Super-Resolved Face Images using Robust Optical Flow

C. Fookes, F. Lin, V. Chandran and S. Sridharan
Image and Video Research Laboratory
Queensland University of Technology
GOP Box 2434 Brisbane, 4001 QED Australia
{c.fookes,fc.lin,v.chandran,s.sridharan}@qut.edu.au

Abstract

Surveillance systems are commonly used to monitor and track individuals in a cluttered environment. Face images that are captured using such a system often suffer from poor resolution and consequently degrade the performance of any face recognition system which may be applied to these images. Super-resolution (SR) is one avenue for overcoming this limitation, however, many existing SR techniques perform poorly in applications involving the human face as faces are non-planar, non-rigid, non-lambertian, and are subject to self occlusion. This paper presents a super-resolution system using robust optical flow in order to overcome these limitations. The optical flow method employed incorporates robust estimation methods to overcome problems associated with violation of the brightness constancy and spatial smoothness constraints. Resolving these issues greatly enhance the quality of the super-resolved images. Experimental results show significant improvement of the image quality and image resolution.

1. Introduction

Biometric research is a rapidly evolving field due to the increased demand on modern society to identify or authenticate an individual [9]. Most current practices, however, only link identification documents to the owner by human inspection. On the other hand, biometric identification has uncovered a wide range of human features and characteristics which may be used to identify or authenticate an individual with much stronger certainty.

One of the current streams of biometric research receiving a substantial amount of interest is the area of face recognition technology. These sophisticated recognition systems involve complex operations such as face detection, segmentation, normalisation, feature extraction and classification to ultimately verify or identify an individual. The performance of the majority of these recognition systems however, decreases significantly if the resolution of the face image drops below a certain level [5]. This drop in resolution decreases the amount of information available for identifying or verifying an individual, ultimately resulting in a severe degradation of recognition performance.

This issue is especially true in surveillance applications that involve the monitoring and the tracking of individuals in a cluttered environment. The majority of face images that are captured by such surveillance cameras have a very low resolution due to the cheap low-resolution (LR) imaging systems available. Also, a person's face usually occupies a small region of interest in the entire field of view of the camera. To perform face recognition in such an environment is an extremely arduous task. Thus, for face recognition to operate in a surveillance environment, a capacity must exist for the generation of higher resolution images of the face. This can be achieved through the use of super-resolution techniques: a signal processing method that can be used to increase the spatial resolution of a scene from a series of LR observations.

Super-resolution techniques have enjoyed good success in a wide variety of applications including medical imaging, satellite imagery, and some pattern recognition applications [8]. Many of these proposed techniques have been developed on the assumption that the system operates in a constrained environment, for example: only rigid objects assumed in the scene or only simple transformations are employed. Consequently, many of these proposed techniques are not applicable to images involving the human face due to the inherent difficulties that exist in this domain. Some of these problems include [1],

- *Non-Planarity:* A large majority of super-resolution systems employ simple parametric transformations in the registration stages, including translations, affine or projective transformations. These types of mappings, however, operate on the assumption that the environment is made up of planar objects. When such approaches are applied to images involving the human face, results are severely degraded as the human face is far from planar.

- *Non-Rigidity:* To further compound this problem of non-planarity, the human face is also inherently non-rigid. Local deformations occur frequently as facial expressions transition from state to state as well as parts such as the eyes, nose, mouth and jaw that move independently from other parts of the face. Most super-resolution approaches fail on non-rigid scenes. Some are capable of handling independently moving objects, however, each object is assumed rigid [11].

- *Occlusions:* Movement of the human face will result in many occlusions as some parts of the face will block the viewing angle to other parts of the face. These occlusions can be a large source of error for many super-resolution techniques.

- *Illumination and Reflectance Variation:* Faces are subject to specular reflections, particularly off certain parts, eg: the cheeks and forehead. These reflections introduce further difficulties into the registration process as well as introduce outliers into the fusion stage.

To overcome some of these problems, particularly the non-planarity and non-rigidity of the face, it is possible to use optical flow techniques to recover a dense flow field that describes a deformation or mapping for every pixel in the scene. By determining these local flows, it is possible to track the motion of a complicated non-planar and non-rigid object such as the human face. The remaining two problems of occlusions and illumination variation can be addressed through robust estimation methods.

This paper presents a super-resolution system using optical flow. A similar approach has been adopted by Baker et al. in [1], however, the optical flow method they adopted performed poorly at motion discontinuities. This resulted in the generation of artifacts around the edges of faces that degraded the final super-resolved image. To overcome this dilemma, this paper presents a super-resolution system that utilises a "graduated non-convexity" algorithm to recover the optical flow [2]. This algorithm is based on robust estimation techniques which addresses violations of the brightness constancy and spatial smoothness assumptions - two issues that severely affect previous optical flow techniques.

The outline of the paper is as follows. Section 2 presents some background to the field of super-resolution. The proposed super-resolution system using robust optical flow is outlined in Section 3. Section 4 provides experimental results on a face image sequence. Concluding remarks are discussed in Section 5.

2. Super-Resolution Background

The fidelity of an image is governed by its spatial resolution which represents the number of pixels per unit area in an image. Generally, images with higher fidelity or higher resolution are visually more gratifying to the onlooking observer when compared to lower resolution images. These higher resolution images not only provide more detailed information, which is important for further investigation in many applications, they are also easier to recognise. However, there is a direct relationship between the cost of an imaging sensor and its spatial resolution and there are often times when high resolution sensors are just too expensive for a given application. Even if the cost is not an issue, there are physical limitations which restrict the maximum spatial resolution of current sensor technology. Increasing the pixel density causes less light to fall upon each pixel resulting in a poorer signal-to-noise ratio (SNR). Conversely, increasing the size of the chip in order to squeeze in more pixels will also increase its capacitance and this in turn will reduce the charge transfer rate and make the chip slower [8].

An alternative approach for obtaining a high resolution (HR) image is by judiciously combining a number of low resolution (LR) images. This process, referred to as "super-resolution (SR) image reconstruction", is an area which has recently experienced a prolific expansion in research. The aim of SR image reconstruction is to estimate a HR image with finer spectral details from multiple LR observations degraded by blur, noise, and aliasing [6]. It is not sufficient to just resample one single observation of the scene as size does not equate to resolution. Increasing the resolution could be viewed as either increasing the signal-to-noise ratio while keeping the size fixed and/or approximating the image at a larger size with reasonable approximations for frequencies higher than those representable at the original size [4]. Such an approach to resolution enhancement using signal processing techniques would result in significant cost savings. Another advantage is that traditional LR imaging systems can be utilised to produce these HR images.

In SR image reconstruction, the LR images represent different observations or "snapshots" of the same scene. These LR images are subsampled (aliased) and contain sub-pixel shifts, containing complementary information which can be merged into a single image with higher resolution than the original observations. The existence of the sub-pixel shifts is the key to this process. If the images only contained integer shifts, then there is no new information which can be combined to form the HR image.

The sub-pixel shifts are generated from relative scene motions between different image captures. These captures may be made from a video sequence, multiple captures from the same camera at different positions, or multiple captures from many cameras at different locations. The difference in locations can be a result of controlled movement, such as an orbiting satellite, or uncontrolled motions such as a vibrating imaging system. Whatever the scenario, if the sub-pixel shifts are known or can be estimated, then it is possible to

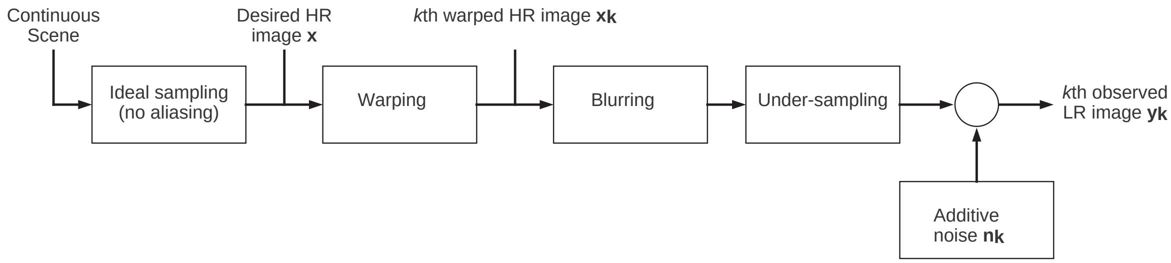

Figure 1. Super-resolution observation model.

generate a HR image from the multiple LR observations.

The field of SR image reconstruction is closely related to image restoration [7]. In the latter case, the task involves removing the effects of noise, blurring, and other artifacts from a degraded image in order to recover a "restored" image of the original scene. The resolution of the image remains the same, however. SR techniques not only have to deal with relative motion between images, they must also deal with blur and noise removal as well. Generally, most SR image reconstruction methods consist of three basic components,

1. Motion compensation (registration),
2. Interpolation,
3. Blur and noise removal (restoration).

There are numerous SR image reconstruction methods proposed in the literature to generate a HR image from a series of LR observations. Consequently, the way in which the registration, interpolation and restoration stages are performed vary according to the method employed. One of the first steps in the analysis in the SR reconstruction problem is the formulation of an observation model. That is, to develop a model that relates the HR image to the observed LR images. Several observation models have been proposed in the literature but generally, the type of model employed depends whether the SR reconstruction problem is being applied to multiple still images or a video sequence. For the former case, the observation model can be expressed as,

$$y_k = DB_kM_kx + n_k, \tag{1}$$

where y_k denotes the $k = 1 \dots p$ low resolution images, D is a subsampling matrix, B_k is the blur matrix, M_k is the warp matrix, x is the original HR image of the scene which we are trying to recover, and n_k is the additive noise that corrupts the image. This scenario is graphically illustrated in Figure 1 which shows how the LR observed image y_k is obtained from the original continuous scene.

As seen from Equation 1, the SR reconstruction problem essentially is an inversion problem as the process lies in the determination of the HR image, x, from multiple low resolution observations, y_k. This scenario is also an ill-posed inverse problem as a multiplicity of solutions exist for a given set of observation images [10]. This is due to an insufficient number of LR images and also ill-conditioned blur operators. Problems such as these are often managed by constraining the space of possible solutions according to a-priori knowledge of the form of the end solution. Typical constraints include image properties such as smoothness and positivity. The incorporation of such constraints allow for improved quality of the HR reconstructions.

3. Super-Resolution Optical Flow

As discussed earlier, previous approaches to super-resolution perform poorly when applied to applications involving the human face as faces are non-planar, non-rigid, non-lambertian, and are subject to self occlusion [1]. A super-resolution system that is based on optical flow, however, is capable of overcoming these problems due to the estimation of a dense flow field that describes a deformation or mapping for every pixel in the scene. The incorporation of optical flow overcomes the principal difficulty of estimating motions of a non-rigid object. The following sections will describe the outline of the proposed system and the details of the individual modules.

3.1. Proposed System Outline

The super-resolution system proposed in this paper takes an image sequence as input and outputs the super-resolution image sequence along with the optical flow between successive frames. This concept is illustrated in Figure 2 which shows the super-resolution system flow diagram.

The individual steps of the algorithm are described as follows and are repeated for all images in the input sequence,

1. Interpolate the original image to twice the input resolution using bilinear interpolation.

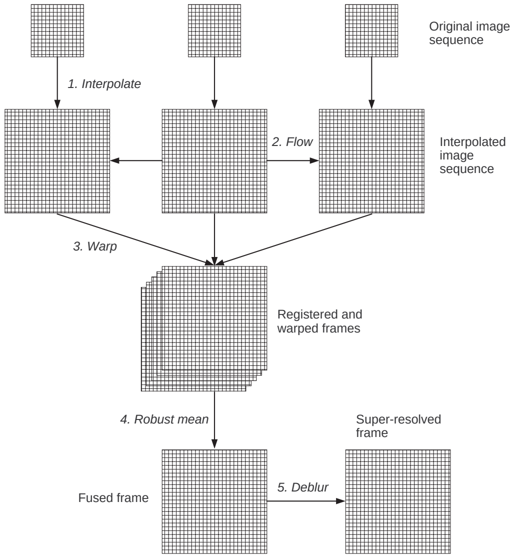

Figure 2. Super-resolution system flow diagram.

2. Compute the optical flow between the current reference image and the two previous images and the two following images.

3. Register the two previous and two following images to the reference image using the displacements estimated from the optical flow stage.

4. Estimate the super-resolution image using a robust mean computed across the reference image and the four registered images.

5. Restore the final super-resolved image by applying a deblurring Wiener deconvolution filter.

3.2. Optical Flow Algorithm

Optical flow techniques relate the 2D projection of the physical movement of points in a scene relative to an observer to 2D displacement of pixels on the image plane. These techniques operate on the concept that an object or point in an image will be observed with an intensity $I(t)$ at a certain time. Although the physical location of this point may change over time, it will always be observed with the same intensity. This is known as the intensity constancy assumption and can be expressed as,

$$I(x, y, t) = I(x + u\delta t, y + v\delta t, t + \delta t), \qquad (2)$$

where (u, v) are the horizontal and vertical velocities (or flow fields) of a point (x, y), and the change in time δt is assumed to be small.

Optical flow techniques also generally operate using a spatial coherence constraint. That is, surfaces have extent and consequently neighbouring pixels in an image are likely to belong to the same surface. This assumption equates to a smoothness constraint which is imposed on the algorithm to ensure that the motion of neighbouring pixels varies smoothly.

Most optical flow algorithms break down when these above two constraints are not satisfied in practice. This occurs regularly, especially when motion boundaries, shadows and specular reflections are present. To avoid these problems in the super-resolution system proposed in this paper, a robust optical flow algorithm was implemented. This algorithm was originally developed by Black et al. in [3] and [2] for robustly estimating optical flow between a pair of images. This method is based on a robust estimation framework and overcomes significant flaws inherent in many optical flow algorithms by addressing these two key assumptions: the brightness constancy and spatial smoothness assumptions. A graduated non-convexity algorithm is proposed to recover the optical flow and motion discontinuities. Readers are referred to [2, 3] for further details on the optical flow implementation.

Example results of the optical flow algorithm are given in Figure 3. This figure shows an image pair with the calculated optical flow fields, u and v, for the x and y directions respectively. The optical flow algorithm also detects data and spatial discontinuities, i.e. pixels that do not satisfy the two assumptions described above. These figures are also shown in Figure 3.

3.3. Robust Mean

Once the frames have been registered or warped onto the coordinate of the central frame using the results of the optical flow, the pixel intensity values are fused together by taking the robust mean. The robust mean algorithm used assumes a normal distribution of intensity values for an individual pixel, taking the mean of values within two standard deviations from the mean (95% of values). This prevents extreme values from corrupting the result.

3.4. Deblurring

Each frame is deblurred individually by convolving it with a spatially invariant Wiener filter. A 5 x 5 Gaussian kernel was used, with a standard deviation of 0.46. The value was chosen as it resulted in the smallest MSE and MAE values across the image sequence. Figure 4 shows the MSE with different values for the standard deviation.

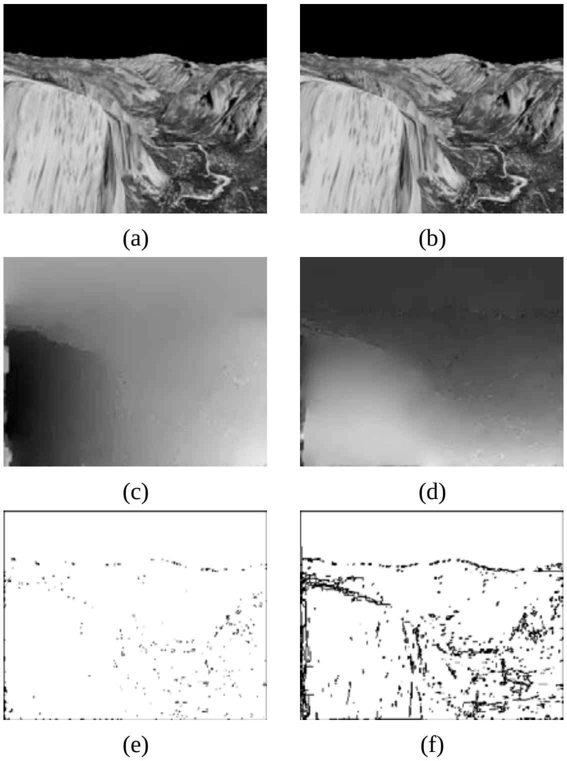

Figure 3. Optical Flow results on example synthetic data. (a,b) Yosemite image pair, (c,d) horizontal and vertical flow fields, (e,f) data and spatial discontinuities.

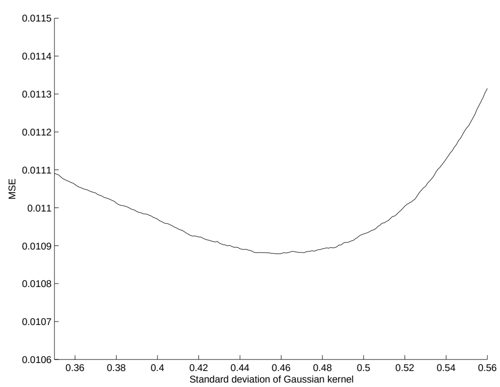

Figure 4. Plot of MSE vs sigma for deblurring.

4. Experimental Results

To test the performance of the proposed system, a 26 frame image sequence was downsampled from 100 x 100 (HR) to 50 x 50 (LR). 5 LR frames were used for computation of optical flow and reconstruction of each SR image.

A single frame of the original image sequence is shown in Figure 5 at HR and LR, along with a 50 x 50 cropped region of interest, i.e. the face region.

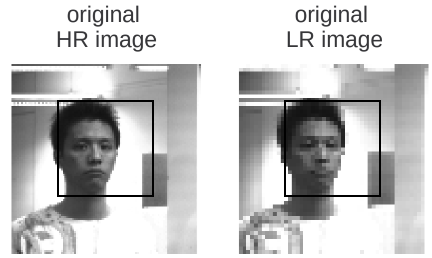

Figure 5. Original Image scene at HR and LR respectively.

Figure 6 shows the LR images, the enhanced images using bilinear and bicubic interpolation, the super-resolved images using the system proposed in this paper, and also the original HR images. The visual improvement of the SR image over simple bilinear interpolation can be easily seen. The facial features in the super-resolved images are sharper than those in the bilinearly interpolated images, especially around the nose and eye regions.

The Mean Squared Error (MSE) and Mean Absolute Error (MAE) were also used to evaluate the different image expansion/reconstruction methods quantitatively. MSE and MAE are defined respectively as,

$$MSE = \frac{\sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}(\hat{z}_{m,n} - z_{m,n})^2}{\sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}(z_{m,n})^2} \tag{3}$$

and

$$MAE = \frac{\sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}|\hat{z}_{m,n} - z_{m,n}|}{\sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}|z_{m,n}|} \tag{4}$$

where $\hat{z}_{m,n}$ is the reconstructed image and $z_{m,n}$ is the original HR image. From Table 1, the average MSE and MAE obtained from the SR image sequence show significant improvement over bilinear, bicubic, and even more over nearest neighbour interpolation.

Method	MSE	MAE
Nearest neighbour	0.0239	0.0735
Bilinear	0.0159	0.0763
Bicubic	0.0161	0.0756
SR reconstructed	0.0109	0.0660

Table 1. Comparison of expansion/reconstruction methods on image sequence

5. Conclusion

This paper has presented a super-resolution system using robust optical flow to recover high-resolution images of the human face from an image sequence. Such a system

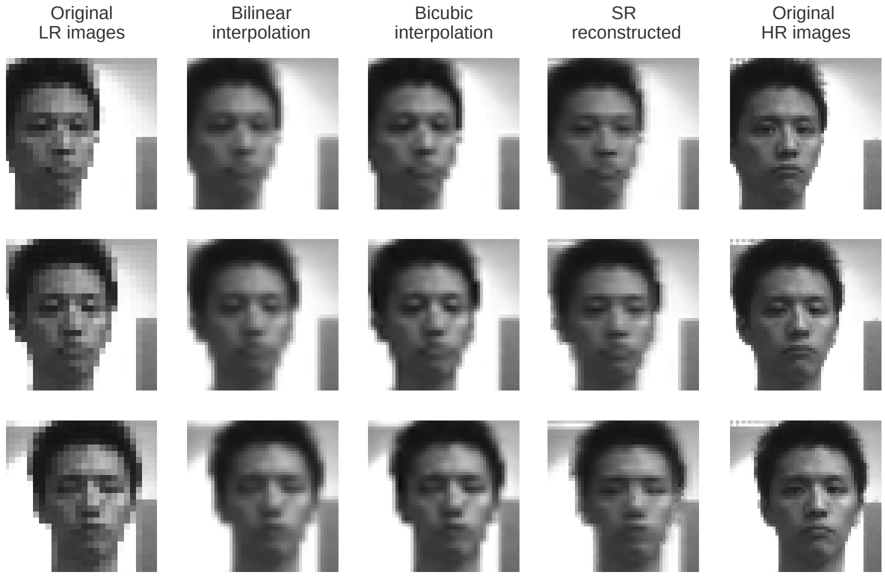

Figure 6. Super-Resolved Results.

is applicable to surveillance applications where the resolution of face images is often a very low quality. This would consequently degrade the performance of any face recognition system which may be applied on the extracted images. The proposed system utilised an optical flow technique to recover the local displacements between image pixels. This helped to overcome some of the problems associated with images involving the human face, as faces are non-planar, non-rigid, non-lambertian, and are subject to self occlusion. The performance of such a super-resolution system is extremely dependent on the effectiveness of the optical flow. Thus, the optical flow technique implemented in this paper was selected due to the incorporation robust estimation methods. These methods help overcome common problems associated with violation of the brightness constancy and spatial smoothness assumptions. Experimental results showed the super-resolution system improved the image content and quality significantly, especially when compared to nearest neighbour or bilinear interpolation methods.

6. Acknowledgements

This research was supported by the Office of Naval Research (ONR), USA, under Grant Award No: N000140310663 and by the Australian Research Council (ARC) through Discovery Grant Scheme, Project ID DP452676, 2004-6.

References

[1] S. Baker and T. Kanade. Super-resolution optical flow. Technical Report CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University, October 1999.

[2] M. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Fourth International Conference on Computer Vision*, pages 231–236, Berlin, Germany, May 1993.

[3] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[4] M.-C. Chiang and T. Boult. Efficient super-resolution via image warping. *Image and Vision Computing*, 18:761–771, 2000.

[5] B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes, and R. Mersereau. Eigen-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, 12(5):597–606, May 2003.

[6] M. Kang and S. Chaduhuri. Super-resolution image reconstruction: from the guest editors. *IEEE Signal Processing Magazine*, 20(3):19–20, May 2003.

[7] A. Katsaggelos. *Digital Image Restoration*, volume 23. Springer-Verlag, Heidelberg, Germany, 1991.

[8] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20(3):21–35, May 2003.

[9] N. Ratha, J. Connell, and R. Bolle. Biometrics break-ins and band-aids. *Pattern Recognition Letters*, 24:2105–2113, 2003.

[10] R. S. S. Borman. Super-resolution from image sequence - a review. In *Proceedings 1998 Midwest Symposium on Circuits and Systems*, pages 374–378, 1999.

[11] R. Schultz and R. Stevenson. Extraction of high resolution frames from video sequences. *IEEE Transactions on image processing*, 5(6):996–1011, June 1996.