatively homogeneous regions. Using a simplifying assumption that these regions are homogeneous, it can be easily shown that the linearity of the blurring operator W_i implies an affine transformation between the intensity values of different sensors near an edge. The full proof is omitted due to space limitations.

The affine approximation fails in some cases, for example when the imaged region contains more than two homogeneous colors, or when the variations in intensity are mainly due to noise, as demonstrated in Fig 3. It is therefore important to test the validity of the model using the correlation measure (Eq. 2) before using it.

3 Experiments

We first tested the algorithm on color samples organized in a Bayer pattern, as presented in Fig. 1. The algorithm was applied to each of the sensors, improving for example the red image using the blue and green images, etc. For the local photometric affine alignment stage we used in all our experiments a window weighted by a Gaussian kernel with standard deviation 1.225, and a correlation threshold $T = 0.8$. The algorithm was not sensitive to changes in these values. For the blurring operator we have used an isotropic Gaussian with standard deviation 1.0 pixels.

We examined whether adding information from other sensors increases the quality of the image. The algorithm was applied to the red image, using the green and blue channels as input, and the result was compared to single-image super resolution on the red channel image, which is equivalent to a high pass filter on the image. The experiment results are presented in Fig. 7.

To demonstrate the quality of the combination of the three super-resolved channels, we compared it with three

Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)
0-7695-1858-3/02 $17.00 © 2002 IEEE

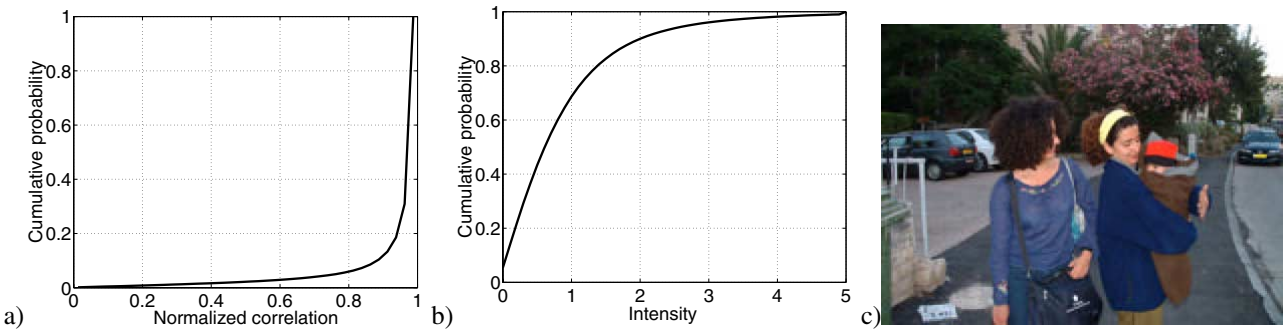

Figure 4. Measuring the validity of the local photometric affine model between different sensors. Fig.4-a Shows a typical cumulative histogram of normalized crossed correlation in regions where the gradient is larger than 10 values (for 8-bit images). Fig.4-b Show the cumulative histogram of prediction error of the local affine model. Fig.4-c Shows the image on which these statistics were computed. These results repeated on different images and among all pairs of RGB sensors.

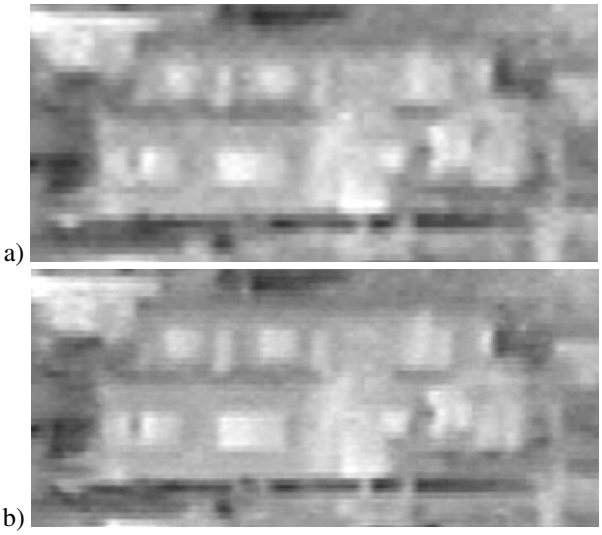

Figure 5. Results of multi-sensor super resolution on IR images of different spectral sensitivity. Fig. 5-a shows the input image (bilinearly interpolated), and Fig. 5-b shows the result of the multi-sensor super resolution.

demosaicing methods: Bilinear interpolation, Freeman's method [7] and color restoration [12]. The results are presented in Fig. 6. It can be seen that the presented method eliminates the zipper artifacts along the edges without oversmoothing the image. Note that most of the intensity information is captured by the green channel, for which the resolution enhancement is minor. Thus the process serves mainly as an anti-aliasing filter.

In a second experiment, we have enhanced a 2-5um IR image using 8-12um IR images of higher-resolution (Fig. 5). Due to the low SNR in the input images, we have used a larger neighborhood for the local affine model, with standard deviation 5.2.

4 Summary

This paper has introduced a new way to combine information from different image sensors. Given non-registered images captured by different sensors, the algorithm improves their resolution, using a local photometric affine alignment along the edges.

Acknowledgments

The authors would like to thank Minolta for providing raw samples from their camera, Elop and Eli Shechtman for providing IR input data and useful comments, and Danny Keren for applying his algorithm on our input.

References

[1] www.3dvsystems.com.
[2] P. Blomgrem and T. Chan. Color tv: Total variation methods for restoration of vector-valued images. *IEEE Trans. Image Processing*, 7(3):304–309, March 1998.
[3] P. Burt and R. Kolczynski. Enhanced image capture through fusion. In *ICCV93*, pages 173–182, 1993.
[4] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *ICPR*, pages Vol I: 600–605, September 2000.
[5] Y. Caspi and M. Irani. Alignment of non-overlapping sequences. In *ICCV*, pages II: 76–83, July 2001.
[6] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Processing*, 6(12):1646–1658, December 1997.
[7] W. Freeman. Median filter for reconstructing missing color samples, united states patent 4,724,395, 1998.
[8] B. Funt and J. Ho. Color from black and white. In *ICCV*, December 1988.
[9] M. Irani and P. Anandan. Robust multi-sensor image alignment. In *ICCV*, pages 959–966, January 1998.
[10] M. Irani and S. Peleg. Improving resolution by image registration. *GMIP*, 53:231–239, 1991.

Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)
0-7695-1858-3/02 $17.00 © 2002 IEEE

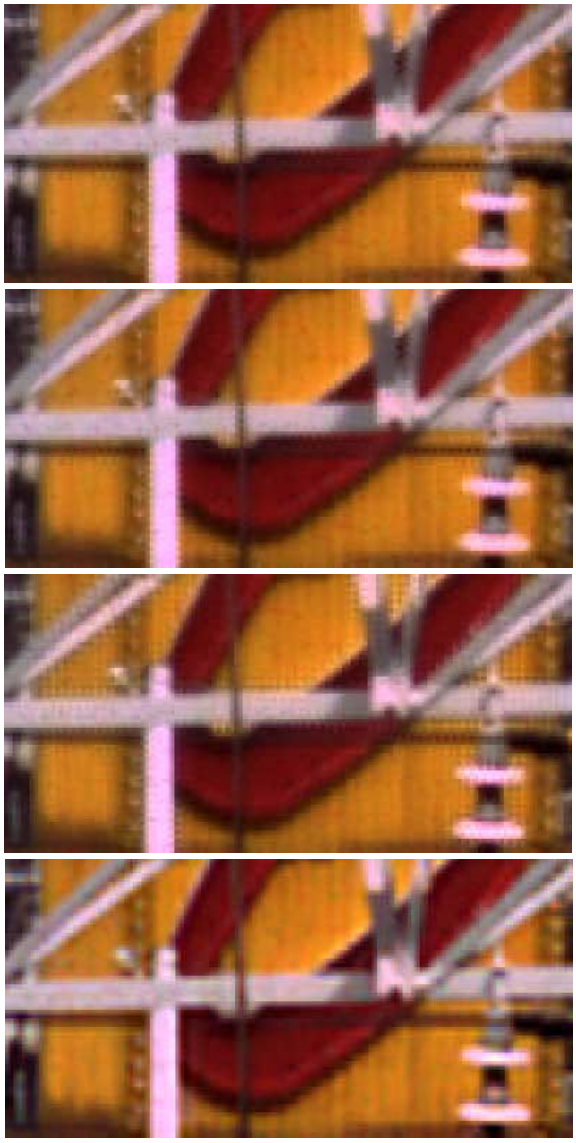

Figure 6. A comparison of multi-sensor super resolution to other demosaicing techniques on a Bayer pattern. a) Bilinear interpolation. b) Freeman's method [7]. c) Multi-channel restoration [12]. d) The presented multi-sensor super resolution. image d) contains no high-frequency zipper artifacts, and is sharper than than a) and b).

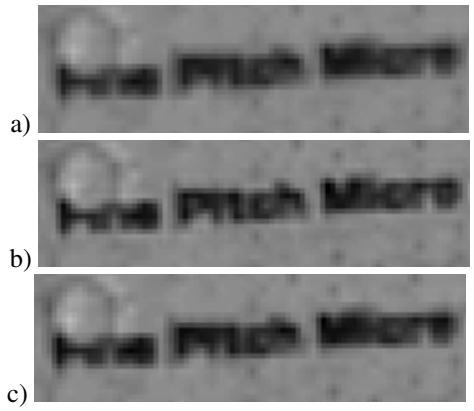

Figure 7. The results on the red image of the "Books" example. a) Magnified regions from the input red image, using bilinear interpolation. b) Multi-sensor Super resolution result. c) Single image super resolution (high-pass filter). The contrast of all images was enhanced for visualization.

[11] D. Keren and Y. Hel-Or. Image processing system using image demosaicing, patent ep1050847, 2000-11-08.

[12] D. Keren and M. Osadchy. Restoring subsampled color images. *MVA*, 11(4):197–202, 1999.

[13] R. Kimmel. Demosaicing: Image reconstruction from color ccd samples. *IEEE. Trans. IP*, 8(9):1221, September 1999.

[14] S. K. Nayar and S. G. Narasimhan. Assorted pixels: Multi-sampled imaging with structural models. In *ECCV*, May 2002.

[15] 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.

[16] R. Ramanath, W. Snyder, G. Bilbro, and W. Sander. Demosaicking methods for bayer color arrays. *Journal of Electronic Imaging*, 11(3), July 2002.

[17] R. Schultz and R. Stevenson. Stochastic modeling and estimation of multispectral image data. *IEEE Trans. Image Processing*, 4(8):1109–1119, December 1995.

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

[19] B. Tom and A. Katsaggelos. Multi-channel image identification and restoration using the expectation maximization algorithm, optical engineering, vol. 35, no. 1, pp. 241-254, jan. 1996.

[20] P. Viola and W. Wells, III. Alignment by maximization of mutual information. *IJCV*, 24(2):137–154, September 1997.

[21] A. Zomet and S. Peleg. Efficient super-resolution and applications to mosaics. In *ICPR*, volume I, pages 579–583, September 2000.

Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)
0-7695-1858-3/02 $17.00 © 2002 **IEEE**