The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Robotic applications of a defensible error-aware super-resolution technique

Jeremy Straub[1]

[1]University of North Dakota, Grand Forks, ND 58202, USA
(Tel: 1-701-777-4107, Fax: 1-701-777-3330)

[1]jeremy.straub@und.edu

Abstract: Many robotics applications that rely on computer vision for long-term route planning can benefit from increasing the resolution of the imagery used for that planning. Increased resolution increases the effective planning timeframe and allows the AI planner to consider obstacles that are more distant. As a best-case scenario, a route that might otherwise be taken just to encounter a distant obstacle that requires significant backtracking could be avoided. Super-resolution image enhancement, however, introduces its own problems as it can create false positive and false negative inclusions. Thus, for many applications, basic super resolution is unsuitable for robotic planning applications as the level of accuracy of the enhanced data is unknown. A framework for reporting confidence in super-resolution enhancement is presented in this paper. This approach includes a numeric confidence map along with the super-resolved data. The AI consumer of the enhanced data can, thus, consider both the data as well as the confidence meta-data. This framework is demonstrated and evaluated via an implementation of a database-based super-resolution technique that also supplies confidence map data. Robotic applications are discussed.

Keywords: Image enhancement, super resolution, decision confidence.

1 INTRODUCTION

Robotic control decision-making can have disastrous effects when it is based on inaccurate data. When a failure happens, it is important to be able to ascertain what caused that failure. Many artificial intelligence (AI) techniques, however, make this difficult. Some techniques train on-the-fly resulting in a decision being made with a framework that existed for a single moment-in-time and may never reoccur. While, theoretically, the data, node weightings, and other factors relevant to a decision could be preserved via storage, this is practically problematic – particularly when one considers that a craft that makes a poor decision may be destroyed (as a result of the decision, or otherwise), making the data unavailable for analysis. While this problem is wide-ranging and affects all areas of AI decision making, this paper deals with error that can be introduced via super resolution (the enlargement of images via computer processing) and how this error can be quantified and its potential factored in to the decision making process.

2 BACKGROUND

Numerous approaches to super resolution have been suggested. These approaches can be divided in to two main categories: those which infer data and those which simply piece together existing data. Piecing solutions [e.g., 2, 3, 4, 7] use several images of a scene or subject and take advantage of camera shifts or movement to allow a higher resolution image to be constructed from two or more low-resolution images.

Inference-based super resolution [e.g., 6], alternately, draws on patterns in the image format, prior knowledge or other sources to create an output image that attempts to replicate what would produced by capturing the original scene or subject at the output resolution. Inference-based approaches can be further divided into approaches which attempt to apply an algorithm to a single image to enhance it without additional information and approaches which base the output on both the input image and prior knowledge. All present forms of super resolution only produce approximations of the actual high-resolution image; these approximations can (and usually do, to various extents) include false positive inclusion and false negative exclusion.

Robotic AI control systems use imagery for immediate decision-making and long range planning. Immediate decision-making can include obstacle avoidance, target location and identification, and such. Long range planning uses image analysis for target identification, prediction of traverse-ability and, terrain type.

A short-range perception failure could result in damage to or loss of the robot (due to misjudging the terrain and falling or flipping). This, however, is unlikely as at close range the obstacles are quite large and would be difficult to miss (to the extent where super resolution would generally be unnecessary). Also, other techniques can be used to avoid these perils (such as a scanning laser or whisker, etc.).

At a longer range, however, super resolution is very relevant and the cost of bad information can be significant. Two scenarios deserve consideration. In the first longrage imagery scenario, enhanced imagery is used for route planning. In this instance, detail removal (via super resolution smoothing) can result in obstacles going unrecognized – conversely, it can also result in a preferable path being smoothed over to look no different than near-by obstacles). This may cause the robot to commit to a path which it will later be forced to backtrack from when higher-resolution imagery correctly detects the true nature of the path.

In the second scenario, super-resolved imagery is used for target location and identification. Particularly if the object size is close to the working-block size of the super resolution (for prior knowledge SR systems), it is possible that one complete image is incorrectly replaced with another. A friendly craft may thus be targeted as a foe or a foe incorrectly ignored as a friend.

3 SUPER RESOLUTION TECHNIQUE

The proposed approach makes three important additions to previous super resolution research. It starts with a trained database-based approach with corner matching loosely based on [5]. It adds to this the integration of a commercial database product for storage and searching, weighting of the search based on pattern occurrence training, and the inclusion of a confidence level map with the produced image.

3.1 Training

Training populates the database. Source images are trained to identify patterns starting at each pixel (excluding those on the far right and bottom where the training box size would be cut off by the image edge) in the image, starting from the top left. Figure 1 demonstrates how a given area of the image actually is used to crate numerous training patterns. The number of patterns produced by a given image is N=(H-A)(W-a), where a is the height and width of the pattern size.

Patterns collected in training are stored in the database at three resolutions. They are stored at the full resolution, a medium resolution (which will equate to the resolution of the images to be presented later) and a low resolution which is stored to aid in pattern matching when processing a presented image. The database contains a column for each pixel-location in each pattern size. An index is applied only to the database columns for the low-resolution image, to speed searching later. A temporary index may be incorporated for the high-resolution patterns to speed the training process; however, this should be removed before presenting images for SR.

For each pattern that is sliced from each training image, the database is searched for a match. If a match is found an 'occurrence' counter is incremented and no further work is done. If no match is found, the medium and low versions of the pattern are created via averaging the corresponding pixels of the high-resolution pattern and stored; the occurrence counter for the pattern is set to 1.

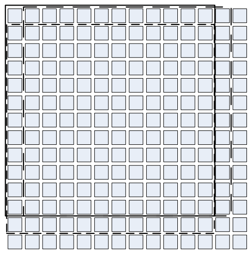

Fig. 1. Source Image Slicing

3.2 Image Enhancement

The image is processed through one medium-size pattern at a time, with a slight overlap. For each location, the medium resolution pattern from the source image is used to make a low-resolution pattern which is used to search the database and identify candidate patterns. Some previous work has attempted to fulfill a particular quota when selecting candidate patterns. In this instance, patterns are searched for based on a maximum difference from the low-resolution pattern (applied on a per-pixel basis). Any patterns which match the criteria are selected for consideration. If a match is not found, the tolerances are increased incrementally until a match is found or a maximum level of tolerance increase is reached. In the case where the search stops without finding a match, the medium resolution pattern is enlarged and placed in to the final image.

If candidates are identified, the medium resolution patterns (associated with the particular small size patterns selected in the search) are then used to select the pattern that will be placed in the final image. The medium patterns are compared to the medium pattern from the presented image and a difference value is computed. They are also compared to the overlap area and a difference value is computed. Finally, the 'occurrence' value stored with each pattern is retrieved. A final merged difference evaluation value is computed:

$$DEV = (a * DM) + (b * DO) - (c * O) \quad (1)$$

where a, b and c are the weighting constants (respectively) for the medium-pattern-difference (D_M), overlap-difference(D_O) and occurrence level (O). The prevailing pattern is placed in to the final image and the confidence map is updated.

3.3 Confidence

For each pattern of pixels placed in to the final image, a corresponding entry is placed in to the confidence map. The map, which is created as an ASCII text file, is depicted visually in figure 4. The confidence value is similar to the DEV used during the enhancement process; however, it is presented as a confidence value (as opposed to a weighted difference). The confidence value is calculated:

$$C=[(w*DM)+((1-w)*DO)]/K * (O/OMax) \tag{2}$$

where K is the maximum difference possible and O_{Max} is the maximum number of occurrences recorded for a single pattern in the current database (O_{Max} could also be set to an arbitrary value). It is expected that this confidence map can be fed into the AI planning routine and incorporated in its internal route confidence assessment process.

4 TECHNIQUE EVALUATION

A frequently used approach to evaluating super resolution is to compare the SR output image to the original image on a pixel-by-pixel basis. While this approach has the benefit of simplicity, it does not provide a particularly useful metric for evaluating the real-world performance of super resolution. In the case of images that will be viewed by humans, the question (generally) isn't one of whether the image is pixel-perfect, but instead relates to whether the image looks to be of high enough resolution. For example, in super resolving video for recreational viewing (such as in [3]), the objective is not to make the image perfect – but to make sure that the image is believably high-res. In other applications (such as the discussed application of AI decision making or barcode enhancement [e.g., 1]), the smoothness and high-resolution-look of the image are unimportant. What is important, however, is the accurate representation of critical features. To this end, a suitable performance evaluation metric is the correlation of low confidence with error areas. We can't expect the SR algorithm to perfectly enlarge every part of every image, every time but we must be able to ascertain how good each super-resolved image area is in order to include it in decision-making.

4.1 Subjective Visual Evaluation

The visual effect of the super resolution process is, obviously, quite subjective. Further, in an application in robotic guidance, the appearance of the super-resolved image to human observers is secondary to its utility for decision-making. However, as this is a commonly used metric for evaluating super resolution, it is appropriate to discuss it briefly. As figures 2 and 3 demonstrate, the algorithm seems (visually) to perform quite well on pattern-filled areas. Edges, however, can prove difficult. Experimentation using a database trained with the high-resolution version of an image that is presented performs exceptionally well. This, thus, demonstrates that the edge problem is not an issue with the approach (that is, the approach can produce a visually pleasing edge-area) but with the data that is stored during training. Increasing the training set size (using images that have minimal correlation with the image to be presented) also demonstrates improvements, confirming this.

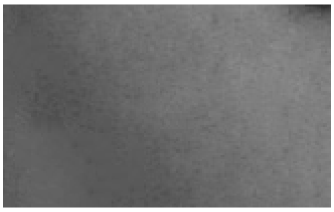
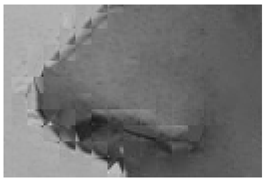

Fig. 2 & 3. Left: Enhanced Face Texture. Right: Pixilation Around

4.2 Pixel Comparison Evaluation

One of the most commonly used methods for evaluating super resolution is to compare the super-resolved image on a pixel-by-pixel basis to the original large-size image. This comparison averaged a 0.9% perfect match. However, given that the images are not black and white, but instead gray scale, it is not prudent to compare only perfect matches (as it is unlikely that a source pattern would have exactly the same shades of grey in pattern – the goal here is to get close). When we apply an error margin of 10% of the color range, the match average jumps to 89.3%. The visual difference of this 10% error margin is barely perceptible. It may, for example, manifest itself as a slightly different shading of an individual's face based on pattern matching of a lighter or darker skinned individual. However, similar effects could easily be caused by lighting in the location of photography or minor skin color change due to tanning, etc.

4.3 Confidence Correlation Evaluation

The evaluation technique that is of greatest interest to AI decision making, however, is confidence correlation. Confidence correlation compares each area of the image to the natural high-resolution image like in pixel comparison evaluation. The error value resulting from this is then compared with the confidence value for the pixel produced for the confidence map (as previously described). The confidence correlation, thus, indicates whether the confidence values should be trusted (while the confidence value indicates the accuracy of the super resolution process).

Confidence correlation is calculated as:

$$L = \frac{\left[\sum \left(\frac{C}{E} \times S \right) \right]}{n} \tag{3}$$

where C is the confidence value from Eq. 2, E is the error discussed in section IV-B, S is a scaling constant and n is the number of observations that are being used to compute the confidence correlation value. We, thus, end up with a scaled average correlation between the confidence and error value.

Confidence correlation was calculated for three cases: high error, medium error and low error. The confidence correlation values obtained through testing appear to be fairly consistent between the different error levels: 0.78, 0.82 and 0.79 for high, medium and low error, respectively.

5 SOURCES OF ERROR

Several sources of possible error were considered. Like most super resolution research, the experimental setup for this research is somewhat contrived. Images from prior facial recognition work were used to train and test the system. The images were headshots against a quasi-consistent background (though the actual color of the background in the images varied, likely due to the impact of lighting and using different cameras for different images). In each image, at least 25% of the image consisted of a lightly patterned background. The presence of this background could potentially increase or decrease accuracy calculations (depending on whether the database contained an exact or very close match) significantly, because of its prevalence. An additional test of two closely cropped face images was performed to assess the potential impact of this. One had a background color that was in the database and the other had a background color that the database was not trained with. While the accuracy metric values differed from the values for presenting the image with the larger background area, they did not fall outside of the range of values generated for testing with backgrounds.

6 CONCLUSION

It appears that a training-driven approach to super-resolution can produce acceptable results when the training and presented images are of similar objects. This would tend to suggest that a general-purpose database could be made that would be able to provide suitable enhancement for a variety of object and background types and which could be supplemented (to provide greater accuracy) with domain-specific databases.

When using a super-resolved image for AI decision-making, the potential error that is inevitably introduced by the super resolution process must be considered. The confidence map has been shown to provide a reasonable estimation of this error that can be considered by the AI as an estimation of the true error (based on comparing the actual high-resolution image to the super resolved one) which is, of course, not available in real world application.

Several topics of future research are suggested from this study. First, it would likely be possible to increase accuracy while reducing database size by storing difference patterns instead of gray scale values in the database. Future research could look at the feasibility of storing a pattern of delta-values from the mean-color-value of the image; [8] has looked at how patterns recur at different levels and thus it is reasonable to assume that they may recur in different shades as well. Second, this same concept could be applied to color images (including the aforementioned change). Finally, it would be prudent to compare the accuracy of decision-making using a higher-resolution super-resolved image with decision-making using a lower resolution image to determine the actual benefit of increasing image resolution.

ACKNOLEDGEMENT

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.

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] Irani, M., and S. Peleg. "Super resolution from image sequences." *10th International Conference on Pattern Recognition*, 1990: 115-120 vol. 2.
[8] Wang, Qiang, Xiaoou Tang, and H. Shum. "Patch based blind image super resolution." *Tenth IEEE International Conference on Computer Vision*, 2005: 709-716.

