J. Zhang and X. Yuan: Analysis of segmentation performance on super-resolved images. In *Texture 2005: Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*, pp. 155–160, 2005.

Analysis of Segmentation Performance on Super-resolved Images

Jian Zhang
Department of Computer Science
West Virginia University Inst. of Tech.
Montgomery, WV 25136, USA

Xiaohui Yuan
Department of Elec. Eng. and Comp. Sci.
Tulane University
New Orleans, LA 70118, USA

Abstract

Image super-resolution is to reconstruct one or more high-resolution images from a set of low resolution images. However, subjective methods based on human visual comparison have been used to evaluate the reconstruction quality. To quantitatively analyze the performance of super-resolution method, we applied False Negative Ratio metric on the result images. In addition, we calculated Receive Operating Characteristic probabilities to quantify the impact of super-resolved images on the segmentation accuracy. We demonstrated that a small set of inputs are sufficient to achieve the sub-optimal segmentation precision and little improvement is brought by using extra source images.

(a) (b)

Figure 1. (a) is simulated low resolution frame 1 ($L_1 = L_2 = 4, \sigma_n^2 = 100$). (b) ALE estimation with 8 low resolution frames.

1. Introduction

The segmentation of images into meaningful and homogenous regions is a key method for image analysis within applications such as content based retrieval [12, 14] and automatic target recognition [21]. However, if the imaging detector array is not sufficiently dense, the resulting images are probably degraded. In addition, blurring factors and discretizing process also hamper reliable segmentation and subsequent analysis. For example, autonomous road following based on segmentation is particularly difficult when the motion is fast, noise levels are high, and the computation needs to happen in real time.

If the low resolution images are captured from the same scene and provide slightly different "looks" at the same object, image super resolution can properly reconstruct a high resolution image from the series of aliased low resolution frames 1. In Figure 1, (a) is one of simulated 16 low resolution frames and (b) is the super resolution result from 8 low resolution frames. As illustrated, this process boosts the image quality and reduce or complete solve the aliasing [26, 13, 10, 3, 5, 9].

In real-time military applications such as Automated Target Recognition (ATR), it is critical to fulfill detection, segmentation and recognition tasks in a timely manner. Therefore, we expect super resolution method to provide a good input for segmentation and subsequent analysis without using too much time. Currently, most researchers use a set of low resolution frames, which number ranges from 4 to 55 [10, 9, 7], to produce one super resolved image. Super resolution results are provided for visual comparison but there are several questions left unanswered. For example, how well does super resolution image recover the features from low resolution images, can super resolution image provide better segmentation result compared to that on low resolution images, and how many low resolution images are necessary to generate comparable segmentation result to that on original high resolution image. This paper endeavors to answer these questions through a set of well designed experiments and set up a framework to compare different super resolution algorithms in future work.

We organize this article as follows. In section 2, we briefly review the edge detection and image segmentation techniques. In this section, we introduce image super resolution methods, specifically the Anti-Lamenessing Engine as well. We describe three performance evaluation methods that hierarchically examine super resolved images from image level, object level to feature level in section 3. In

section 4, experiments are performed on super resolution images and analysis are given. This article is concluded in section 5.

2 Background

2.1. Edge Detection

In order to segment an object from the background, edge detection is needed primarily to establish accurate borders. The human perceptual system places great importance on edges. It has been suggested that one of the most fundamental organizing principles of the visual system is the detection and description of discontinuity.

Roberts [22] proposed the first widely-used edge detector based on taking the first spatial derivative. His process consisted of convolving a pair of masks approximating the first spatial derivative with the source image and then identifying responses with magnitudes above a set threshold. Two improvements on his approach were proposed by Sobel [25] and Prewitt [20].

An alternative approach is to search for zero crossings in the second spatial derivative. The Laplacian, $\bigtriangledown^2 f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2$, was introduced and shown to be isotropic by Rosenfeld and Thurston [24]. Filters discussed thus far have been linear FIR filters. Their initial development was universally in the spatial domain.

A more advanced and widely used edge detector was developed by Canny [4]. Building on the first spatial derivative approach, the Canny detector takes a source image, applies Gaussian smoothing, then applies a first spatial derivative operator to create "ridges" corresponding to edges in the source image. The detector then uses non-maximum suppression, zeroing all but the maximum pixels along each ridge. Finally, hysteresis is adapted to join edge fragments.

We use Canny detector in this paper to examine the feature retaining performance of super resolution images. By calculating the False Negative Ratio (FNR) [30] on Canny detected edges, we show how well the feature/edge is recovered by iteratively fusing multiple low resolution images.

2.2. Image Segmentation

Image segmentation is an important research area in computer vision and hundreds of segmentation algorithms have been proposed in the last 30 years. Obviously, a detailed review on segmentation methods is out of the scope of this paper. However, many segmentation methods are based on two basic properties of the pixels in relation to their local neighborhood: discontinuity and similarity. Methods based on pixel discontinuity are called contour-based methods, whereas methods based on pixel similarity are called region-based methods. Nevertheless, contour-based or region-based methods alone often fail to produce accurate segmentation results [19]. In fact, color and texture are fundamental features in defining human visual perception. Hence, complementary information such as brightness, color and texture is taken into consideration [15, 16, 17].

Image segmentation based on texture is to partition an image into homogeneous regions and identify the boundaries which separate regions of different texture. One of widely used segmentation methods is to apply filters on image and calculate the gradient in texture feature space. Based on the observation that only several distinct filter responses are needed to represent textures, Malik etc. [16, 17] group the filter responses into a small set of prototype response vectors which are called "textons". Thereafter, they developed a texture gradient algorithm using these textons to generate a soft boundary map for image segmentation.

To provide a scientific basis for research on image segmentation, UC Berkeley Computer Vision Group maintains a segmentation dataset and benchmark [18, 1]. The human segmented test images provide the ground truth boundaries. Comparing texture based segmentation results on super resolution images to the ground truth boundaries, our experiments in this paper are performed on images taken from their dataset.

2.3. Super Resolution and ALE

In last two decades, several super resolution methods are proposed in the video processing, video compression, and image processing areas and are receiving continuous attention in machine vision, pattern recognition and remote sensing [26, 13, 6, 10, 3, 5, 9, 7].

Combining information from multiple low resolution images to form a single high resolution image. The widely used observation model for super resolution is in the form:

$$Y[m,n] = [H_{img}X(x,y)] \downarrow + n[m,n] \qquad (1)$$

where H_{img} is the imaging system, $X(x,y)$ is the unknown high resolution image, $\downarrow$ is the discretizing operator, $n[m,n]$ is additive noise, and $Y[m,n]$ is the observed discrete noisy and blurred image.

The characteristic of these methods is that they all make use of the imaging and motion information from a set of observed low resolution images. The difference among those images is represented by subpixel shifts. By estimating the subpixel displacement, back-projection [13] and bayesian methods [6, 10] are applied to estimate a high resolution image.

Since the objective of this paper is to evaluate the performance of super resolution method, we use an image processing software Anti-Lamenessing Engine (ALE) [11] to super resolve low resolution images without describing its

technical mechanism in detail. Interested readers are encouraged to visit ALE website for more information.

3. Performance Evaluation

The performance of the super resolution procedure is essentially twofold, that is, processing speed and image quality. This work attempts to show that in time critical systems, how the number of low resolution images affects feature preservation and image segmentation. In this section, we describe three performance evaluation methods that hierarchically examine super resolved images from image level, object level to feature level.

3.1. Image Level Evaluation

On image level, a reference image is used to compare with the super resolution result. This kind of comparison has been widely used in many applications such as image compression and communications. Examples include mean square error (MSE), Equation 2 and mutual information (MI), Equation 3. A detailed survey can be found in [8] and the references therein.

$$MSE = \frac{\sum_{(j,k)} |\mathcal{I}_r(j,k) - \mathcal{I}_f(j,k)|^2}{M \times N} \tag{2}$$

where $\mathcal{I}_r$ and $\mathcal{I}_f$ are the reference image and the super resolution result respectively. $M \times N$ denotes the size of images. $h(\cdot)$ is the histogram function.

Given two images A and B, *mutual information* is defined in terms of entropy as follows

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{3}$$

where $H(A)$ (or $H(B)$) is the Shannon entropy of image A (or B)

$$H(A) = -\sum_{a \in A} P(a) log P(a)$$

and $H(A,B)$ is the joint entropy of image A and B.

$$H(A,B) = -\sum_{a \in A, b \in B} P(a,b) log P(a,b)$$

Assume we have two images as input, mutual information is maximized when they are identical images and the root MSE is minimized at the same time.

Due to the clear physical meanings of these metrics, they have been adopted to evaluate the quality of image registration [28, 27, 23]. The reference image is usually assumed available. We use the original high resolution image as the reference image.

3.2. Texture based Image Segmentation

As mentioned in 2.2, a local boundary model is essential to higher level image segmentation algorithms, no matter whether they are based on grouping pixels into regions or grouping edge fragments into contours. However, Canny detector models boundaries as brightness step edge. It is unable to detect boundary between textured regions where there is slight change in average brightness.

A common practice in texture analysis is to apply a bank of filters on an image and the filters are tuned to various orientations and spatial frequencies. After filtering, a texture descriptor is constructed using the empirical distribution of filter responses in the neighborhood of a pixel. This approach is employed to identify feature and object in natural scene images [31, 29].

Martin etc. [17] used a filter bank consists both even and odd filters at multiple orientations as well as a radially symmetric center-surround filter. A vector of filter responses is associated with every pixel in the image. These vectors are clustered using k-means and the cluster centers are called *textons*. Each pixel is therefore assigned to one of the textons. To measure the degree of texture variation at a location (x, y) in direction θ, they use a disk of radius s centered on (x, y) and divided in two along the diameter at direction θ. Texture similarities can then be computed by comparing the histograms of textons in the two disc halves.

We use the same method to calculate texture gradient [17]. Let g_i and h_i represent the number of pixels of texton type i in each half disk, texture gradient is defined as the χ^2 distance between these two histograms

$$\chi^2(g,h) = \frac{1}{2} \sum \frac{(g_i - h_i)^2}{g_i + h_i} \tag{4}$$

The resulting soft boundary map from texture gradient method contains boundaries valued from zero to one where high values means greater confidence in the existence of a boundary. This soft boundary map is converted to binary boundary map by thresholding. Because we use images from UC Berkeley segmentation dataset, we are able to choose this threshold from their benchmark. We then compare the binary boundary map to the ground truth boundaries and calculate the true positive and false positive probabilities for boundary map matching.

Let the number of corrected detected boundary pixels ("true positive") be $N(TP)$, the number of incorrectly detected boundary pixels ("false positive") be $N(FP)$, and the total number of boundary pixels in ground truth boundaries be $N(B)$. The probability of detected boundary being correct and the probability of detected boundary being incorrect are computed as:

$$P(TP) = N(TP)/N(B) \tag{5}$$

and

$$P(FP) = N(FP)/N - N(B) \quad (6)$$

where N is the total number of pixels in the entire image.

3.3. Feature Recovery Analysis

Besides segmentation based on image texture, we are concerned about how features are recovered in the super resolution images. In general, edges are considered important image features and have been used for evaluating image quality [2].

Using the reference image, we are able to quantify the amount of erroneous feature on the super resolution image. By analogy, erroneous features can be categorized into false negative features and false positive features. False negative features are features that exist in the reference image but diminish in the super resolution image, whereas false positive features are those that appear in the super resolution image but not in the reference. The false positive features are essentially noise and its quantification overlaps with the segmentation analysis stated in the previous section. Hence we focus on only the false negative features. We propose a new quality metric, the false negative ratio, as follows.

False Negative Ratio

Given a reference image $\mathcal{I}^*$, an edge map, denoted as ψ^*, is extracted, where edge positions are marked as 1 and 0 elsewhere. Similarly, edge maps of processed images are generated and denoted as ψ_j, j is the index of images. For a pair of edge maps, $\{\psi^*, \psi_j\}$, the False Negative Ratio (FNR) is computed as follows

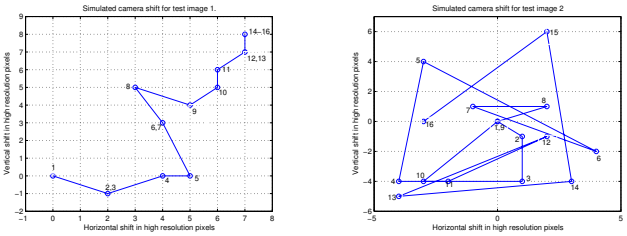
$$\text{FNR} = \frac{\sum_{\psi^* - \psi_j > 0} (\psi^* - \psi_j)}{\sum \psi^*} \quad (7)$$

FNR measures the amount of missing features in an image. The smaller the FNR value, the better the feature retaining performance. Two extreme cases of FNR are 1 and 0, which correspond to the total lose of image feature and the optimal feature recovery.

4. Experiments

Two simulated sequences of random translational shifts are shown in Figure. 4 and are applied to two images chosen from the Berkeley dataset, which are shown in Figure. 2(a) and Figure. 3(a), respectively. Note that these translational shifts are given in terms of high-resolution pixel spacing. Using these shifts, a sequence of 16 translated images is generated. The original image is blurred to simulate the low resolution and subsampled by a factor of four on horizontal and vertical directions. Gaussian noise with $\sigma_n^2 = 100$ is also added to each low resolution frame.

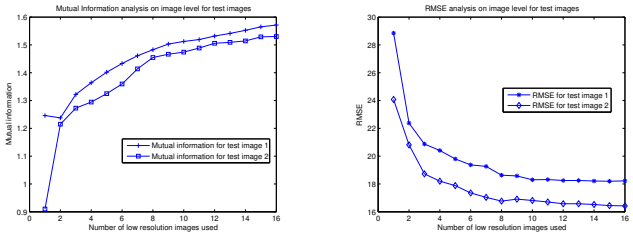
For comparison with the multiframe super resolution, the images formed with bicubic interpolation of the single frame are shown in Figure. 2(b) and Figure. 3(b), respectively. Finally, Figure. 2(c) and Figure. 3(c) show the high resolution image by super resolving 4 low resolution images. An obvious visual improvement is observed compared to bicubic interpolation of single frame.



(a) (b)

Figure 4. Simulated camera shifts in terms of high resolution pixels for 15 frames. (a) Shifts applied on Figure. 2(a); (b) shifts applied on Figure. 3(a).

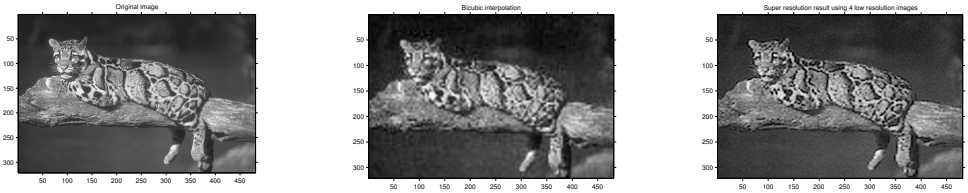
Quantitative analysis of super resolution using different numbers of frames are provided in Figure. 5, Figure. 6 and Figure. 4.



(a) (b)

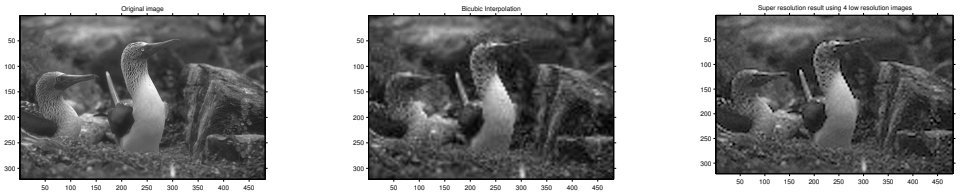
Figure 5. (a)RMSE for super resolution results with different number of frames. (b)Mutual information for super resolution results with different number of frames.

With only one frame, the ALE's performance is comparable to that of the bicubic interpolator. However, with additional frames, ALE estimate becomes significantly improved with respect to the single frame interpolator on image level, object level and feature level. However, the relationship between CPU time usage of ALE and the numbers of low resolution images can be approximated by a linear function, which means the more low resolution images the longer user has to wait to get a high resolution output. By studying the performance evaluation illustrated in Figure. 5,



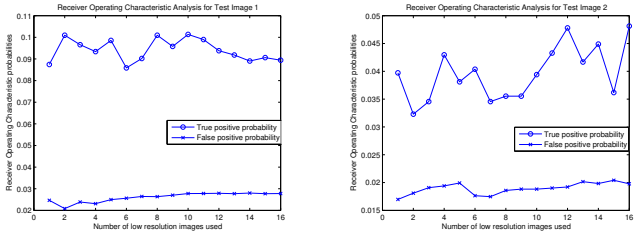
(a) (b) (c)

Figure 2. (a) Original image. (b) Bicubic interpolation of frame 1. (c) ALE estimate with 4 low resolution frames.



(a) (b) (c)

Figure 3. (a) Original image. (b) Bicubic interpolation of frame 1. (c) ALE estimate with 4 low resolution frames.



(a) (b)

Figure 6. Receiver Operating Characteristic probabilities for super resolution results with different number of frames. (a) Results for test image 1. (b) Results for test image 2.

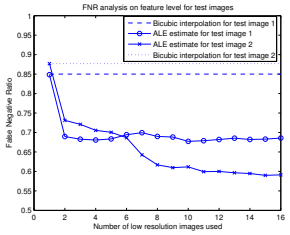


Figure 7. False Negative Ratio for super resolution results with different number of frames.

Figure. 6 and Figure. 4, we believe that a number of eight low resolution images is adequate to achieve the satisfying result for image segmentation based on texture features and that result is comparable to the same operation on the original high resolution image.

5. Conclusions

In this paper, we describe a framework that can systemically evaluate the performance of super resolution methods. On the image level, we compute the mutual information and RMSE between original image and iteratively generated super resolution images on the highest level. Then we calculate the Receiver Operating Characteristic probabilities on object level and False Negative Ratio on feature level. Our results show that super resolution can improve the precision of segmentation based on texture and image feature recovery. However, not many low resolution images are needed to achieve this performance. This is a very important induction for real-time applications such as battle field ATR and security surveillance system.

Our experiments are conducted using ALE software. However, different super resolution methods can easily be compared with each other using the methods described here. Also, the results presented are from image segmentation

aspect of view. How does super resolution affect the object identification and target recognition results are not discussed. We plan to study super resolution methods from these applications' viewpoint in the future work.

References

[1] The berkeley segmentation dataset and benchmark. http://www.cs.berkeley.edu/projects/vision/grouping/segbench/.

[2] B. Avcibas, B. Sankur, and K. Sayood. Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, pages 206–223, 2002.

[3] S. Borman and R. Stevenson. Super-resolution from image sequences - a review. Technical report, University of Notre Dame, 1998.

[4] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, Nov. 1986.

[5] D. Capel and A. Zisserman. computer vision applied to super resolution. *IEEE Signal Processing Magazine*, pages 75–86, 2003.

[6] P. Cheesman, B. Kanesky, R. Kraft, and J. Stutz. Super-resolved surface reconstruction from multiple images. Technical report fla-94-12, NASA Ames Research Center, 1994.

[7] F. J. Cortijo, S. Villena, R. Molina, and A. Katsaggelos. Bayesian super-resolution of text image sequences from low resolution observations. In *IEEE 7th International Symposium on Signal Processing and its Applications*, pages 421–424, 2003.

[8] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, Dec. 1995.

[9] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004.

[10] R. C. Hardie, K. J. Barnard, and E. E. Armstrong. Joint map registration and high resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(12):1621–1633, 1997.

[11] D. Hilvert. Anti-lamenessing engine (ALE). http://auricle.dyndns.org/ALE/.

[12] W. Hsu, S. T. Chua, and H. H. Pung. An integrated color-spatial approach to content-based image retrieval. In *MULTIMEDIA '95: Proceedings of the third ACM international conference on Multimedia*, pages 305–313, 1995.

[13] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graph. Models Image Processing*, 53(3):231–239, 1991.

[14] A. Kam, T. Ng, N. Kingsbury, and W. Fitzgerald. Content based image retrieval through object extraction and querying. In *MULTIMEDIA '95: Proceedings of the third ACM international conference on Multimedia*, pages 91–95, 2000.

[15] W. Y. Ma and B. S. Manjunath. Edge flow: a framework for boundary detection and image segmentation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 744–749, 1997.

[16] J. Malik, S. Belongie, T. K. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.

[17] D. Martin, C. Fowlkes, and J. Malik. Learning to detect image boundaries using brightness and texture. In *Proceedings of NIPS*, pages 1255–1262, 2002.

[18] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[19] T. Pavlidis and Y. Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 225–233, 1990.

[20] J. M. S. Prewitt. Object enhancement and extraction. In B. S. Lipkin and A. Rosenfeld, editors, *Picture Processing and Psychopictorics*. Academic Press, New York, 1970.

[21] C. Rasmussen. Combining laser range, color, and texture cues for autonomous road following. In *IEEE International Conference on Robotics and Automation*, pages 4320–4325, 2002.

[22] L. G. Roberts. Machine perception of three-dimensional solids. In T. J. Tippet, editor, *Optical and Electro-Optical Information Processing*. MIT Press, Cambridge, MA, 1965.

[23] M. Robinson and P. Milanfar. Fundamental performance limits i image registration. In *Proceedings of the 2003 IEEE International Conference on Image Processing (ICIP'03)*, pages 323–326, Barcelona, Spain, 2003.

[24] A. Rosenfeld and M. Thurston. Edge and curve detection for visual scene analysis. *IEEE Trans. on Computers*, 20(5):817–830, May 1971.

[25] I. E. Sobel. *Camera Models and Machine Perception*. PhD thesis, Stanford University, Palo Alto, CA, 1970.

[26] R. Tasi and T. Huang. Multiframe image restoration and registration. In *Advances in Computer Vision and Image Processing*, pages 317–339, 1984.

[27] P. Thévenaz and M. Unser. An efficient mutual information optimizer for multiresolution image registration. In *Proceedings of the 1998 IEEE International Conference on Image Processing (ICIP'98)*, volume I, pages 833–837, Chicago IL, USA, 1998.

[28] S. van Engeland, P. R. Snoeren, J. Hendriks, and N. Karssemeijer. A comparison of methods for mammogram registration. *IEEE Transactions on Medical Imaging*, pages 1436–1444, 2003.

[29] J. Yuan, J. Zhang, X. Yuan, and B. P. Buckles. Multi-scale feature identification using evolution strategies. *Image and Vision Computing*, pages 555–563, 2005.

[30] X. Yuan and B. P. Buckles. Adaptive image fusion. *IEEE Trans. on Systems, Man and Cybernetics, B*, submitted 2005.

[31] X. Yuan, J. Zhang, and B. P. Buckles. Evolution strategie based image registration via feature matching. *Information Fusion*, pages 269–282, 2004.