Neil Patrick Del Gallego   04/16/15
Edward Patrick Tighe   CSC781M
Oliver Bob Urag

CSC781M Final Project Report
Superresolution Techniques and their Fundamental Limitations

Introduction

This paper tackles the problem of image resolution enhancement. Given a downsampled image, we attempt to upscale the image by a factor of 4 and observe if the enhanced image has achieved better overall composition and natural structures of the images are properly recovered. This paper covers two major techniques of superresolution, the multiple-image SR and the single-image SR. An experimental approach to single-image SR is also discussed on this paper.

The multiple-image SR is a technique wherein several low resolution images are combined to form a higher resolution image. The first step is to align these images such as computing the offset of pixels from one another, which is called the image registration. Once the image registration is complete, the fusion step attempts to recover the high resolution image by factoring the constraints and modelling the image formation process.

The single-image SR technique only attempts to consider as little as one image for attempting to create a higher resolution image. Techniques involve slowly resizing the image through iterations while imposing sharpening or deconvolution operators per iteration in order to bring back the natural structures of the image that is otherwise lost through distortion via interpolation. Another technique is to identify common pixel patches (5x5) that should also occur on its higher image resolution counterpart.

However, while numerous work has already been done, the results of the algorithms are mixed. There are certain parts of an image that is difficult to recover [1]. We will attempt to discuss these limitations alongside the techniques mentioned in detail.

Supperresolution Techniques

Related work on superresolution(SR) techniques are discussed in detail here. There are many techniques that have been proposed by numerous researchers that aims to solve the SR problem. We can group this into two major categories, the multiple-image SR and the single-image SR technique.

Multiple-Image SR

Multiple-Image SR technique involves having a set of low resolution images as input for the higher resolution image to be processed. An image resolution is mainly affected by the physical camera sensor. The goal of multiple-image SR is that increasing the sampling rate could potentially yield a better higher resolution image. These means that the camera should obtain more samples of a scene representing them as sequences of displaced images [2].

These image sequences are then combined and aligned using transformations that considers the geometric displacements of the images as well as its photometric components (such as lighting, changes in color, contrast, etc.) to yield a common reference frame. Thus, the first step is to properly "register" the images to create a common base frame. Once this process has been performed, a formulation of an SR estimator is considered that attempts to model the SR image by reversing the steps caused by image degradation caused by the image formation process. Important details that are present in the set of low resolution images are estimated and propagated to the higher resolution image [3]. We shall discuss the first step, **image registration**, followed by the **SR estimation** in detail.

- **Image Registration**

The first key component of a multi-image SR technique is the image registration process. Given a set of low resolution images, attempt to combine them to form a common reference frame using transformations. The overall accuracy of this process directly affects the outcome of a multi-image SR algorithm. This problem, according to [3], can be stated as follows: given N different images of a scene, for each image point in k view ($k \in N$), find the image point in $k + 1$ view which has the same properties of k (meaning it corresponds to the same point in the scene).

There are three kinds of registration process, the **geometric** registration type, **feature-based** registration type and the **photometric** registration type. The **geometric** registration considers displacements of images in its coordinates found and defined by the planar homography. The planar homography is dicussed in [3]. We shall briefly describe the geometric registration as follows: for each x in image k where $k \in N$ (N is the number of images of a scene), x' in $k+1$, x'' in $k+2$, and so on should correspond to point X found in the world when we attempt to map it into a plane.
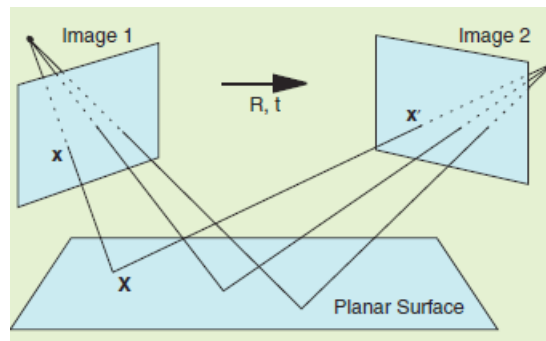

Figure 1: How images should correspond to a
certain point in a planar surface as described in [3]

Feature-based registration appears to be most accurate according to [3] as it has the ability to cope to widely disparate views and has excellent robustness to illumation changes. Furthermore, feature-based registration allows easier derivation of an SR estimator (using maximum likelihood or ML). In feature-based registration, interest points (or corners) are detected using an algorithm such as the Harris feature detector

Photometric registration refers to the procedure of considering photometric transformations between images. Such transformation include changes in illumination, intensity variations and changes in color composition. In this case, using a photometric transformation uses affine transformation under the constraints imposed by the contrast and brightness of the image, or the RGB channel [3]

- **SR Estimation**

Once a set of low-resolution images are properly aligned using the common registration techniques discussed in the previous set, the challenge would be to properly invert the image formation process to generate the SR image. The details present in the SR image should also be present across all the low-resolution images. This means that each low-resolution image corresponds directly to the SR image modeled.

We shall model the super resolution process mathematically as discussed in [4]. Given N low resolution images, $\{I_L^k\}_{k=1}^N$, of size L_1 X L_2. Find a high resolution image I_H of size H_1 x H_2 with

$H_1 > L_1$ and $H_2 > L_2$ which minimizes the cost function:

$$E(I_H) = \sum_{k=1}^{N} \left\| P_k(I_H) - I_L^{(k)} \right\| \tag{1}$$

where $P_k(I_H)$ is the projection of IH onto the coordinate system. IL represents a low resolution image. PK is modeled by four linear transformations that affects the high resolution image, namely the motion, camera blur, down sampling and additive noises. We can use a matrix-vector notation for the following:

$$I_L^k = D_k B_k W_k I_H + e_k \tag{2}$$

where DK is a down-sampling matrix, Bk is a blurring matrix, Wk is a warping matrix and ek is a noise vector. This simply means that a low-resolution image is affected by four factors caused by generating these images from the camera sensor. We should **inverse** this process to yield a HR image. The inversion of this image process is referred to as superresolution.

Real world scene   Motion   Blur   Down sampling   Noise

Figure 2: Inversion of the image formation process is called superresolution

For simplicity of this discussion, we shall only discuss one method for SR estimation. We refer to this as the ML or **maximum likelihood** estimation. Given such estimate, the resulting SR image when projected into the LR images via the image formation process, the difference between the actual and the "predicted" observations should be minimal.

This technique (although not specifically mentioned) is also similar in the technique discussed in [2]. This estimation technique is performed in an iterative manner. Given an initial SR image (which can be formed using bicubic interpolation or other interpolation methods), a set of low resolution images are determined. These set of LR images are compared to the original LR images entered. If these LR images are sufficiently equivalent (by measuring an error function), then we can say that the SR image is the "best" SR image to be returned as an output. Otherwise, we continue to refine the SR image until we minimize the error function.

For most multiple-image SR techniques, the error function is computed as follows [2]:

$$e^{(n)} = \sqrt{\sum_{k} \sum_{(x,y)} (g_k(x,y) - g_k^{(n)}(x,y))^2}. \tag{3}$$

The difference in images ($g_k - g_k(n)$) are computed and used to improve the "guess" of the SR image. Succeeding iterations should minimize the error function. Once a sufficient amount of the error function has been reached, or the error function has converged, then the iteration stops and outputs the final SR image. Variations for minimizing the error function exists on various works, but the goal is similar.
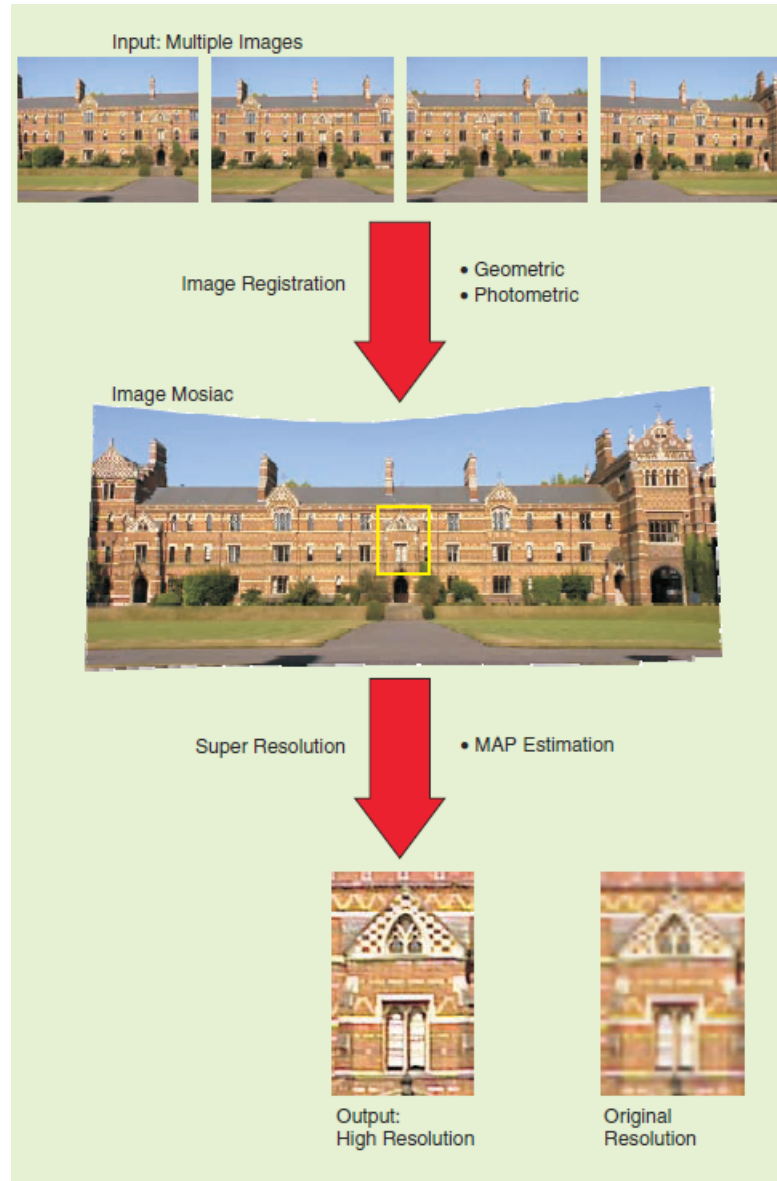
Figure 3: The general multi-image SR process as described by [3]

Figure 3 shows the general multi-image SR process as described in detail in the previous sections. An image mosaic is generated from a set of low input resolution images accurately aligned using one of the three common registration types (or a mixture of these), geometric registration, feature-based registration and photometric registration. The image mosaic generates an initial SR image which is further refined by considering its error function as discussed in the previous section.

Figure 4 shows the general error minimization process for multiple-image SR algorithms. Given an initial SR image (we call it the *SR_candidate)*, the set of potential LR images are generated (*LR_set)*. The *LR_set* and the original LR images (*LR_initial_set)* are compared using the error function formula mentioned in the previous section to get a function *e(n)*. If a certain threshold has been reached, *SR_candidate* becomes the final SR image. Otherwise, *SR_candidate* should be refined to minimize the error function *e(n)* until it reaches a certain threshold. A threshold of 0 theoretically means that the LR

images back-projected by the *SR_candidate* are exactly similar to the input LR images, but this is generally impossible to be performed.
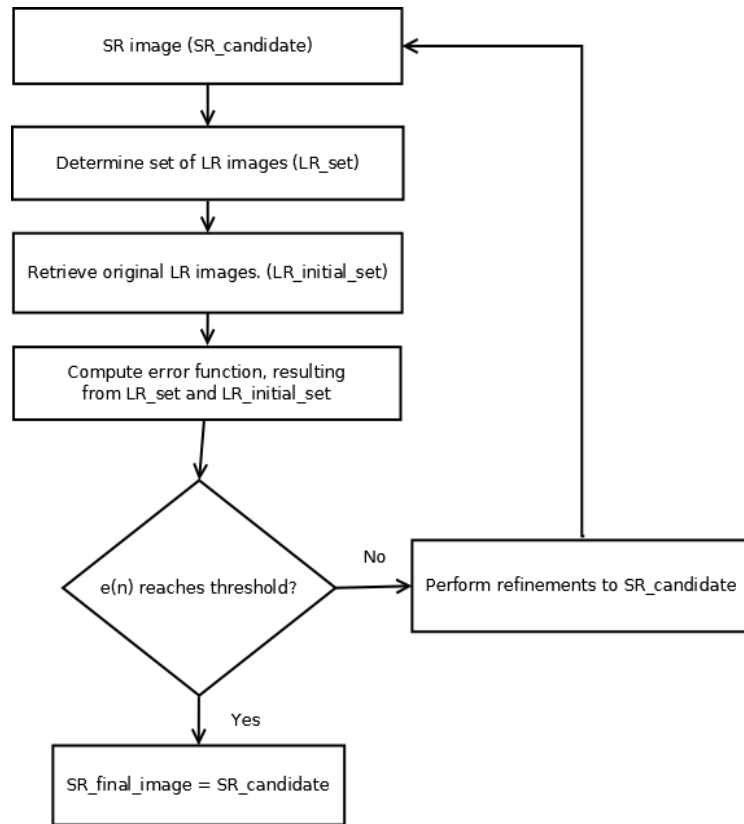

Figure 4: The general error function minimization process

Multiple-Image SR Examples and Limitations

To show proof of concept and idea to readers, this section contains sample images taken from various literature that uses a multi-image SR approach to generate a high-resolution image. We will also discuss our findings and observations regarding these images and how multiple-image SR approaches contain various limitations.

Oppo, a smartphone manufacturer, managed to release a commercial software product that implements a multiple-image SR technique to create 50MP photos from the phone's camera (Oppo Find 7 and Oppo Find 7a). The LR images entered for processing only takes four of the initial input images of 10. The four best shots of image sequences are only considered for the ML estimation. Figure 5 shows a sample image, given a 13MP initial LR image, and the output of 50MP image. Edges of the images are smoother on the output image which proves to be fulfilling. Noises present in the original 13MP are removed from the output image as well. The technique employed by Oppo (which is confidential as of this writing) also has certain limitations. It fails to emphasize the finer details in an image that seems to be affected by the overall luminance of the input images. Let us consider the image in Figure 6.

Figure 5: Oppo Super Zoom Feature. 13MP above. 50MP below.

Figure 6: Oppo Super Zoom Feature. 13MP above. 50MP below. (zoomed to a certain area)

Notice the lack of details when we attempt to zoom in on a particular area in Figure 6. Ideally, a high-resolution image should contain finer details but it was not really present in the output SR image. We also compare an output SR image to an SR image generated by a simple bicubic operation in. Notice that while the multiple-image SR technique by Oppo is fairly better than an interpolation technique, it also fails to recover the finer details present in the texts of the image.

Figure 7: Oppo Super Zoom feature. 50MP left vs 50MP right. Left photo uses the multiple-image SR technique. Right photo uses bicubic interpolation.

Consider the image in Figure 8 taken from [3]. Notice that as the magnification factor increases, details of the image are not really recovered. It has been explained in [1] that the performance of multiple-image SR techniques get worse as the magnification factor is increased. It shows that some of the reconstruction constraints imposed by the four transformations mentioned in the previous section become ireversible. According to [1], a set of low resolution images to reconstruct an SR image is no longer applicable beyond a point wherein the magnification factor is great (greater than 4 on most related work) as it is now affected by a wide range of factors which is discussed in their paper.

So how can an SR image be improved further? Clearly, the multiple-image SR technique poses a limitation (let us say a scaling factor of 4) for it to be actually effective. Other works regarding SR has shown that it is quite possible to recover an SR image given only one input LR image by attempting to "guess" the missing pixels or to repeatedly deconvolve a given image.
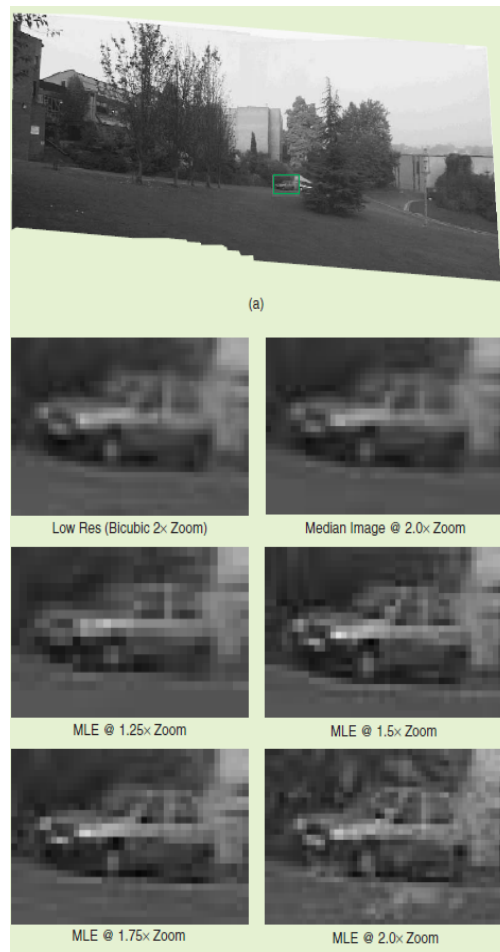
Figure 8: Given an image mosaic, an increasing magnification factor makes it more difficult to recover structures.

References

[1] Baker, Kanade (2000). Limits on Superresolution and How to Break Them. IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE Conference Proceedings. Carnegie Mellon University

[2] Irani and Peleg (1990). Super Resolution From Image Sequences. IEEE Transactions 1990. IEEE. The Hebrew University of Jerusalem

[3] Capel and Zisserman (2003). Computer Vision Applied to Super Resolution. IEEE Signal Processing Magazine. .

[4] Mitzel, Pock, Schoenemann, and Cremers (2008). Video Super Resolution usingDuality Based TV-L1 Optical Flow. IEEE. IEEE Conference Proceedings. Institute for Computer Graphics and Vision, University of Bonn, Germany