

Super Resolution: A Database Driven Inference Approach

Monica TRIFAS, Jeremy STRAUB

Department of Mathematical, Computing and Information Sciences

Jacksonville State University

700 Pelham Road N., Jacksonville, AL 36265, USA

ABSTRACT

A revised super-resolution technique is presented and evaluated in this paper. This technique incorporates prior knowledge gained via training subject domain specific or general purpose images prior to presenting an image for super-resolution image resolution enhancement. When presented with an image to enhance, the engine selects candidate patterns using a low-resolution search mechanism and then uses the higher source-image resolution to select a winning candidate for inclusion in the super-resolved image. This proposed technique is evaluated from an application-agnostic perspective using several evaluation metrics. The failure of common super-resolution evaluation metrics to adequately measure application goal success is discussed. This failure is also directly tied to engine performance as, in many cases, the success metric and the pattern selection metric are tightly aligned. Thus, by developing new super-resolution evaluation metrics and other enhancements, not only can evaluation of training-based super-resolution be improved but also the super-resolved image product as well.

Keywords: digital images, resolution, super-resolution, evaluation of super-resolution, image enhancement.

1. INTRODUCTION

Super resolution techniques, generally, produce one or a set of high-resolution images from one or a sequence of low-resolution frames. Super-resolution has applications in machine vision, remote sensing, intelligence, medical and a plethora of other fields. Each application has its own requirements; however, the techniques that can be used to enhance an image are applicable across multiple application fields.

A super resolution technique is, generally, evaluated based on its ability to produce a replica of the original high resolution image. Many studies involve down-scaling images, enhancing the produced lower-resolution versions and comparing the super-resolved output image to the original on a pixel-by-pixel basis. Alternate approaches, which take in to account the desired application outcome may, however, be more suitable metrics.

2. BACKGROUND

Image resolution refers to the number of pixels contained in an area of an image. The resolution of a digital image is determined by how it is collected (resolution of camera, etc.) and any transformations applied to it (e.g., resizing, etc.). The physical resolution of an image refers to the actual size of the image as stored on disk. Images can have a lower effective resolution than their physical resolution if they have been

blurred or resized (without super resolution techniques) to a larger size than the original. Super resolution is used to increase the effective resolution of the image (or set of images).

Super resolution approaches can be classified in to two main categories: those which infer data and those which simply piece together existing data. Inference-based super resolution may draw on patterns in the image format, prior knowledge or other sources to create an output image that is similar (to some extent) to what would be produced by capturing the original scene or subject at the output resolution. Piecing solutions [e.g. 3, 4], on the other hand, use several images of the scene or subject and take advantage of camera shifts or movement to allow a higher resolution image to be constructed from several lower resolution ones. Most piecing techniques are computationally complex and numerically ill-posed problems. However, early work [e.g., 7] in this area showed that the aliasing effects in the high-resolution fused image can be reduced or removed, if a relative sub-pixel motion exists between the under-sampled input images. Both approaches produce an approximation of the actual high resolution image. Problems can include false positive inclusion and false negative exclusion.

3. DISCUSSION OF PROPOSED METHOD

The proposed approach makes an inference from a single source image based on prior knowledge. It uses a database of known patterns that it is trained with prior to being presented with the source image. Each database record contains a high, medium and low resolution version of the applicable pattern. When an image is presented for super resolution, each area of the image is compared to the low and medium size patterns in the database and the best match is selected to be the large size pattern in the final image. The proposed approach is loosely based on [5].

A. Pattern Sizing

A critical set of decisions that must be made with this super resolution approach is the size of the small, medium and large patterns. These patterns can, theoretically, be of any size; however, for the super resolution methodology to be effective (for most applications), pattern size should be quite small compared to the features in the image. The pattern should represent a part of an image feature (so that patterns will recur throughout various similar features) as opposed to covering a whole feature or (worse) a set of features. In a remote aerial/satellite sensing application, for example, the pattern could be a small part of a building or terrain feature.

Because the proposed technique involves inference, it may create false positives or false negatives when the closest match is not an accurate representation of the true image. If a pattern comprised the whole image feature or set of features, this miss-

match could result in the inclusion of a non-existent feature in to the image, the removal of an actually-existing feature, or the replacement of one feature with another. By keeping the pattern size small, relative to the size of image features, the amount of damage that a single miss-match can do to an image is minimized which decreases the likelihood of an undetected false negative or false positive.

Given this, it is necessary to choose the size of the patterns with the details of application and source images in mind. Multiple databases of trained patterns may be required to handle images with different effective resolutions and feature sizes.

B. Training

During training, one or more high resolution images are presented to the system and used to populate a database of corresponding high, medium and low resolution patterns. Thus, each source image is split into sections of the large pattern size; medium and low sized patterns are produced by down-scaling the large size pattern. The software iterates through the image (from left to right and top to bottom) with a one pixel offset for each row and column. Thus, the image will produce significantly more than $H/L \times W/L$ (where H is the image height, W is the image width and L is the large pattern size) patterns. Allowing for the fact that incomplete patterns will not be produced on the bottom and right of the image, the total number of patterns that will be collected from a source image is:

$$N = (H - (L - 1)) \times (W - (L - 1)) \quad \text{Eq. (1)}$$

It is necessary to collect all of the possible patterns (as opposed to simply cutting the image up so that each pixel is part of only a single pattern, a technique referred to as simple slicing) as there is no reason to presume that the image that will later be presented will match the artificial grid created by a simple slicing approach. Thus, simple slicing, would potentially fail to collect the patterns that may later serve to provide the best match as part of the image enhancement process.

In many cases, training will involve the presentation of multiple images. From a logical perspective, the training process involves simply adding additional patterns to the database from each image that is presented. The actual implementation of this, however, must be somewhat different. This is due to the fact that the training time for each subsequent pattern increases as the database must be searched to ensure that the pattern is non-duplicative (and, potentially, to increment an occurrence counter if a duplicate pattern is detected). This becomes particularly time consuming when the database size increases to the point where the index searches are requiring disk access (as opposed to simply searching data in RAM). To resolve this, all patterns could be loaded without checking for duplication and then duplicates could be removed (and occurrence counts set) later. Alternately, the approach that was taken was to train a single image at-one-time and then use the SQL MERGE command to combine the single image databases (and increment the occurrence counter, if desired).

The use of the merging approach has several benefits. First, it eliminates the need to create a routine to search for duplicates and to disable and later re-enable and populate the database index. Second, and most importantly, it allows the various training image datasets to be combined in various combinations for testing and evaluation without requiring re-training.

For a deployment of the proposed method, the database could be customized to contain patterns particular to the specific application, patterns from a wide variety of sources (for general use), or a hybrid of the two. The use of a customized database would, generally, be recommended for applications where images are constrained to one subject. For example, there would be little value in having up-close pictures of plant life in a remote sensing application as they would likely not match (or produce an erroneous match). Some previous research [e.g. 10], however, has demonstrated that patterns can recur at various levels of nature and human construction. This could, potentially, make patterns from one source useful in another. A general database, conversely, would be of greatest use when a particular image subject is not enhanced frequently enough to warrant the creation of a specialized database or a specialized database is not available. A hybrid approach combines a specialized database with a general one. This allows superior coverage of the specialized subject and provides for coverage of elements of the image that are not included in the specialized database.

C. Image Enhancement

When an image is presented for super resolution, it is processed through in $M \times M$ sized chunks (where M is the size of the medium pattern). In each instance, the medium pattern from the source image is resized to the small size. This produced small-size pattern is used to search the database for candidate match patterns. These patterns are selected by selecting patterns which fall within a range of $V_s \pm T$ (where V_s is the source image value and T is the selected tolerance value) on a pixel by pixel basis (that is, each pixel in the pattern must be within the tolerance range of the source image pattern for it to be selected as a candidate). If a set of candidates is not found within tolerance, the tolerance value is incrementally increased until a match or matches are found (once an expanded tolerance value is set, all matches at that level of tolerance, not just the first identified, are considered).

Once candidate patterns are identified, the medium sized image for each candidate pattern is then compared to the original medium-sized pattern from the source image. The best match from the candidates is then used in the output super-resolved image. An optional enhancement (which was used in the results presented in this paper) is the consideration of adjacent areas when matching (as opposed to considering only the presented medium-sized pattern). This is accomplished by overlapping the pattern with previously enhanced areas. In this case, a weight is applied to the match of the pattern to the candidate and the candidate to the overlap area. The candidate with the best score for the combined selection metric (SM) is then incorporated in the output image.

$$SM = (a \times D_M) + (b \times D_O) \quad \text{Eq. (2)}$$

An additional metric can be added to this selection process for larger data sets. Additional weight can be used to emphasize patterns that appeared more frequently in training. This can be incorporated to create a more robust difference evaluation value (DEV) of

$$DEV = (a \times D_M) + (b \times D_O) + (c \times O) \quad \text{Eq. (3)}$$

where a , b , and c are the weights for the difference between the compared medium patterns (D_M), difference between the overlap areas (D_O) and occurrence level (O) respectively.

If an occurrence value is to be considered in pattern selection, a significant number of images must be presented to generate a robust set of occurrence values. Alternately, the training process could be modified to group similar patterns together under an averaged pattern and increment its occurrence value. This potentially would have the added benefit of reducing database size while providing additional weight to patterns that are common (while not necessarily recurring identically each time).

D. Use of Commercial Database Product

The approach is designed to incorporate a commercially available database product. Microsoft's SQL Server 2010 was used during testing; however, any commercially available database product could potentially be substituted. Therefore, the approach, takes advantage (without development cost) of features in commercial database products which include searching, caching and optimization. The availability of the MERGE command, discussed previously, is one such example of previously created functionality which was used to speed the development and testing of the software.

The use of a common database format may also allow greater interoperability with other potential providers or users of patterns. Database tools also aided the combination of various training image configurations that were used in testing, as discussed below.

E. Configuration Learning

Three key factors influence the successfulness of the super-resolution process, under this approach. First, the size of the small, medium and high resolution patterns can be varied as desired. It is important to consider each pattern size both relative to the image feature size and to the other pattern sizes. The pattern versus feature size, as previously discussed, can have a significant impact on the occurrence (and detection) of false positive and negative inclusions. The relative size of the medium and large patterns determines the amount by which the resolution of presented images will be able to be increased. The enhancement factor (EF) is computed as $EF = L / M$.

Second, the pattern composition of the database that is used for enhancement can be altered. A database that is domain specific, as discussed previously, will generally perform better than a generic one. Finally, metrics related to matching (how many candidates to select) and closeness (what is considered a possible / acceptable match) can be changed. The adjustment of these factors should allow maximization of the two key super resolution performance evaluation metrics: speed and accuracy. A potential enhancement and topic for future research could involve using an automated neural network (ANN) with feedback (based on speed and accuracy metrics) and backpropagation to optimize these settings.

4. DISCUSSION OF SUPER RESOLUTION EVALUATION TECHNIQUES

It was desired to evaluate the effectiveness of the proposed super resolution method's application without regard to application.

This however, is quite problematic as the success metrics vary from application to application. In a military remote sensing scenario, avoiding false positives and negatives would in many cases be of primary importance. In this context, a false negative could result in an enemy position being missed (possibly rendering friendly forces exposed to an unexpected attack). Alternately, a false positive could result in (in the best case) a waste of munitions on a non-existent target and at worse the unnecessary and undesired loss of human lives.

On the other hand, research has been previously conducted regarding using super resolution to reduce bandwidth needs for streaming video [e.g., 3]. In this context, the presence of false negatives and positives is less important and dwarfed by the need to avoid video discontinuities.

A third scenario to consider would be when super resolution is being used to enhance an image for navigation route-planning purposes. In this instance, discontinuities are less important and even false positive and negative readings are not problematic unless they impact the route generation process. Obviously, both a false positive and false negative inclusion have the possibility to do this. However, the impact of this may be able to be mitigated by the incorporation of a confidence value for each part of a super-resolved image.

Given the above, it is clearly not appropriate to judge all super resolution techniques by a single application-agnostic standard. The conventionally used pixel-difference metric values are exactly this type of standard. As such, this approach fails to capture the multitude of possible application impacts and instead renders judgments that may be of little use.

5. EXPERIMENTAL METHODOLOGY

With the goal of evaluating the proposed super resolution technique in a way that is applicable to at least several (if not most) applications, several experimental scenarios were developed. An implementation of the super-resolution technique (as previously described) was created. Next, training was conducted. The training was common to all experimental scenarios. The scenarios, however, differ in the number and set membership of the trained images and in the evaluation technique used.

The first scenario involved using (and for the two and three image cases, combining – using the SQL MERGE command, as discussed previously) images from the same set as the image that would be presented. The images used in this study were collected during research related to facial recognition. As such, a set of images consists of images of a single individual's face taken at different angles. Thus, in this first scenario, the super-resolution engine was trained with one, two or three images from a given set and then another (not trained) image from the set was presented for super-resolution.

In the second scenario, one, two or four images from different sets were trained (one image per set was selected for the one and two image trainings and two from two sets were trained for the four image training). An image from an alternate (not-trained) set was presented for super-resolution. The subjects of these images, thus, had different skin completions and features as compared to the subject of the image that was presented for super-resolution.

A final control scenario involved training an image and then re-presenting the image for super-resolution. This scenario is significant as it allows evaluation of the performance of the engine (as opposed to the dataset, as the presentation of a trained image guarantees that all necessary patterns for super-resolution will be present in the pattern database).

The super-resolved result images from each of these scenarios (and each number of trained images) were evaluated in three ways. First, the conventional average pixel difference comparison approach was used. This involved calculating the difference in value for each pixel and averaging them. These numbers are presented in table 1 as a percent of the maximum possible difference. This allows the results to be compared across images of various color depth.

Second, in recognition of the fact that for many applications a close enhancement is a complete (not partial) success, a threshold-adjusted pixel difference comparison was computed. This computation treats any value within $\pm V$ (acceptable variance) as a perfect 0-difference match. Any value outside of this threshold is averaged as per the standard average pixel difference computation.

Finally, a visual appraisal of the resulting images was conducted. This evaluation technique was included to evaluate the proposed technique's utility for applications (such as entertainment) where human perception of how good the resulting images look is the primary success metric.

6. RESULT SET

# Trained		Pixel Comparison Value	
		Same Set	Across Sets
	1	5.926173%	6.199269%
	2	5.877611%	6.129308%
	3	5.876358%	N/A
	4	N/A	6.104011%

# Trained		Threshold Px Comparison Value	
		Same Set	Across Sets
	1	0.3939018%	0.5331753%
	2	0.377937%	0.4959146%
	3	0.3775162%	N/A
	4	N/A	0.473204%

Table 1 Evaluation of the super-resolved images

7. DISCUSSION OF RESULTS

A. Visual Evaluation

The images that are produced by the proposed approach are not as visually attractive as one would desire (see Figure 1). The proposed approach performs very well over patches of continuous pattern (for example, hair, skin texture and the background texture). It also performs well when there are features within this type of a texture (for example, wrinkles and fold lines on a face). It even performs well on boundaries between textures (for example, between the hair and face or between face and chin) when a shadow is not present. The technique requires improvement for the images having dramatic

edges (particularly those with shadows), in order to be used in an application where human viewing pleasure is a key metric. Specifically, the approach produces visible discontinuities in enhancing the eyes and the face-shadow-wall boundary.

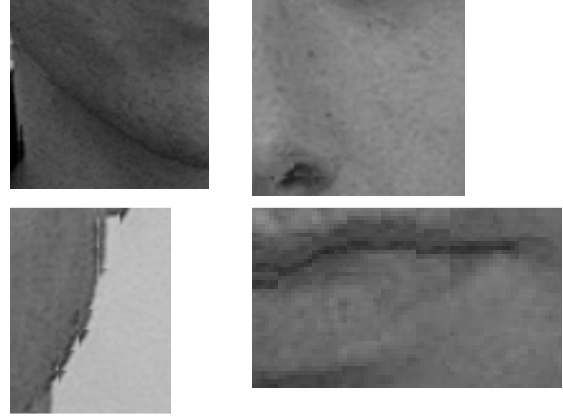


Figure 1a-d The effect of using super-resolution on four areas of an image.

B. Pixel Comparison

The pixel comparison evaluation technique directly mirrors the mechanism that is used to select the candidate patterns and choose (from amongst the candidates) the pattern that will be incorporated in the final super-resolved image. Given this, it is not surprising that the proposed approach performs very well with regards to this metric. Thus, by the conventional method of evaluating super-resolution performance, the proposed approach is performing acceptably.

C. Pixel Comparison with Threshold

Given that the pixel comparison with threshold approach is actually just a relaxed version of the pixel comparison approach, superior performance for the proposed approach under this metric would be expected (given the performance previously discussed for the pixel comparison approach). This is exactly what the test results have confirmed. For applications where achieving a proximal result represents a complete success, this metric is nearly perfect. As previously discussed, application-specific testing is necessary to ensure that the thresholds selected and results are appropriate to the particular use of the engine.

D. Alternate Evaluation Techniques

The pattern selection mechanism that is used has a significant bearing on the outcome of a super-resolution technique for any given metric. If the pattern selection mechanism and the evaluation technique are aligned (such as with the alignment that is demonstrated between the selection technique used in this proposed approach and the pixel comparison evaluation method), then it is obvious that the engine will perform well for the given metric. The identification of metrics, thus, is absolutely critical to any super-resolution application. If a computer-evaluable metric can be developed for a super-resolution goal, it is likely that the engine can be optimized and thus perform well for the application. Given this, it appears future studies of super-resolution would benefit from the

development of different metrics for success in addition to those presented in this paper and currently in use.

8. SOURCES OF ERROR

Virtually all super-resolution research uses experimental techniques which are necessarily somewhat contrived. Small and large source images are required, for many performance analysis techniques, requiring that an image be down-scaled, super-resolved and compared to the original high-resolution image. In this instance, images collected for previous facial recognition research were used to test the super-resolution image. These images incorporate a significant amount of foreground (skin) and background area (the painted wall) which has a similar texture. This may skew the results. Given that, for many applications, a significant amount of a quasi-similar pattern may exist, this is not an error per-se; however, it is one way that the results of this testing may not be applicable to other applications. Similarly, the choice of images used may make the results presented uncharacteristic of various applications.

9. CONCLUSION

The general suitability of the proposed approach and the engine developed to implement has been demonstrated in this paper.

As an inference-based methodology, the proposed approach is susceptible to producing both false positive and false negative inclusions. However, this can be mitigated through the use of appropriate pattern size values. Additionally, the training occurrence heuristic should (generally) push the engine towards avoiding these conditions as long as the training data is representative of (and inclusive of the feature types within) the image or images being presented for enhancement.

Unlike pure un-trained algorithm approaches, the proposed approach is also able to incorporate two benefits available only via training-based engines. First, it is able to provide a real-world-based occurrence heuristic with the benefits previously discussed. Second, the results can be customized and optimized for various image types via the customization of the training images presented. Output performance can also be improved via the adjustment of the pattern size settings and coefficients discussed previously.

10. TOPICS FOR FUTURE RESEARCH

Several areas of possible future research are suggested by the research that has been conducted. First, additional work is needed to develop metrics for evaluating super-resolution performance. This includes developing application-domain-specific metrics and other general-purpose metrics such as a surrogate metric for human image discontinuity perception.

Second, an artificial neural network (ANN) could potentially be used to optimize the engine's configuration parameters by incorporating metric feedback on engine performance across image enhancement runs. Research should be conducted into the feasibility of augmenting the proposed approach with a self-learning capability.

Third, a change to the training process may improve performance. Specifically, grouping together and storing highly similar patterns as a single pattern (as incrementing the

occurrence counter correspondingly) may result in a smaller database size (and thus faster super-resolution) as well as making the occurrence heuristic more useful. A closeness threshold parameter would need to be incorporated. This parameter's setting could also be optimized using an ANN, as previously discussed.

Finally, for super-resolved images to be useful as an input to other artificial intelligence (AI) systems (such as collision avoidance, route planning and such), it would be useful for them to include information regarding how confident the engine is about the enhancement of each particular area. Future research can focus on the development of a confidence value to convey this information.

11. ACKNOWLEDGMENT

Credit is hereby given to the Massachusetts Institute of Technology and to the Center for Biological and Computational Learning for providing the database of facial images.

12. REFERENCES

- [1] Bailey, Donald G. "Super-resolution of bar codes." *Journal of Electronic Imaging*, 2001: 213-220.
- [2] Bishop, Christopher M., and Andrew Blake. "Super-resolution Enhancement of Video." *Proc. Artificial Intelligence and Statistics*, 2003.
- [3] Borman, S., and R.L. Stevenson. "Super-resolution from image sequences-a review." *1998 Midwest Symposium on Circuits and Systems*, 1998: 374-378.
- [4] Elad, M., and A. Feuer. "Super-resolution reconstruction of image sequences." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999: 817-834.
- [5] Freeman, William T., Thouis R. Jones, and Egon C. Pasztor. "Example-Based Super-Resolution." *IEEE Computer Graphics and Applications*, 2002: 56-65.
- [6] Glasner, D., S. Bagon, and M. Irani. "Super-resolution from a single image." *2009 IEEE 12th International Conference on Computer Vision*, 2009: 349-356.
- [7] Huang, T.S. and Tsai, R.Y. "Multi-frame Image Restoration and Registration", *Advanced Computer Vision Image Processing 1*, p. 317-339.
- [8] Irani, M., and S. Peleg. "Super resolution from image sequences." *10th International Conference on Pattern Recognition*, 1990: 115-120 vol. 2.
- [9] Oktem, Rusen, and Levent Oktem. "A Superresolution Approach for Bar Code Reading." *Bilkent University Department of Electrical and Electronics Engineering*. <http://www.ee.bilkent.edu.tr/~signal/defevent/papers/cr1113.pdf>
- [10] Wang, Qiang, Xiaou Tang, and H. Shum. "Patch based blind image super resolution." *Tenth IEEE International Conference on Computer Vision*, 2005: 709-716.
- [11] Yang, Jianchao, John Wright, Thomas Huang, and Yi Ma. "Image super-resolution as sparse representation of raw image patches." *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008: 1-8.