VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

RECTIFICATION OF FIGURES AND PHOTOS IN CAMERA-CAPTURED IMAGES OF PRINTED DOCUMENTS

A. Keerthi Ravi and Annie Julie Joseph
Department of Computer Engineering, Model Engineering College, Ernakulam, Kerala, India
E-Mail: keerthiravimec@gmail.com

ABSTRACT

A method for the rectification of figures and photos in printed documents using a single camera-captured image. The algorithm requires a bounding box for the objects in a single-view image. On receiving the bounding box, the desired image can be cropped for further processing. The main feature of segmentation method is that, it exploits the properties of printed figures to detect the boundaries using an optimization scheme. The boundary is extracted to obtain the boundary points and rectifying the distorted image is carried out using a dewarping method. The method improves the quality of output by largely removing perspective distortions. The rectified image having low resolution is converted to a higher resolution image using an edge directed interpolation method. Thus, an enhanced rectified output image is obtained as a result.

Keywords: image processing; segmentation; feature extraction; rectification; super-resolution.

INTRODUCTION

The topic is motivated by the advantage of using digital camera for capturing photos of printed documents like books, posters, manuscripts etc. Compared with a scanner, a camera has much more portability. Using a camera, one can take the image instantly in a more casual manner. With a scanner, one can make the documents flat simply by pressing them intentionally on the glass board [1].However, the conveniences of using digital camera are also accompanied with some image quality problems. With a camera, the natural warping existing in the document surface would not disappear and the image will be deformed. In such cases, the process of document analysis and recognition will become complicated and unreliable.

However, conventional methods have drawbacks in rectifying figures. For example, use of multiple input images, requires depth-measuring hardware and high computing power are needed, strict conditions on illumination (like a single light source) are imposed, or the boundaries of documents are assumed. Although the imperfection is usually not a severe problem for OCR (Optical Character Recognition), but it becomes critical when users try to digitize figures and photos in documents. It is because even small amount of skew and/or curved boundaries are very annoying [2].

BACKGROUND

Conventional methods focus on dewarping of camera-captured images of complete pages. Therefore in order to get the digitized copies of printed pictures, segmentation and rectification are needed. As a first step, dewarping is accomplished with the help of a transformation model which maps the projection of a curved surface to a 2D rectangular area. The projection of the curved surface is delimited by the two curved lines which fit the top and bottom text lines along with the two straight lines which fit to the left and right boundaries. As a second step, fine dewarping is achieved based on words detection. All words are normalized by the lower and upper word baselines [1].

Text-line based approaches are easy, but picture rectification is not possible. It is because text-line based approaches need a number of text-lines in order to estimate document surfaces and this assumption does not hold form any interesting cases such as movie posters and pictures without sufficient text-lines. A number of text-lines are needed for document surface estimation and it does not hold for pictures [3].

A Coons surface patch based rectification method is defined by four smooth curve segments which form a curvilinear rectangle. The Coons patch is obtained by interpolation of these four input curves. The Boolean sum of convex combinations of two opposite curve segments and a correction term containing the four corner points are generated. The resulting surface interpolates the two given curve segments on the opposite boundary sides and one boundary curve of each given patch on the other pair. The bilinear Coons patch is obtained when the four input curve sand the blending functions are linear, and all the curves are cubic splines.

The desired figure after extraction and rectification will lead to lose of quality. Hence, the image obtained after rectification is enhanced to a higher resolution image. Edge-directed methods try to detect the edge directions and interpolate along the detected edge directions.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

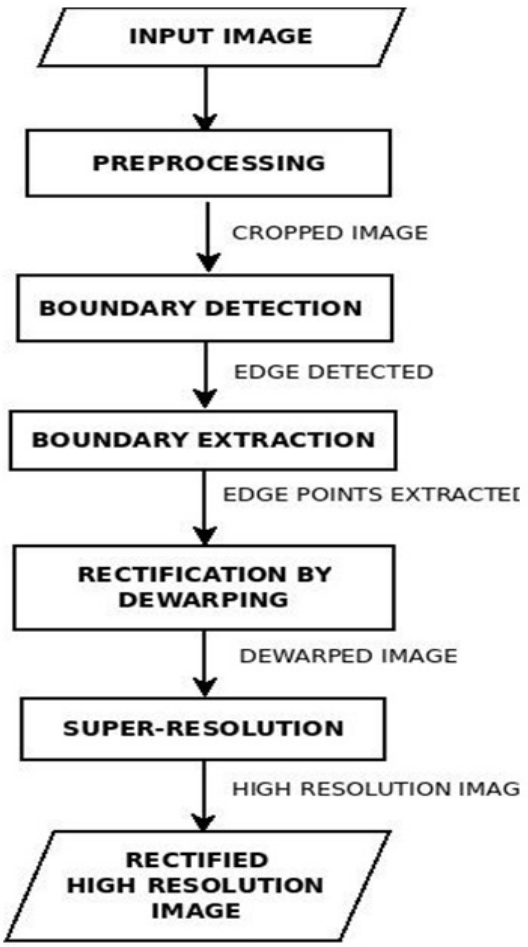


Figure-1. Overall block diagram of the methodology.

THE METHODOLOGY

The method suggests drawing a bounding box on a target figure by dragging the mouse. Then the algorithm segments the figure to obtain a binary image which detects fore ground and background of the cropped image using an energy minimization function. Progressive Probabilistic Hough Transform is used for boundary extraction. PPHT minimizes the amount of computation needed to detect lines by exploiting the difference in the fraction of votes needed to reliably detect lines with different numbers of supporting points. After boundary extraction, rectify the extracted figure using Coons patch bilinear interpolation. The Coons patc his obtained by interpolation of the four input curves. The resulting surface interpolates the two given curve segments on the opposite boundary sides and one boundary curve of each given patch on the other pair. The rectified image which is of low resolution is converted to high resolution image. A new edge-directed image interpolation method is used to obtain a high quality image.

Figure-1 shows the overall flow of the method of rectification of distorted images.

The method involves the following steps:

Preprocessing

Pre-processing is applied on images at the lowest level of abstraction and its aim is to reduce undesired distortions and enhance the image data which is useful and important for further processing [2]. The desired picture or photo is selected by creating a bounding box and it is cropped. The cropped image is the input image for further processing. Figure-2 shows the input document image. Figure-3(a) shows the bounding box created on the given image and 3(b) shows the cropped image.

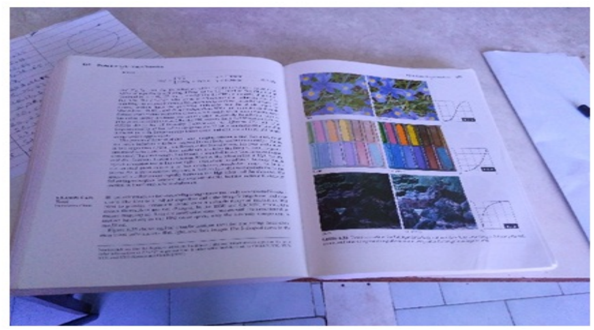


Figure-2. Input image of a printed document image.

(a) (b)

Figure-3. Preprocessing (a) Bounding box creation (b) Cropped image.

Segmentation

Segmentation partitions an image into distinct regions containing each pixels with similar attributes. To be meaningful and useful for image analysis and interpretation, the regions strongly relates to depicted objects or features of interest [4]. Segmentation is the first step from low-level image processing transforming a grey-scale or color image into one or more other images to high-level image description in terms of features, objects, and scenes. The motivation for using the term ‘Energy’ is that typical object detection/segmentation tasks are posed as an Energy minimization problem.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

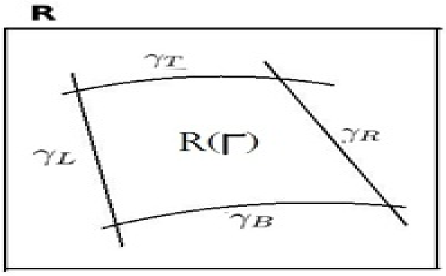


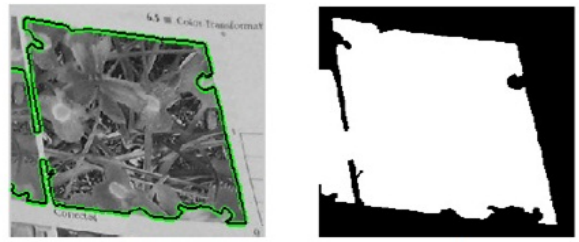
Figure-4. Notations used in segmentation: R represents the bounding box and Γ represents four bounding curves $\gamma_T, \gamma_B, \gamma_R$ and γ_L and R(Γ) is the region corresponding to printed picture.

- Energy minimization function is given by
  $$E = arg\ min\Phi(\Gamma,\ \theta)$$
  Where $\Gamma = (\gamma_T, \gamma_B, \gamma_L, \gamma_R)$ are the four curves.
  Where θ is the set of parameters for GMMs.
- Representation of the curves:
  $$x = \gamma_L(y)$$
  $$x = \gamma_R(y)$$
  $$y = \gamma_T(x)$$
  $$y = \gamma_B(x)$$
- Energy function =Region Term + Bounding Term
  $$\Phi(\Gamma,\ \theta) = \Phi_L(\Gamma,\ \theta) + \lambda\Phi_S(\Gamma)$$
  Where λ is a constant.
- Boundary term determines whether the pixels lie on the boundary or not.
- Region term determines whether the pixels are foreground pixels or background pixels.

Active contour models have been extensively applied to image segmentation. The algorithm 1 explains how to segment the input image to obtain optimized boundaries [5].

Algorithm 1-Segmentation using energy minimization

1. Define an energy function.
2. Generate an initial boundary by selecting edge points.
3. Compare energies of each point on the boundary with the energy of points in its neighborhood.
4. For each current boundary point, move the boundary to the neighboring point that has the lowest energy.
5. Perform steps 3 and 4 on all points on the boundary to complete one iteration.
6. Repeat steps 3, 4 and 5 until it causes no further movement of the boundary.



(a) (b)

Figure-5. Segmentation (a) Boundary detected (b) Segmented output as binary image.

Figure-4 represents the notations used in the segmentation process. Figure-5(a) shows the boundary detected on the input image and 5(b) shows the segmented binary image.

Feature Extraction

The Progressive Probabilistic Hough transform [6] is a technique used to detect lines, curves, and objects in an image using the concept of parameter space. The polar representation of the line in Hough space is shown in Figure-6.

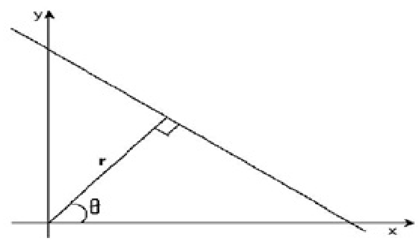
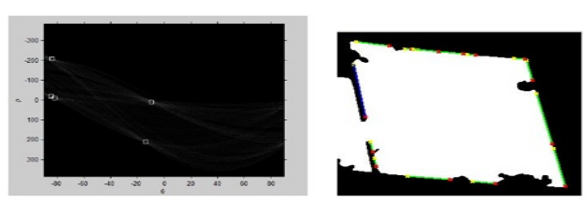


Figure-6. Polar form of the Hough transform for the line.

The linear Hough transform algorithm uses a two dimensional array, called an accumulator, to detect the existence of a line described by

$$\rho = x\cos\theta + y\sin\theta$$

For each pixel at (x, y) and its neighborhood, the Hough transform algorithm determines if there is enough evidence of a straight line at that pixel. If so, it will calculate the parameters (ρ, θ) of that line, and then look for the accumulator's bin that the parameters fall into, and increment the value of that bin. By finding the bins with the highest values, typically by looking for local maxima in the accumulator space, the most likely lines can be extracted [7]. In Figure-7(a) Hough-matrix is shown and in Figure-7(b) boundary extracted image is shown.



(a) (b)

Figure-7. Feature Extraction (a) Hough matrix (b) Boundary points extracted.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Algorithm 2 describes the PPHT method to extract the boundary.

Algorithm 2-Boundary extraction using PPHT

1. Edge points along the boundaries are considered.
2. Select a set of n edge points.
3. Initialize accumulator H to all zeros.
4. Choose a point from the set.
5. Compute $\rho = x\cos\theta + y\sin\theta$ and update the accumulator.
6. Remove the point from the set.
7. Check if the highest peak in accumulator is greater than threshold $p = \left\lfloor \frac{n}{2} \right\rfloor - 1$, if not goto 4.
8. The detected line in the image is given by $\rho = x\cos\theta + y\sin\theta$
9. Look along the corridor specified by peak in accumulator and check if the longest segment that has gaps less than a threshold $g = \left\lfloor \frac{n}{2} \right\rfloor$. If not, goto step 3.
10. If the line segment is longer than the minimum length, $l = n - 1$ add it to the output list, else goto step 4.
11. Check if the input set is empty. If not, goto step 4.
12. Unvote from the accumulator for the pixels that was previously voted.
13. Check if all the edge points are processed, if not goto 2.
14. Output the extracted lines.

Rectification

In the dewarping of the distorted image, Coons patch bilinear interpolation method [10] is used. By extracting the boundary points, the mapping of distorted image to an undistorted image is possible. With the help of a 2D bilinear blending function the mapping is performed [11]. Algorithm 3 shown is the rectification method used to dewarp the distorted image. Figure-8 shows the distortion rectified image.

Figure-8. Distortion rectified image.

Algorithm 3- Rectification of distorted image

1. Given the points along the boundaries.
2. The boundary curves $\gamma_T, \gamma_B, \gamma_L$ and γ_R are obtained where γ_i =[x(t) y(t)], t ∈ [0; 1]
3. Geometric correction can be performed by point-wise Coons bilinear blending function given by

$$\gamma(u',v') = \begin{bmatrix} 1-u & u \end{bmatrix} \begin{bmatrix} \gamma_L(v) \\ \gamma_R(v) \end{bmatrix} + \begin{bmatrix} \gamma_T(u) & \gamma_B(u) \end{bmatrix} \begin{bmatrix} 1-v \\ v \end{bmatrix}$$

$$- \begin{bmatrix} 1-u & u \end{bmatrix} \begin{bmatrix} \gamma_T(0) & \gamma_R(0) \\ \gamma_B(1) & \gamma_L(1) \end{bmatrix} \begin{bmatrix} 1-v \\ v \end{bmatrix}$$

Where $u' \in [0,1]$ and $v' \in [0,1]$

4. The dewarped image $\leftarrow \gamma(u',v')$.

Super-Resolution

A low-resolution image obtained after rectification is converted to a high resolution image by using an image interpolation technique. An edge directed interpolation is employed to find the missing pixels along the edge direction. Edge-directed bi-cubic interpolation [14] interpolates to find the missing pixels in the same direction (horizontal or vertical) and so non-horizontal or non-vertical edges are smoothed. The low resolution image having size NxN is converted to high resolution with size 2Nx2N. Figure-9 shows the super-resolved image as final output.

Figure-9. High-resolution image.

QUANTITATIVE MEASUREMENT

Quantitative measurement is to examine the measurement equivalence. The different quantitative assessments are MSE, PSNR and SSIM. The difference with respect to other techniques such as MSE or PSNR is that these approaches estimate perceived errors; on the other hand, SSIM considers image degradation as perceived change in structural information.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Structural Similarity

The structural similarity (SSIM) index is a method for measuring the similarity between two images. The SSIM index is a full reference metric that is, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. The formula for calculating the similarity measure between two windows x and y for a window of size nxn is given by:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

Where μ_x is the mean of x,

μ_y is the mean of y,

σ_x^2 is the variance of x,

σ_y^2 is the variance of y,

σ_{xy} is the covariance of x and y,

$c_1 = (K_1L)^2$ and $c_2 = (K_2L)^2$ are two variables to stabilize the division with weak denominator,

L is the dynamic range of the pixel-value, L=255.

K_1=0.01 and K_2=0.03 by default.

The resultant SSIM index is a decimal value between -1 and 1, and value 1 is only reachable in the case of two identical sets of data. Typically it is calculated on window sizes of 8x8. The window can be displaced pixel-by-pixel on the image.

Table-1 and Table-2 shows the comparison of structural similarity measures of distorted image and rectified image with an undistorted reference image for the test cases in Figure-10 and Figure-11.

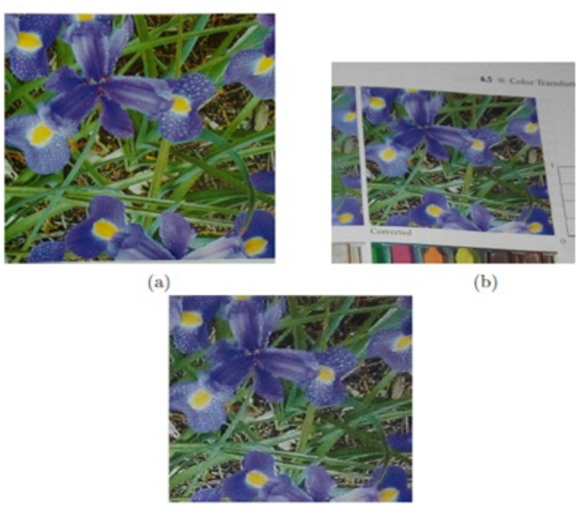


Figure-10. Test images (a) Undistorted reference image (b) Distorted image (c) Rectified image.

Table-1.

Images Compared	SSIM Value
Undistorted Reference Image Vs Distorted Image	0.1144
Undistorted Reference Image Vs Rectified Image	0.1737

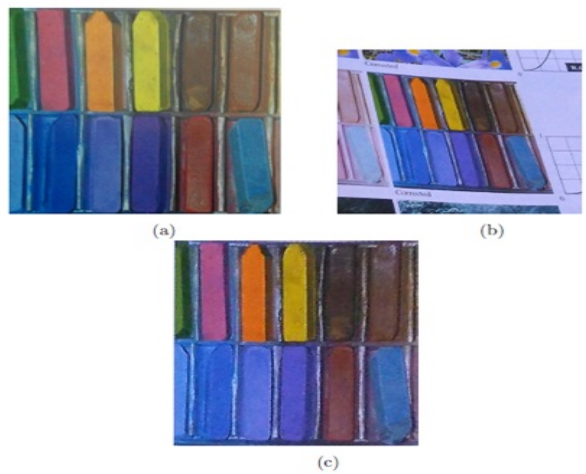


Figure-11. Test images (a) Undistorted reference image (b) Distorted image (c) Rectified image.

Table-2. Structural similarity index for Figure-11.

Images Compared	SSIM Value
Undistorted Reference Image Vs Distorted Image	0.1859
Undistorted Reference Image Vs Rectified Image	0.3838

The above examples show that the undistorted image is structurally more similar to the rectified image as the structural similarity value increased.

CONCLUSIONS

It is a real time method for the images of printed documents containing figures and photos, which can be rectified and enhanced. The methodology applies an approach to rectify the distortions in images of printed document images. The aim is to selectively extract the figures and photos in document images and rectifying them. The process of segmentation involves the setting of an initial boundary. Comparing the energies of neighboring pixels and optimizing the boundary gives a binary output. Using the segmented binary image, the boundary is extracted with the help of PPHT. The rectification is performed with Coons patch bilinear blending function to obtain the dewarped image. Not only the rectification but also enhancing the output image to a higher resolution is performed. By applying an edge-directed super-resolution technique, a higher resolution image is obtained.

The structural similarity measured between the undistorted image and the rectified shows an increased similarity value. The application for the method is that it can rectify any type of Figures and photos in document images like books, posters, panels etc.

REFERENCES

[1] Hyung Il Koo. 2013. "Segmentation and Rectification of Pictures in the Camera-Captured Images of Printed Documents",IEEE transactions on multimedia.

[2] Huaigu Cao, Xiaoqing Ding and Changsong Liu. 2003. "Rectifying the Bound Document Image Captured by the Camera: A Model Based Approach", Proceedings of the Seventh International Conference on Document Analysis and Recognition.

[3] F.Zirari, A.Ennaji and S.Nicolas D.Mammass. 2013. "A simple text/graphic separation method for document image segmentation", Proc. of International Conference on Computer Vision.

[4] Panos M. Pardalos and David Shalloway. 1993. "Optimization Methods for Computing Global Minima of Nonconvex Potential Energy Functions", DIMACS Technical Report.

[5] Chunming Li, Chiu-Yen Kao, John C. Gore and Zhaohua Ding. 2008. "Minimization of Region-Scalable Fitting Energy for Image Segmentation" , IEEE Transactions on Image Processing, Vol. 17, No. 10.

[6] Richard O. Duda and Peter E. Hart. 1972. " Use of the Hough Transformation To Detect Lines and Curves in Pictures", Communications of the ACM, January, Vol. 15, No. 1.

[7] C. Galambosi J. Matasts and J. Kittlert. 1999. "Progressive ProbabilisticHough Transform for line detection", IEEE.

[8] Michael S. Brown and W. Brent Seales. 2001. "Document Restoration Using 3D Shape: A General Deskewing Algorithm for Arbitrarily Warped Documents", IEEE.

[9] N. Stamatopoulos, B. Gatos, I. Pratikakis and S. J. Perantonis. 2008. "A Two-Step Dewarping of Camera Document Images", The Eighth IAPR Workshop on Document Analysis Systems, IEEE.

[10] Yau-Chat Tsoi and Michael S. Brown. 2004. "Geometric and Shading Correction for Images of Printed Materials A Unified Approach Using Boundary", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[11] Wenfa Qi, Xiaolong Li and Bin Yang. 2009. "Bilinear Coons patch and its application in security pattern design", Fifth International Conference on Intelligent Information Hiding andMultimedia Signal Processing.

[12] Adrian Ulges, Christoph H. Lampert and Thomas M. Breuel. 2006. "Document Image Dewarping using Robust Estimation ofCurled Text Lines", IEEE.

[13] Huaigu Cao, Xiaoqing Ding and Changsong Liu. 2003. ",A Cylindrical Surface Model to Rectify the Bound Document Image", Proceedings of the Ninth IEEE International Conference on Computer Vision.

[14] Zhou Dengwen. 2010. An Edge-Directed Bicubic Interpolation Algorithm, 3rd International Congress on Image and Signal Processing.