ON THE IMPROVEMENT OF IMAGE REGISTRATION FOR HIGH ACCURACY SUPER-RESOLUTION

Michalis Vrigkas, Christophoros Nikou and Lisimachos P. Kondi

Department of Computer Science, University of Ioannina, Greece
{mvrigkas, cnikou, lkon}@cs.uoi.gr

ABSTRACT

Accurate image registration plays a preponderant role in image super-resolution methods and in the related literature landmark-based registration methods have gained increasing acceptance in this framework. However, their solution relies on point correspondences and on least squares estimation of the registration parameters necessitating further improvement. In this work, a maximum *a posteriori* scheme for image super-resolution is presented where the image registration part is accomplished in two steps. At first, the low-resolution images are registered by establishing correspondences between robust SIFT features. In the second step, the estimation of the registration parameters is fine-tuned along with the estimation of the high resolution image, in an iterative procedure, using the maximization of the mutual information criterion. Numerical results showed that the reconstructed image is consistently of higher quality than in standard MAP-based methods employing only landmarks.

Index Terms— Super-resolution, image registration, mutual information, scale invariant feature transform (SIFT).

1. INTRODUCTION

The objective of image super-resolution (SR) is to reconstruct a high-resolution (HR) image from a sequence of low-resolution (LR) images. The goal is to improve the spatial resolution by fusing the set of LR images to produce an image with more visible detail. The LR images experiences different degradations such as motion, point spread function blurring, subsampling and additive noise. The HR image is estimated from a sequence of LR aliased images which is possible if there exists sub-pixel motion between the LR images. Thus, each frame of the LR sequence brings complementary information on the original HR image.

The direct inverse solution from interpolation, motion compensation and inverse filtering is ill-posed due to the existence of additive noise, even in cases of perfect motion registration and accurate knowledge of the point spread function of the acquisition system. A large family of methods is based on a stochastic formulation of the problem which imposes a prior distribution on the image to be reconstructed and provides estimates either in a maximum *a posteriori* (MAP) framework, where the posterior distribution of the HR image is maximized [1, 2, 3, 4, 5, 6] or in a fully Bayesian framework by integrating out any unobserved variables [4, 7, 8, 9, 10, 11].

A key issue in the quality of the super-resolved image is the accuracy of the employed image registration technique. Also, knowledge of the involved motion model facilitates the task. This may include simple translational, rigid body or affine motion as well as projective or even photometric transformations. The standard approach is to estimate the registration parameters separately from the HR image [5, 6], either by aligning the LR images once, at the beginning of the algorithm or iteratively before or after each update of the HR image [1, 2, 3]. However, there exist techniques where the registration parameters are assumed to be random variables and they are marginalized in a Bayesian formulation [4, 7]. Apart from using block matching or phase correlation techniques, the majority of the registration methods used in the SR literature are related to standard optical flow methods and their variants.

Following the trends in computer vision, feature matching has also been used [5]. The parameters of the geometric transformation between the LR images are estimated by automatic detection and analysis of corresponding features among the input images. Typically, some hundreds of points of interest, such as the Harris corner features [5], are detected with subpixel accuracy and correspondences are established by examining the image neighborhoods around them. Finally, the estimation of the registration parameters is obtained by optimization of a non linear cost function.

Landmark-based registration is accurate but limited to least-squares based solutions. In the last 15 years, the maximization of the mutual information (MI) has revolutionized image registration theory and applications as it considers the whole gray level image information and provides consistently sub-pixel precision [12]. However, to our knowledge, it has not extended its application domain to image super-resolution. Probably, the main reason is that if it is not initialized close to the global maximum, local extrema impede the registration process [13] and, more importantly, they rule out subpixel accuracy.

Relying on the above observations, we propose a MAP scheme for image super-resolution where the registration part is accomplished in two steps. At first, the LR images are registered by establishing correspondences between robust features computed by the scale invariant feature transform (SIFT) [14]. In the second step, the estimation of the registration parameters is fine-tuned along with the estimation of the high resolution image, in an iterative scheme, by the maximization of the mutual information between the HR image and each of the upscaled (deblurred and upsampled) LR images. Numerical results showed that the reconstructed image is consistently of higher quality than in standard MAP-based methods employing only SIFT features and this improvement is on average 1.5 dB in terms of peak signal to noise ratio (PSNR).

2. IMAGE FORMATION MODEL

The image degradation process [3] is modeled by motion (rotation and translation), a linear blur, and subsampling by pixel averaging along with additive Gaussian noise. We assume that p LR images, each of size $M = N_1 \times N_2$, are obtained from the acquisition process. The following observation model is assumed, where all images are ordered lexicographically

$$\mathbf{y} = \mathbf{W}\mathbf{z} + \mathbf{n}. \tag{1}$$

The set of LR frames is described as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_p^T]^T$, where $\mathbf{y}_k$, for $k = 1, ...p$, are the p LR images. The desired HR image $\mathbf{z}$ is of size $N = l_1 N_1 \times l_2 N_2$, where l_1 and l_2 represent the up-sampling factors in the horizontal and vertical directions, respectively. The term $\mathbf{n}$ represents zero-mean additive Gaussian noise. In (1), the degradation matrix $\mathbf{W} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \ldots, \mathbf{W}_p^T]^T$ performs the operations of motion, blur and subsampling. Thus, matrix $\mathbf{W}_k$, for the k-th frame, may be written as

$$\mathbf{W}_k = \mathbf{D}\mathbf{B}_k\mathbf{M}(\mathbf{s}_k), \tag{2}$$

where $\mathbf{D}$ is the $N_1 N_2 \times N$ subsampling matrix, $\mathbf{B}_k$ is the $N \times N$ blurring matrix, and $\mathbf{M}(\mathbf{s}_k)$ is the $N \times N$ rigid transformation matrix with parameters (rotation angle and translation vector) denoted by $\mathbf{s}_k$ for the k-th frame.

Formulating the super-resolution problem in a probabilistic framework [1, 2, 3], we define a smooth Gaussian prior for the HR image:

$$p(\mathbf{z}) = \frac{(\alpha|\mathbf{Q}^T\mathbf{Q}|)^{N/2}}{(2\pi)^{N/2}} \prod_{i=1}^{N} \exp\left(-\frac{1}{2}\alpha(\mathbf{Q}\mathbf{z})^T(\mathbf{Q}\mathbf{z})\right), \tag{3}$$

where $\mathbf{Q}\mathbf{z}$ is the Laplacian of image $\mathbf{z}$ and parameter α controls the precision (inverse covariance) and consequently the shape of the distribution. Given the HR image $\mathbf{z}$ and the registration parameters between the LR images $\mathbf{s} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_k\}$, the likelihood of the LR images is also a Gaussian [3]:

$$p(\mathbf{y}|\mathbf{z}) = \frac{1}{(2\pi)^{\frac{pM}{2}} \sigma_\eta^{pM}} \exp\left(-\frac{(\mathbf{y} - \mathbf{W}\mathbf{z})^T(\mathbf{y} - \mathbf{W}\mathbf{z})}{2\sigma_\eta^2}\right), \tag{4}$$

where σ_η^2 is the variance of the observation noise $\mathbf{n}$.

Employing a MAP approach and maximizing $p(\mathbf{z}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z})$ leads to the following MAP functional to be minimized with respect to the HR image $\mathbf{z}$ and the rigid transformation parameters $\mathbf{s}$:

$$L(\mathbf{z}, \mathbf{s}) = \sum_{k=1}^{p} ||\mathbf{y}_k - \mathbf{W}_k(\mathbf{s}_k)\mathbf{z}||^2 + \lambda||\mathbf{Q}\mathbf{z}||^2. \tag{5}$$

where $\lambda = \alpha/\sigma_\eta$. Notice the change in notation to explicitly underpin the dependence of matrix $\mathbf{W}_k$ on the registration parameters $\mathbf{s}_k$.

Considering that the registration parameters $\mathbf{s}$ are known and using a gradient descent method with a properly calculated step size it can be shown that the update equation minimizing (5) can be written as

$$\hat{\mathbf{z}}^{n+1} = \hat{\mathbf{z}}^n - \varepsilon^n \nabla_{\mathbf{z}} L(\mathbf{z}, \mathbf{s})|_{\mathbf{z}=\hat{\mathbf{z}}^n} \tag{6}$$

Parameter ε^n is the step size at the n-th iteration which may be obtained in closed form from the data [1]. In general, the estimation of the regularization parameter λ which depends on the noise standard deviation σ_η^2, and the parameter α controlling the variance in the prior (3), is a difficult task. In order to avoid a blurred version of the high-resolution image these parameters are automatically computed from the data as described in our previous work [2, 3].

3. IMAGE REGISTRATION

A standard approach in MAP super-resolution algorithms is to register the LR images prior to the computation of the HR image. This is performed once and the registration parameters are fixed during the iterative estimation of the super-resolved image. A typical solution to the registration problem is the computation and correspondence of corner features [5]. Although the extracted features are robust, this procedure is prone to small registration errors as the registration parameters are computed in the least squares sense.

The maximization of mutual information, originally proposed for medical image registration, is considered to be one of the most accurate methods for image registration [12] as it provides subpixel accuracy. It relies on gray level information by considering each image as a random variable.

Let A and B be the two images with marginal probability density functions (computed from their histograms) $p_A(a)$ and $p_B(b)$ respectively. Let also their joint density be $p_{AB}(a, b)$. The mutual information between A and B measures the degree of dependence between them and it is defined by

$$\begin{aligned} I(A, B) &= H(A) + H(B) - H(A, B) \\ &= \sum_a \sum_b p_{AB}(a, b) \log \frac{p_{AB}(a, b)}{p_A(a) \cdot p_B(b)} \end{aligned} \tag{7}$$

where $H(A)$ and $H(B)$ are the marginal entropies of the random variables A and B and $H(A, B)$ is their joint entropy. If the images are correctly registered their mutual information is maximized.

In order to provide invariance to the overlapping areas between the two images, a more robust measure is the normalized mutual information (*NMI*) [15]:

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)}. \tag{8}$$

A drawback of the mutual information (and *NMI*) is that if it is not initialized close to the optimal solution it is trapped by local maxima [13]. To overcome this issue a good initialization is important.

Therefore, we propose to estimate the registration parameters in two steps. In the first step, the registration procedure is initialized by a landmark-based registration scheme. To this end, to register the LR images, we employ SIFT features [14] extracted from the LR images. SIFT features are generally more robust than corner features. Considering a LR image as the reference, the rigid transformation parameters (translation and rotation) are estimated through minimization of the mean square error between the locations of the features between the reference image and each LR image [16]. Thus, we obtain a good initialization.

In the next step, during the iterative update of the HR image, a fine tuning of the registration parameters is accomplished by the maximization of the mutual information between the current estimate of the HR image, provided by (6) and each upscaled LR image. Upscaling is performed by deblurring (inverse filtering) and upsampling. As the estimate of the HR image changes at each iteration, the registration parameters are updated based on this estimate. By these means, the registration accuracy is improved at each iteration step. The overall algorithm is summarized in Algorithm 1.

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed methodology, experiments were conducted on synthetic data sets. Sequences of low resolution images were created by rotating, translating, blurring, down-sampling

Algorithm 1 Super-resolution image reconstruction algorithm.

- Extract SIFT descriptors from the LR images and establish correspondences.
- Estimate rotations and translations using least squares [16].
- First estimate of the HR image $\hat{\mathbf{z}}^0$ by bicubic interpolation of a randomly selected LR image.
- $n := 1$; $\hat{\mathbf{z}}^n = \hat{\mathbf{z}}^0$;
- do
  - do
    - Random selection of a LR image $\mathbf{y}_k$.
    - Register by mutual information the upscaled $\mathbf{y}_k$ to $\hat{\mathbf{z}}^n$.
    - Update $\hat{\mathbf{z}}^n$ using (6) only for the the visited $\mathbf{y}_k$.
    - Declare $\mathbf{y}_k$ visited.
  - until all $\mathbf{y}_k$ are visited.
  - $n := n + 1$;
  - Declare all $\mathbf{y}_k$, $k = 1, ...p$ visited.
- until $\|\hat{\mathbf{z}}^{n+1} - \hat{\mathbf{z}}^n\| / \|\hat{\mathbf{z}}^n\| < \epsilon$.

and degrading by noise an original image. Translation parameters were randomly drawn from a uniform distribution in $[-3, 3]$ (in units of HR pixels) and rotation angles were also uniformly selected in $[-5, 5]$ (in degrees). The images were then downsampled by a factor of 2 (4 pixels to 1). Then, a point spread function of 5×5 Gaussian kernel with standard deviation of 1 was applied. Finally the resulting images were degraded by white Gaussian noise in order to obtain signal to noise ratios of (i) 20 dB and (ii) 30 dB. In all of the experiments, in order to have a first estimate of the HR image, a LR image was chosen at random and it was upscaled by bicubic interpolation. A quantitative evaluation of the obtained HR images is given by the peak signal to noise ratio (PSNR) defined by:

$$\text{PSNR} = 10 \log_{10} \frac{255}{||z - \hat{z}||}$$

where $\hat{z}$ is the estimated HR image and z is the ground truth.

The numerical results are summarized in Table 1, where the mean values, the standard deviations and the median values of the PSNR for each image are presented. These values are obtained through 10 random realizations of the experiment in each case. As it can be seen, the combination of SIFT-based initialization of the registration parameters followed by fine tuning by the maximization of the MI criterion provides consistently higher accuracy in terms of PSNR. This improvement is approximately 1.5 dB on average for both noise scenarios (20 dB and 30 dB).

Convergence of the super-resolution algorithm was achieved when $\|\hat{\mathbf{z}}^{n+1} - \hat{\mathbf{z}}^n\| / \|\hat{\mathbf{z}}^n\| < 10^{-5}$. Another advantage of the proposed scheme is that not only the reconstructed HR image is of better quality but also the algorithm converges faster. This is depicted in Figure 1 where the cost function (5) is drawn with respect to the iteration number for the compared methodologies. Representative results of the reconstructed HR images along with a LR frame is shown in Figure 2.

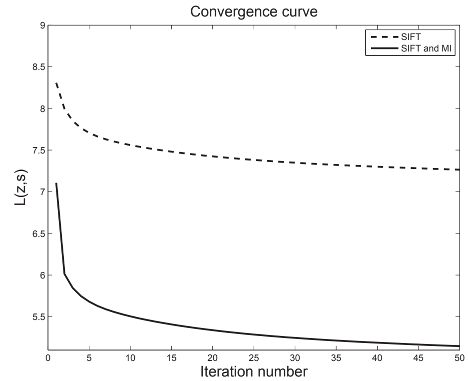

Fig. 1. The cost function $L(\mathbf{z}, \mathbf{s})$, with respect to the iteration number for the compared methods.

5. CONCLUSION

We proposed an image registration framework that improves the performance of super-resolution image reconstruction. The hybrid approach is based on the synergy of SIFT-based image registration whose result is forwarded to a maximization of mutual information algorithm. The first step provides a robust least squares parameter estimation and the second step of the method achieves a a high precision registration result. By these means, the main drawback of mutual information, which is the large number of local maxima is overcome. Therefore, a solution of high accuracy is obtained for the super-resolved image and the overall reconstruction algorithm converges faster than the standard solution based only on landmark correspondence and registration [5].

Finally, let us notice that we have also tried to register the LR images by the MI method only, without initialization by the SIFT-based registration. In all cases the resulting estimation of the registration parameters was erroneous leading to a HR image of very low quality.

6. REFERENCES

[1] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP image registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.

[2] H. He and L. P. Kondi, "Resolution enhancement of video sequences with simultaneous estimation of the regularization parameter," *SPIE Journal of Electronic Imaging*, vol. 13, no. 3, pp. 586–596, 2004.

[3] H. He and L. P. Kondi, "An image super-resolution algorithm for different error levels per frame," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 592–603, March 2006.

[4] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Overcoming registration uncertainty in image super-resolution: Maximize or marginalize?," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. Article ID 23565, 14 pages, 2007.

[5] D. Capel and A. Zisserman, "Computer vision applied to super-resolution," *IEEE Signal Processing Magazine*, 2003.

[6] G. K. Chantas, N. P. Galatsanos, and N. A. Woods, "Super-resolution based on fast registration and maximum a posteriori

Table 1. PSNR statistics (in dB) for the compared super-resolution methods.

Image	30 dB additive Gaussian noise						20 dB additive Gaussian noise					
	SIFT			SIFT and M. I.			SIFT			SIFT and M. I.		
	mean	std	median	mean	std	median	mean	std	median	mean	std	median
Books	25.22	0.28	25.09	**25.34**	0.42	**25.45**	23.45	0.19	23.44	**23.64**	0.31	**23.52**
Front page	24.34	0.34	24.31	**26.78**	0.41	**26.52**	22.39	0.30	22.34	**24.83**	0.36	**24.85**
Car	26.67	0.41	26.45	**27.80**	0.59	**27.69**	21.23	0.19	21.22	**21.56**	0.19	**21.47**
Eye chart	25.47	0.42	25.35	**27.01**	0.11	**27.06**	23.30	0.34	23.24	**24.59**	0.26	**24.64**

(a) *Books* (PSNR = 26.06 dB)

(b) *Front page* (PSNR = 26.14 dB)

(c) *Car* (PSNR = 28.13 dB)

(d) *Eye chart* (PSNR = 27.33 dB)

Fig. 2. Representative low resolution degraded observations (SNR=30 dB) and the reconstructed high-resolution images using the proposed method.

reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1821–1830, July 2007.

[7] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," in *Advances in Neural Information Processing Systems*. 2006, pp. 1089–1096, MIT Press.

[8] M. E. Tipping and C. M. Bishop, "Bayesian image super-resolution," in *Advances in Neural Information Processing Systems*. 2003, pp. 1303–1310, MIT Press.

[9] A. Kanemura, S. Maeda, and S. Ishii, "Edge-preserving Bayesian image superresolution based on compound Markov random fields," in *Proceedings of the 17th International conference on Artificial Neural Networks (ICANN)*, Porto, Portugal, 2007.

[10] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Transactions on Image Processing*, vol. 13, pp. 898–911, 2004.

[11] S. Farsiu, M. Elad, and P. Milanfar, "Multi-frame demosaicing and super-resolution of color images," *IEEE Trans. on Image Processing*, vol. 15, pp. 141–159, 2006.

[12] P. Viola and W. M. Wells, "Aligment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[13] J. Pluim, A. Maintz, and M. Viergever, "Interpolation artefacts in mutual information-based image registration," *Computer Vision and Image Understanding*, vol. 77, no. 9, pp. 211–232, 2000.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[15] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image aligment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[16] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 376–380, 1991.