7th International Symposium on Digital Earth
Perth, Australia
23 - 25 August 2011

IMAGE RECONSTRUCTION IN THE CONTOUR DOMAIN

GABRIEL SCARMANA

Faculty of Engineering and Surveying
University of Southern Queensland
gabriel.scarmana@usq.edu.au

ABSTRACT

Digital image enhancement systems are essentially pixel-based. In this paper, a novel method of multi-frame image enhancement employs contours rather than pixels as the primitive working unit. The feasibility of this proposal is based on the fact that gray-scale images can be sufficiently and accurately represented by their contour maps if a suitable contour model and scale selection method is used [5].

In this process several low-resolution shifted frames of the same scene are (a) registered (b) transformed into their contour domain, and (c) grouped or mapped globally within a singular framework to create an enhanced contour composite. Inverting this improved contour representation yields a high-fidelity reconstruction of an image in pixel format with more details than any of the original input frames.

The mapping of the low-resolution contour images within a singular framework takes advantage of the differences (or shifts) existing between their contour representations. These shifts are computed using an innovative image registration technique that can achieve accuracies of up to 0.01 of a pixel. The technique is based on DCT (Discrete Fourier Transforms) and normalized cross-correlation.

Controlled tests show that this global contour mapping approach, with subsequent converting to a compact and improved raster image, provides for an efficient and simplified method for multi-frame image enhancement. As compared to pixel-based multi-frame image enhancement processes, the proposed method may be computationally more efficient as it allows the detection, filtering and elimination of redundant contour data which may have little or no effect in the final enhancement.

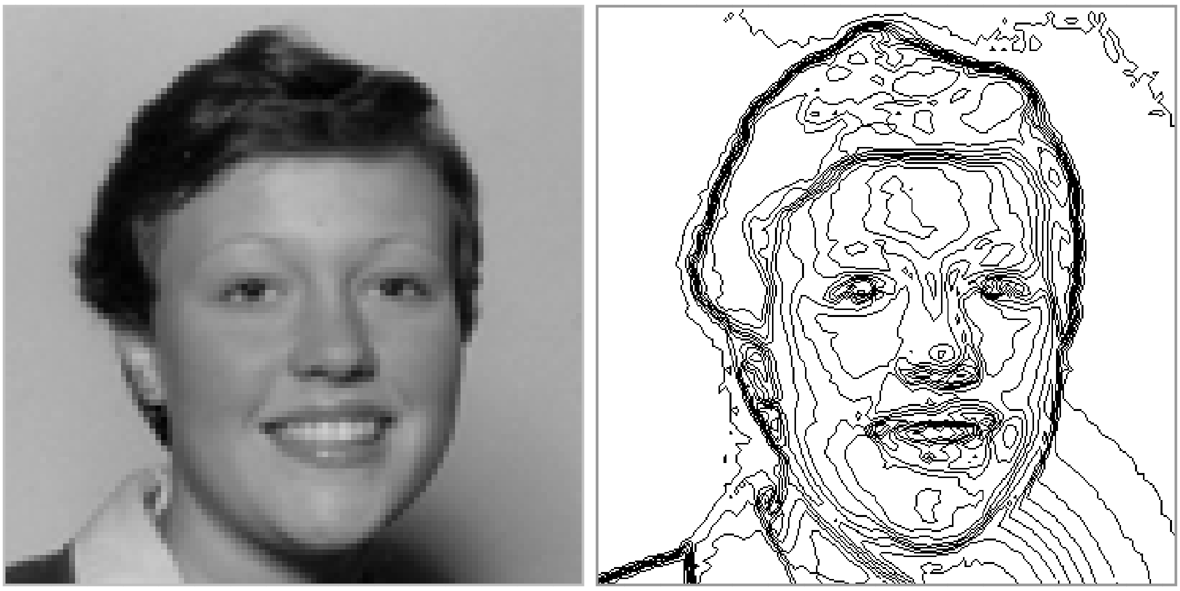
Although still in the experimental stage and restricted to gray-scale images, the proposed method is particularly viable in applications involving facial enhancements of compressed video frames.

1. INTRODUCTION

A digital image is a discrete approximation of an image obtained by sampling points with discrete coordinates and quantizing the values of each sample. It is generally formed by a finite number of sample elements equally spaced over a square grid with a rectangular shape [4].

Each element is called a pixel and has an intensity value. The rows and columns of elements determine the spatial coordinates of the pixel; and the intensity determines its -scale or gray-scale value. In a -scale image, all pixels have shades of ranging from black to white. In the case of colour images, the intensity determines the colour of each pixel according to some colour model, such as RGB (Red, Green, Blue).

Hence, a digital image can be represented as a three-dimensional surface or by using contours, as shown in Figure 1(a), where each point (i.e. a vertex) of a given contour corresponds to a position (x,y) having a constant gray-scale intensity or elevation z, respectively. The definition of contour used in this work relates to the term used in cartography and surveying where a contour line joins points of equal elevation (height) above a given reference level. Note that this definition may differ from the one commonly used in image processing where contours relate to shapes, edges or to lines and object boundaries [2].



(a) (b)

Figure 1 - An image (100^2) of the face (a), and a contour representation (b) using contour increments of 10 gray-scale intensity values on a scale 0-255.

In the general process of recording an image, there is a natural loss of spatial resolution caused by the non-zero physical dimension of the individual sensor elements, the non-zero aperture time, optical blurring, noise and motion [14]. However, this loss of spatial resolution may be largely reduced by combining a number of images of the same scene via pixel-based methods of multi-frame image enhancement [6], [11].

This is the particular case of video images where multiple images of the same scene taken in rapid succession can be extracted from the footage. Video sequences usually contain a large overlap between successive frames and regions in the scene are sampled in several images at different sub-pixel positions. This multiple sampling can often be combined to achieve images with a higher spatial resolution [2].

Differently from pixel-based multi-frame image enhancement methods, the approach here is to use contours as the primitive working unit. In this context, Figures 2(a) and 2(b) shows a contour representation of the face in Figure 1(a). Although the two images represent the same object they display a number of differences. For the same number of contour increments, contours lines in one image have disappeared while new contours have emerged in the other.

Hence, a composite image can be constructed by combining the contour representations of a sequence of images of the same object within a single reference framework. The result of this process will include more contour lines (hence information) than any of the initial input frames. Once a composite contour map is attained a reverse process is applied to transform this composite into a raster format (pixels).

Figure 2 – Two contour representations of the same scene sampled at different sub-pixel positions. For the same number of contour intervals, the differences between the two frames are evident.

In summary the proposed image enhancement method involves the following steps:

1. *Image registration*: Estimation of sub-pixel shifts among the different low-resolution input images depicting the same scene.

2. *Contour detection*: Each low-resolution frame of the sequence is converted to a contour format. The contour increments will depend, in general, on the images' dynamic range (largest and smallest possible values of pixel intensities within the input frames).

3. *Contour grouping*: The contour lines extracted from the input images are merged and/or mapped within a common reference frame using the shifts computed in step 1.

4. *Image reconstruction:* The combination of data is then filtered so that redundant contours are eliminated from the process of reconstructing an enhanced image in raster format for final display, analysis and use.

This succession of steps were automated and assembled with a package of basic image processing functions and tools included in Matlab 7.1 (www.mathworks.com, 2005).

2. CONTOUR DETECTION

The process of generating contours comprises the step of assigning coordinate values to pixels in the raster format image data and interpolating between pixels to find the coordinates of points in the path of a contour having a gray-scale intensity value.

Each pixel can be assigned x and y coordinates, which may conveniently be based on the coordinates of the centre of each pixel. Each pixel then has at least a pair of coordinates and a pixel gray-scale or intensity value. If two adjacent pixels have gray-scale values of 40 and 50 respectively, then a contour for gray-scale value 45 may lie between these two pixel centres. The coordinates in pixel space of a point on this contour can be estimated by interpolation, thus enabling the contours and/or the contour bounds to be estimated to sub-pixel accuracy.

For the class of scale facial images investigated in this work, the contour interval or increments were selected based on the dynamic range and/or on image histograms. Image histograms provide a convenient, easy-to-read representation of the concentration of pixels versus pixel brightness in an image. Using this graph it is possible to discern how much of the available dynamic range is used by an image [7].

The dynamic range is represented by how many levels in the scale are occupied. For instance, a group of pixels falling between values 25 an 210 (within a range of 0 to 255, see Figure 3), with none in the other regions, indicates a relatively wide dynamic range of brightness thus, in general, requiring a larger contour representation than an image with a narrow scale distributions (small dynamic range).

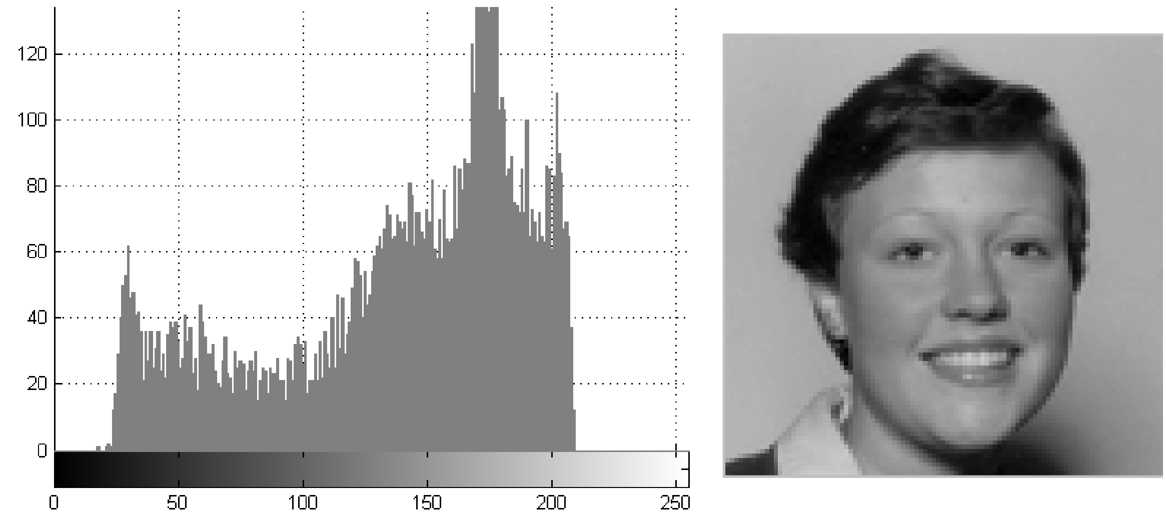


Figure 3 - The histogram on the left shows the distribution of pixel intensity values for the image displayed on the right.

Figure 4 illustrates a series of contour lines as produced from an array (4x4) of uniformly spaced pixel intensity values. In this instance the contours increment was 4 units. The dots connecting the contour segments are referred to as vertices. There exist a number of methods and computer packages used for constructing contour lines from equally distributed data points [1]. The contours in Figure 4 were derived by interpolation using a weighted average.

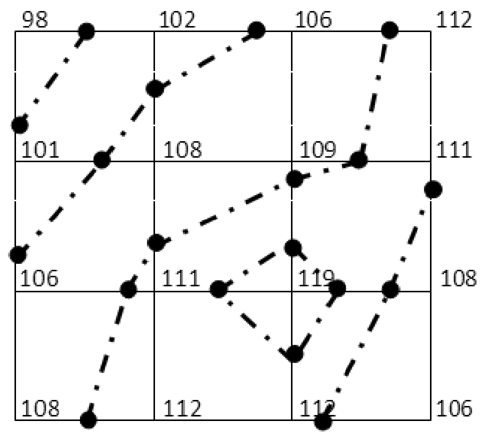


Figure 4 – Contour representation (with contour increments of 4 units) of an array of 16 uniformly spaced pixel brightness values.

A number of tests by the author have shown that contour increments of 5 scale units may be sufficient to reconstruct images of faces with relatively large dynamic range. This was confirmed by way of comparing the results of reconstructing a set of facial images from their contour representations to their original raster format. These tests were carried out using three major interpolants, that is, Weighted Average (WA), Bilinear and Bi-cubic [9].

The graph in Figure 5 is typical of the degree of accuracy to be expected when reproducing a raster image from its contour representation. The graph is based on the image of the face in Figure 1. The y axis indicates the r.m.s. (root mean square) error of the differences between the original image of the face and the same images reconstructed using the vertices of -scale contour increments ranging from 1 to 10.

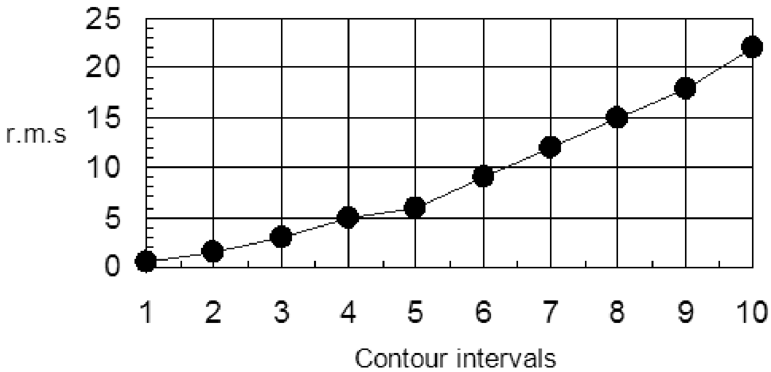


Figure 5 - The graph shows the r.m.s. of the differences between the original image of the face in Figure 1 and the same image regenerated using contour lines with increments ranging between 1 and 10.

The graph shows an exponential degradation of the r.m.s. value as the contour interval is increased. The reverse interpolation method used in this example was Weighted Average [1]. In all cases, the number of vertices required to reconstruct the original image of the face was less than the total number of pixels of the image itself (100^2). By way of comparison, the .txt file needed to store the vertices required to reproduce the face (for contour increments of 5 units) used approximately the same amount of memory (i.e. 10KB) of a JPEG compressed version of the original image (i.e. 10KB), while reproducing a more accurate image with more visual details.

3. SUB-PIXEL IMAGE REGISTRATION

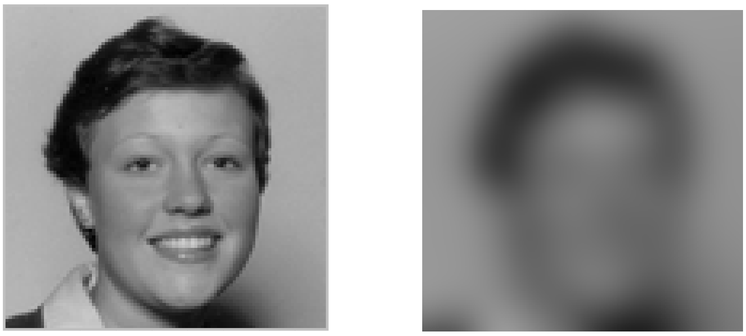
There exists a number of methods for determining sub-pixel translation parameters (i.e. in x and y) existing between two shifted images depicting the same scene. The method used here was devised in [8] and is based on DCT (Discrete Fourier Transforms) and normalized cross-correlation registration techniques [13].

The technique allows images to be registered or matched without using control points in the registration procedure. The accuracy of the registration is user selectable and can achieve accuracies of up to 0.01 of a pixel. For two given images *F* and *G* the registration outputs the normalized root-mean-squared error (n.r.m.s.e.) between *F* and *G*, their global phase angle and the row (x) and column (y) shifts between the two images respectively.

For a more accurate detection of the shifts or offsets between two images, the image must contain some features that make it possible to match two under-sampled images. Very sharp edges and small details are most affected by aliasing, so they are not reliable to be used to estimate these shifts. Uniform areas are ineffective, since they are translation invariant [11]. The best features are slow transitions between two areas of grey values.

These areas are generally unaffected by aliasing and such portions of an image need not to be detected specifically, although their presence is very important for an accurate result. In this context, before a given sequence of images of the same scene is registered, a low-pass filter may be applied uniformly to each image. The purpose of a low-pass filter, as shown in Figure 6, is to smooth (1) sharp edges and small details, (2) sudden changes of intensity values, and (3) the distortions created by a compression process.

The motion estimator (registration procedure) adopted in this research determines the x and y shifts between any two images, but what is really required is the relative positions of a sequence of images. By calculating the shifts with respect to a single reference image, only one realization of the relative positions is obtained. By repeating the procedure for another reference image, a second estimate for the relative positions is made.



(a) (b)

Figure 6 - The original unfiltered image of the face (a) and (b) after applying a low-pass filter (Gaussian, 18 pixels radius).

Reiterating this process for all images in the sequence, a better estimate of the relative shifts, image to image, can be found. The statistical measure used to determine the 'best' possible value for all possible combinations of the motion vectors between a set of shifted low-resolution images is the vector median. If the vector mean was taken instead of the median, then the final motion vector would be an entirely new vector, and not one of the vectors originally estimated. In addition, the mean is less robust than the median if outliers are present [10].

4. IMAGE RECONSTRUCTION - FROM CONTOURS TO PIXELS

Upon determining the shifts between the images, the vertices of the contours are projected or mapped on a uniformly spaced high-resolution grid (see Figure 7). Depending on the final resolution required, this grid may be selected so as to create more pixels in x and y than any of the original input images.

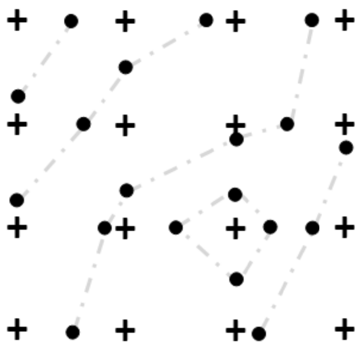


Figure 7 – Contour vertices (dots) are used as the randomly spaced contour locations needed to determine the pixel values which would exist at the intersections of a regular grid (crosses).

Only those contours that provide additional and/or relevant information are used in the process. This is important because the input images may contain multiple redundant contours (i.e. intersecting or nested contours), any of which might logically be used as a basis for discrimination or, on the other hand, used in the reconstruction of an improved image composite. In other words, contours may be filtered so as to eliminate those contours which may have a small or no contribution in the accuracy of the final reconstruction, thus saving substantial processing time.

A potential test for whether a contour is a candidate for removal is to compare the centroid values, the area these contours occupy and the shorter and longest distance to the centroids [13]. These parameters may offer the best statistical constraint for deciding whether a contour is very similar to another and thereby could be removed.

Also, within any group of at least three close nested contours it may possible to test for redundant contours by removing the middle contour and interpolate and/or diffuse between the other two. However, this process may have the disadvantage of smoothing and/or distorting particular high frequency (sudden changes of brightness values within an image) information necessary for a more accurate reconstruction of the enhanced composite.

As illustrated in Figure 8, string filters may also used to remove vertices from contours. Using a user defined tolerance; string vertices that are within an offset tolerance of straight lines are removed. That is, a string filter removes vertices from contour lines that do not deviate by more than a specified offset tolerance from straight lines joining successive string vertices. Optional filtering techniques may be considered, alone or in combination. For instance, adjacent string vertices may be removed if they are equal to a given tolerance either in plan position only (i.e. have similar x and y co-ordinates), or equal to a given tolerance for x, y and z co-ordinates (where z is the pixel intensity value).

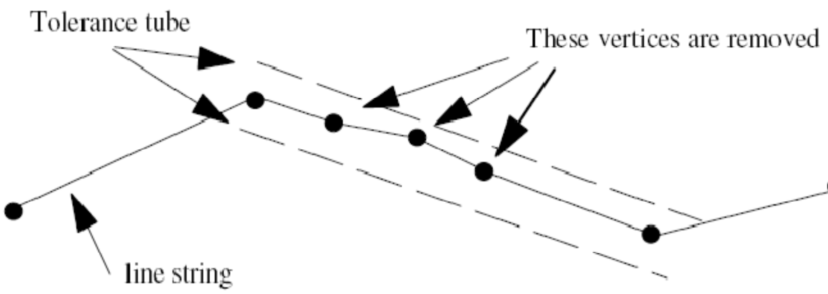


Figure 8 - A string filter option is used to remove surplus vertices from contours lines.

To determine the pixel brightness which would exist at the intersections of a regular grid using randomly spaced contour vertices locations, several interpolators may be used depending on the application and accuracy requirements. One of the methods for interpolating scattered data to a uniform refined grid is referred to as Kriging. This geo-statistical method was used because it is a statistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown areas.

A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated. The interpolation process determines the value of the points at grid intersections to have decimal figures. Hence, the values are rounded to the nearest integer which is the norm for representing pixels brightness within a raster image.

Kriging allows the user to derive weights that result in optimal and unbiased estimates. It attempts to minimize the error variance and set the mean of the prediction errors to zero so that there is no over- or under-estimates. An important feature of Kriging, as compared with other image or surface interpolators, is that it gives an estimation of the error at each interpolated point, thus providing a measure of confidence in the modelled surface. A thorough theoretical explanation of Kriging interpolation is beyond the scope of this paper and the reader is referred to [3] for the theory and applications of this interpolation technique in the particular areas of digital imaging and remotely sensed data.

5. TESTS AND RESULTS

The proposed enhancement process was tested for video imaging applications using simulated data. In this controlled experiment, the ‘true’ image was known prior to the enhancement and thus the accuracy of the enhancement and some of the factors that would degrade the process could be investigated and quantified. In general, the aim was to evaluate visually and statistically the performance of the proposed enhancement.

A high-resolution image referred to as the face was scanned and stored as a .tif file with no compression applied to it. This image is shown in Figure 9(a). The initial image was captured as a colour image and was modified to contain only 255 levels of grey-scale using the image processing toolbox of Matlab 7.1. The size of the image was 100^2.

Fifty (50) low-resolution .tif images were acquired by randomly repositioning and scanning the original image, avoiding unnecessary rotations. The scanner resolution was set to create coarse images having 4 times less pixels linearly than the original, thus creating ten images having a size equal to 25^2. The 50 images were then MPEG (Moving Pictures Expert Group) compressed for a compression ratio of approximately 10.

The main reason a conventional off-the-shelf scanner, rather than a standard digital camera, was used for these tests was the flexibility the scanner offered with respect to the selection of the format of image storage and the scanner’s optical resolution. Uncompressed images could be retrieved at various scanner resolutions and be stored in any desired image format such as TIF.

An image thus obtained was considered to be an image which would preserve the basic integrity of its colour information. The digital cameras to which the author had access compressed their images and may have introduced artefacts during that compression process. The use of the scanner allowed any set of images enhanced by the algorithm to be directly compared if required with an ‘equivalent’ uncompressed TIF image from the scanner.

Rotations were not applied to the images as they would have added extra parameters to the enhancement process and may have detracted from the strength of the conclusions reached in the experiments. Correlation obviously exists amongst a video camera and orientation parameters such as tilts, rotations and affinity/obliquity of the sensor, and, in a controlled experiment where the aim is to demonstrate the use of a process to enhance image resolution *per se*, it was thought imprudent to introduce such complications.

An example of the low-resolution image and its associated contour version are shown in Figure 9(b) and 9(c) respectively. The contour increment for all the down-sampled images was selected to be 5 gray-scale intensity values ranging between 25-225, thus producing approximately 40 contours per low-resolution image. This interval would include all the scale variations of the input images. Figure 9(d) shows visually the results of merging the contour representations of the 50 low-resolution images.

The enhancement process described in section 4 was applied to the 50 compressed images using a spatial enhancement factor (ratio between low-resolution and high resolution grid sizes) of 4. The r.m.s. of the differences between the enhanced composite and the original .tif image of the *face* was ± 5.4 with maximum and minimum values ranging between +12 and -9 grey-scale levels respectively.

In addition, the number of contour vertices processed to attain the result in Figure 9(d) was approximately 45% less than the total number of pixels included in the initial 50 input images, and only 35 of the initial 50 contour representations were used as more than this number did not contribute in improving the accuracy of the enhancement.

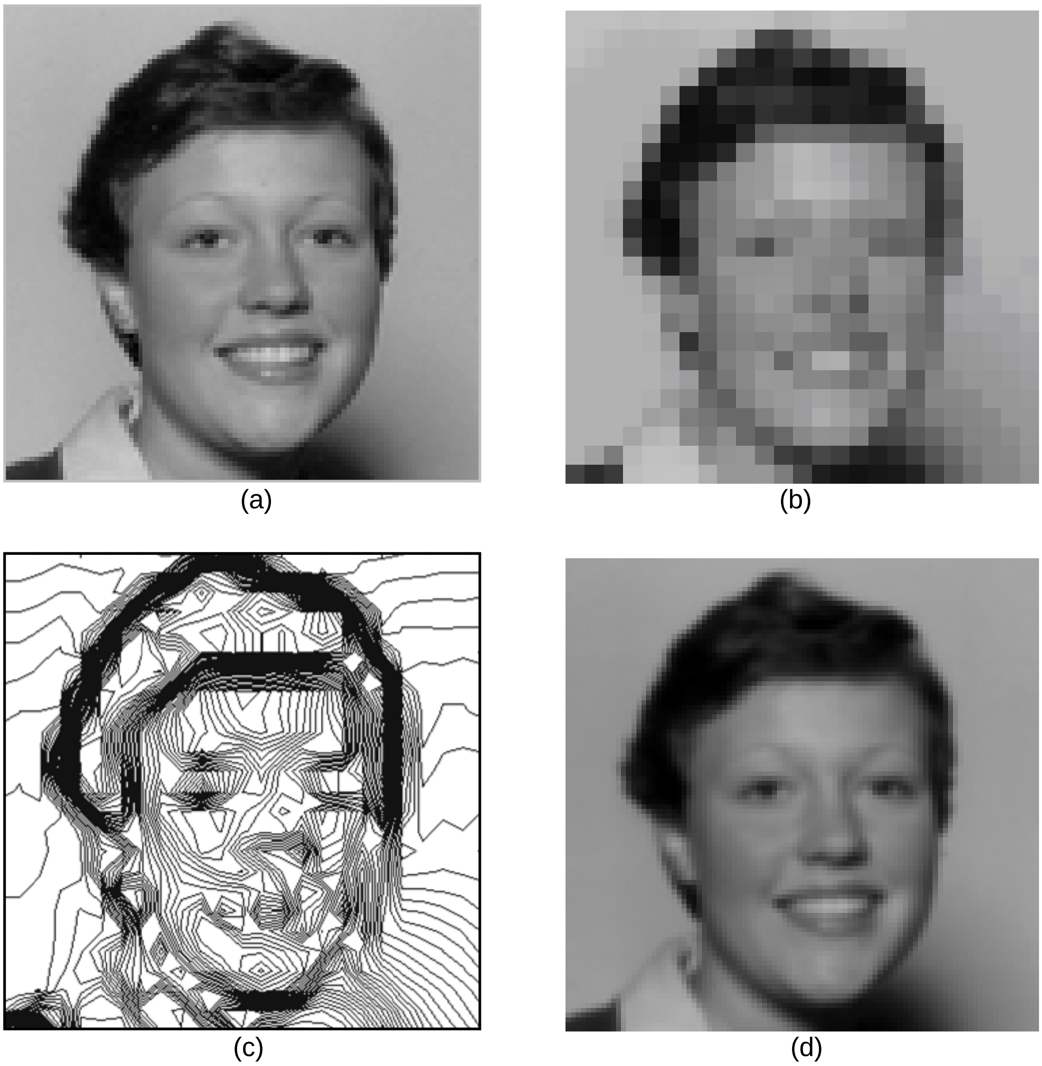


Figure 9 – The original high resolution image (100^2) of the face (a) and (b) one of the 50 down-sampled and compressed images (25^2) used in the enhancement. (c) is a contour representation of (b) whereas (d) is the enhanced composite derived from merging the contour representations of 50 low resolution images as in (b).

6. DISCUSSION AND CONCLUSIONS

Notable findings encountered in this study are:

- The conversion of a set of low-resolution digital images of faces into contour maps, and combining these map representations, may be used to produce a composite raster image having more details than any of the individual low-resolution images.
- The process of generating contour data comprises the step of assigning coordinate values to pixels in the raster format image data and interpolating between pixels to find the coordinates of points in the path of a contour having a given brightness value.
- A method for enhancing the resolution of a compressed sequence using a Kriging interpolation approach has been presented where the alignment or registration of the low-resolution images required for the enhancement relies on a novel registration method that can achieve fractional accuracies of +/-0.01 pixels.
- The application of the enhancement process has been demonstrated in 2D using controlled experiments which simulated a video sequence.
- The contour approach to multi-frame image enhancement allows for redundant contour data to be filtered and discriminated from the final enhancement process thus improving processing time while producing accurate results.

- Refinements to the proposed method are being undertaken to increase the accuracy achievable for a variety of scenes and dynamic ranges (including bi-tonal imagery).
- More research is required to assess the accuracy of the enhancement process in the presence of (a) rotations amongst the input images and the presence of random noise.
- Tests related to investigating the performance of a contour approach to image enhancement whereby both sensor and object are dynamic, and the illumination is non-uniform, are presently being undertaken.

REFERENCES

[1] Bonham-Carter G. F.: Geographic information systems for geoscientists: modelling with GIS. Vol. 13 of Computer Methods in the Geosciences. Elsevier, Butterworth-Heinemann. 398 pages. 2006.

[2] Bovik A.: Handbook of image and video processing. Elsevier Academic Press. 1372 pages. 2005

[3] De Jong S. M. and van der Meer F.D.: Remote Sensing Image Analysis Including the Spatial Domain, Kluwer Academic Publishers, pages 360. 2004.

[4] Duperet, A.: Image Improvements, Chapter 2, Digital Photogrammetry." Michel Kasser and Yves Egels editors, Taylor & Francis Publishers. p.p. 351: 79-100. 2002.

[5] Elder J. H. and Goldberg R. M.: Image Editing in the Contour Domain. IEEE Transactions on Pattern Analysis and Machine. Vol. 23 No. 3. 2001.

[6] Fryer J. and McIntosh, K.L.: Enhancement of Image Resolution in Digital Photogrammetry". Photogrammetric Engineering & Remote Sensing", Vol.67, No.6, pp.741-749. 2001.

[7] Gonzalez R. C. and Woods R. E.: Digital Image Processing. Prentice Hall, 954 pages. 2008.

[8] Sicairos G M, Thurman S. T. and Fienup R. J.: Efficient sub-pixel image registration algorithms" Opt. Lett. 33, 156-158. 2009.

[9] Smith M. Goodchild F. and Longley P.A.: Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. Third Edition. The Winchelsea Press. 560 pages. 2009.

[10] Spiegel M. R., Stephens L.: Theory and Problems of Statistics, Schaum's Outline Series, McGraw-Hill Book Company. 556 pages. 1999.

[11] Vandewalle P., Süsstrunk S. And Vetterli M.: A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution". EURASIP Journal on Applied Signal Processing (special issue on Super-resolution), Vol. 2006, Article ID 71459, pp. 14. 2006.

[12] Wang S., Ge F. and Liu T.: Evaluating edge detection through boundary detection. EURASIP Journal on Applied Signal Processing, vol. 2006, pp. 1–15. 2006.

[13] Weeks, A. R.:Fundamentals of Electronic Image Processing. SPIE Optical Engineering Press, Bellingham, Washington. .570 pages. 1996.

[14] Zitova B. and Flusser J.: Image registration methods: a survey. Image and Vision Computing. Elsevier Publisher, 21, June, pp. 977-1000. 2003.