International Journal of Applied Research 2016; 2(3): 705-710

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2016; 2(3): 705-710
www.allresearchjournal.com
Received: 22-01-2016
Accepted: 26-02-2016

Nabanita Basu
Department of Computer Science and Engineering, University of Calcutta, West Bengal, India.

Samir K Bandyopadhyay
Department of Computer Science and Engineering, University of Calcutta, West Bengal, India.

Correspondence
Nabanita Basu
Department of Computer Science and Engineering, University of Calcutta, West Bengal, India.

Automatic perspective rectification of documents photographed with a camera

Nabanita Basu, Samir K Bandyopadhyay

Abstract
Using an incorrect lens, or camera shake can cause the perspective of photographs to be tilted or skewed. The perspective may be distorted, and is more evident in photographs containing continuous vertical lines or geometric shapes. This paper deals with perspective distortion together with skew distortion rectification of a document image file to make its content more easily comprehendible for an end user.

Keywords: Sobel Operator, Corner detection, Segmentation, *Vanishing point detection* and Perspective Rectification

Introduction

In photography, there are two types of distortions: optical and perspective. Both result in some kind of deformation of images – some lightly and others very noticeably. While optical distortion is caused by the optical design of lenses (and is therefore often called "lens distortion"), perspective distortion is caused by the position of the camera relative to the subject or by the position of the subject within the image frame. And it is certainly important to distinguish between these types of distortions and identify them, since you will see them all quite a bit in photography. The goal of this article is to explain each distortion type in detail, with illustrations and image samples.

Perspective distortion has nothing to do with lens optics and thus, it is not a lens error. When projecting three dimensional space into a two dimensional image, if the subject is too close to the camera, it can appear disproportionately large or distorted when compared to the objects in the background. This is a very normal occurrence and something you can easily see with your own eyes. If you take a smaller object like your mobile phone, then bring it very close to your eyes, it will appear large relative to say your big screen TV in the background (and the farther your phone is from your TV, the smaller the TV will appear relative to your phone). The same thing can happen when photographing any subject, including people.

Factors that Induce Perspective Distortion are:

- The 'Angle of view ' at which the image is captured by the camera
- The angle of view at which the photograph of the subject is presented or viewed

Document scanner is widely used to capture text and transform it into electric form for further processing. As camera resolution rises in recent years, high-speed non-contact text capture through digital cameras is becoming an alternative choice. Unfortunately, perspective distortion coupled with captured images by digital camera brings up a new problem to the traditional optical character recognition (OCR) system. Similar to the skew compensation operation required after the scanning process, perspective distortion must be removed before the document image is fed to the OCR system.

Skew distortion is rotation-induced, which means that the top line and base line of each text line are still parallel to each other within the scanned document image. Therefore, none of them can handle the perspectively distorted document images where the parallel relation between top line and base line of text lines is totally destroyed.

In present day India, people from all walks of life have access to mobile phones with inbuilt camera device. Most of these mobile phones are economically priced and call for low maintenance. Taking images of pets, quotable quotes, friends, content of books etc. is a

'trend' among most youngsters. However, images clicked are often subjected to skew distortion, perspective distortion, blurring and other sorts of noise artefacts.

It often happens that students are not allowed to carry home certain textbooks, magazines etc. from the University library. On a few other occasions, students do not have access to a Xerox machine near the University premises. Again, students might not have access to the library over the weekend or for extended hours on working days. This work is largely aimed to help students to study at a pace suitable to them without having to worry about book return and availability deadlines at the library.

Related Works

The use of portable image readers over fixed-mount image readers is increasing and these portable image readers are seeing applications in many industries. One of the main challenges with portable image readers however is the perspective distortion caused by inconsistent image reading positions. With fixed-mount systems, such as a document scanner, the image reader is placed in such a manner that the optical path of the image reader is perpendicular to the image plane. With portable systems, however, the position of the image reader is dependent on a human operator. It is difficult for an operator to know the ideal point from where to capture an image of a target such as a document. More often than not, the user captures the image at an oblique angle, i.e. the image reader is not in a plane parallel to the plane of the document, and the captured image is skewed.

In 2011 Tian Jipeng has introduced a novel method for skew correction using Hough Transform. Here they have used the voting procedure to detect the straight-line. This approach can detect skew angle up to -90 to +90 degrees [1]. In 2013 Mandip Kaur, Simpel Jindal has introduced the integrated skew detection and correction using Fourier transform. In this method they applied DCT compression and thresholding technique on image. So the time required for recognition is decreased. Then after that the Fourier spectrum is obtained. Here the spectrum is split into four equal parts and the detected skew angle of all parts is measured. At the last the input image is rotated using the Bilinear Interpolation method [2].

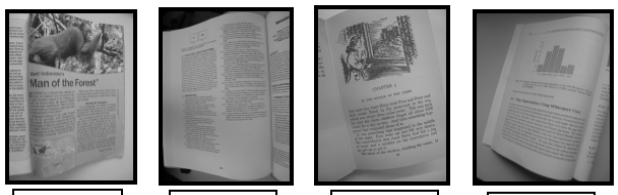
In 2013 Prakash K Aithal, Rajesh G, U Dinesh Acharya has presented the skew estimation approach using Radon Transforms. This approach gives faster and better results as compared to other methods [3]. In 2007 Shutao Li, Qinghua Shen and Jun Sun has proposed projection profile analysis technique for skew detection using wavelet decomposition method. Here first the skewed document image is decomposed using wavelet transform and then a projection profile is used to measure the skew angle [4].

In 2003 Yue Lu and Chew Lim Tan has introduced improved Nearest Neighbor based approach to accurate Skew detection. Here the largest possible number of nearest-neighbor pairs is selected, and their curves are measured to detect the skew angle of document image [5]. In 2003 Yang Cao and Shuhua Wang has introduced a technique for skew detection and correction based on straight line fitting. And a concept of Eigen point is also introduced. Here the relationship between successive Eigen points in every text line is analyzed from a particular sub-region and the Eigen points from most suitable to lay on the baselines of the selected samples for the straight-line fitting. Then the average of all these baseline directions is computed. Hence the degree of skew of the complete document image is detected [6].

In 2009 Atallah Mahmoud Al-Shatnawi and Khairuddin Omar has been introduced a skew detection and correction method of document image based on the Centre of gravity. The proposed work contains an accessing the text into the document by an arbitrary polygon and deriving the baseline of polygon's centroid [7]. In 2012 A.V. Narasimha Rao, M. Ramanjaneyulu and M. Devaraju has introduced a technique permutation invariant RAS Transform for skew character recognition. This approach is based on a permutation invariant Rao Alaka Shift Transformation (RAST) and implemented for several printed alphanumeric characters in different fonts [8].

Data Set

The first dataset, to our knowledge, of camera-captured document images is the "DFKI-1 dataset", proposed by Shafait *et al*. [9] and its evolution, the "IUPR dataset" [10]. Each dataset is composed of approximately 100 images, and represent many different documents classes captured in different situations with various distortions: perspective, warping, lighting, curl, skew, etc. The ground truth associated with each image is manually created and contains, along with a scanned version of each page, pixel-level accurate tagging of lines, text zones, and content type. ASCII text is also available. During the de-warping contest at CBDAR 2007 [3], de-warping methods were compared against what a commercial OCR system would produce on de-warped images. Such approach is not viable in our case, as we cannot rely on the presence of a majority of textual content within each document. Furthermore, the manual annotation of each document image is not feasible when dealing with tens of thousands of video frames. The data set from IUPR is presented below.



Doc 071 Doc 044 Doc 006 Doc 074

Document ID	Camera Setting		Document Content	
	Viewing Angle	Mega Pixel	Book Type	Book Thickness
Doc 071	Left	15	Magazine	1.5 cm
Doc 044	Right	15	Conference Proceedings	2.5 cm
Doc 006	Top-right	9	Old Story Book	2.0 cm
Doc 074	Bottom-Right	15	Bound Pages (Technical)	1.0 cm

Proposed Method

In computer vision, most perspective correction methods rely on vanishing point detection. And these methods involve extracting multiple lines and their intersections, or using texture and frequency knowledge this paper presents a method for automatic perspective correction on distorted images. The flow of the overall methodology used to rectify the distorted image is given in Figure 1.

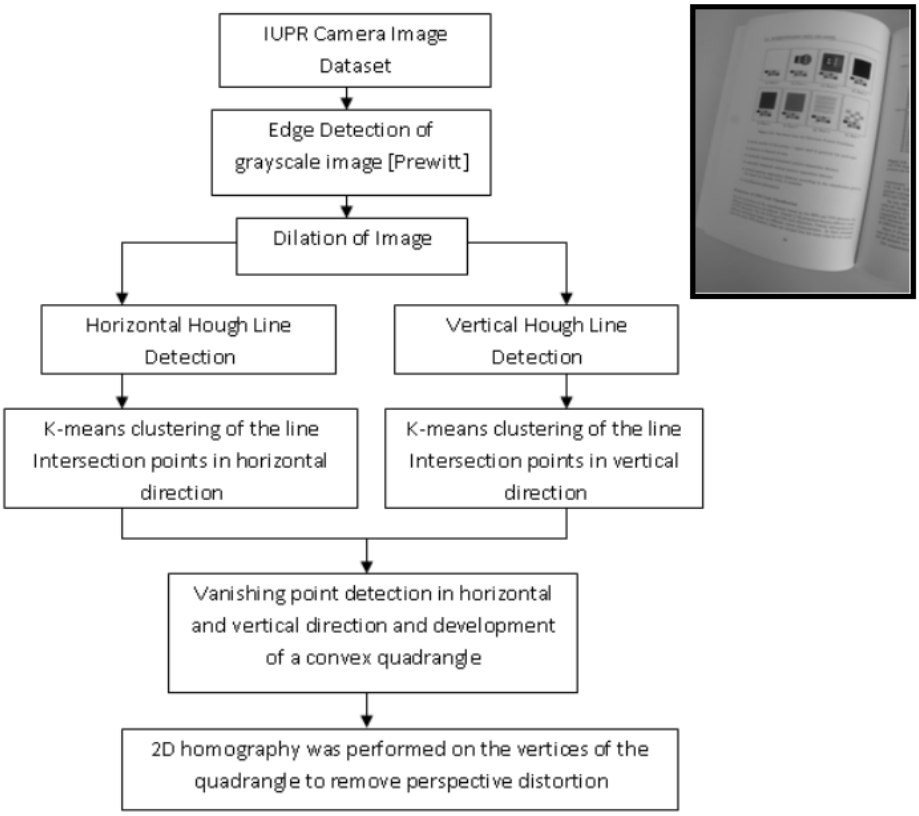


Fig 1: Flow of the overall methodology

Images captured in pinhole cameras are perspective in nature. In a perspective image, objects of similar size appear bigger when it is closer to view point and smaller when it is farther. Hence the number of pixels to represent farther objects in the image requires lesser pixels as compared to the number of pixels required to represent the closer objects in the image. Figure 2 describes the basic difference between both views. Thinning is a process used to remove selected foreground pixels from binary images. It is a morphological operation on binary images. Thinning is applied only on binary images and it produces another binary image. It is used for edge detection by reducing all lines to a single pixel thickness. It can also use for further processing such as skeletonization. Smoothing is the process that is mainly used to decrease the noise in an image or to produce a less pixilated image. It can also be used to prepare image for further processing such as Segmentation.

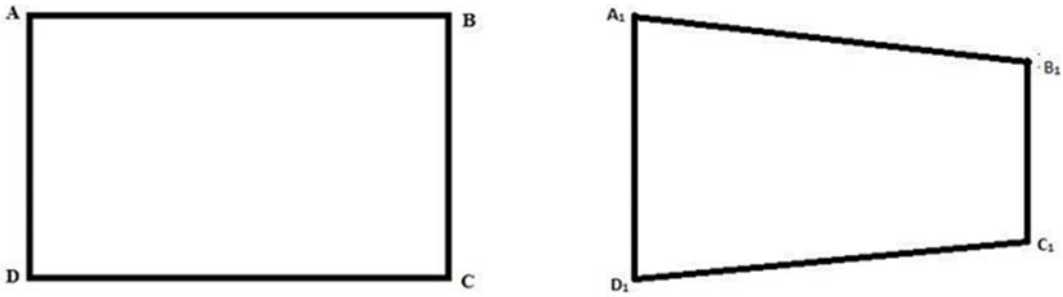


Fig 2: (a) Rectangular Plane (b) Perspective distortion

Pre-processing is performed to segment the portion to be corrected. The desired portion is segmented and further used in corner estimation. The steps involved are as follows:

a) The given colour image is converted to gray image
b) Skew Detection using Hough Transform (HT)
c) Apply sobel operator to detect edges.
d) Segment the image by applying the dilation and erosion methods

The first step in image analysis was the conversion of the RGB image into grayscale using standard techniques. After grayscale conversion, the noise within the image was smoothed using a median filter. A median filter modifies an image pixel such that the new pixel value is selected as the median of the neighbouring values within a range. This filtering has the effect of removing random noise in images while still maintaining the overall integrity of image regions and boundaries. This type of filter is typically categorizes as a blurring filter in many image processing toolkits.

In the next step removal of watermark and shadow effects in image are done. Also extract of foreground object while removing the unwanted background. The result is shown figure 3.

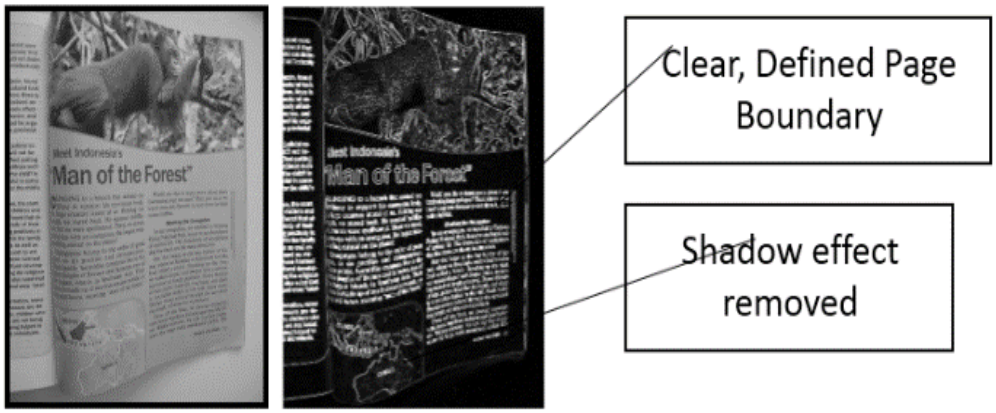


Fig 3: Shadow Effects

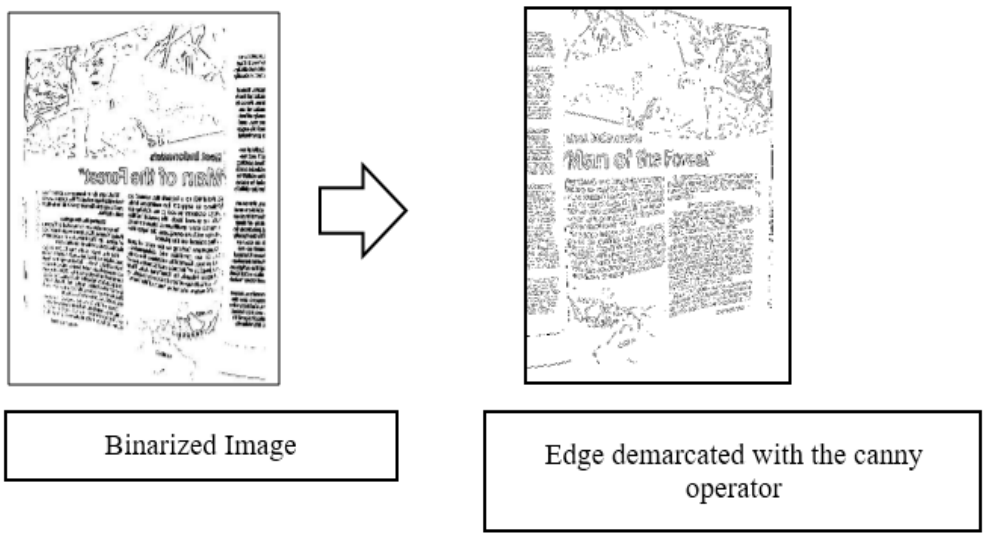


Fig 4: Image Binarization and Edge Detection.

Hough transform is a feature extraction technique that is used to isolate /find imperfect instances of objects within a certain class of shapes by a voting procedure'. Hough line transform particularly deals with the identification of straight lines in an image. The lines identified by hough line transform algorithm are commonly referred to as hough lines.

The Hough transform is a feature extraction technique. The classical HT was related with the identification of lines in the image, but later the HT has been used to identify positions of arbitrary shapes such as circles or ellipses. The HT provides a better technique for skew detection in document images. The classical Projection Profile approach for skew detection works like, first the input image is rotated to a range of angle and then from each angle projection profile is calculated. After that the feature are extracted from each projection profile and the skew angle is calculated.

The Modern Projection Profile is a histogram method of skew detection. The histogram of a document image is taken. It contains number of pixels from the document image and projection profile. The document image takes the highest peak when the skew angle is zero. That approach is not suitable for skew detection of document images contains multiple fonts or non-texts such as, images, figures, and table's etc. Algorithm to detect skew using HPP as follows,

Algorithm

1. Input image.
2. Find out the HPP θ ranging from -90to 90 degree.
3. Find the peak point from HPP for each angle.
4. The highest peak is considered as θ, when the skew angle is zero.
5. Find out the rotation from maximum peak.
6. Calculate the rotation angle as the skew angle.

Morphology is a broad set of image processing operations that process images based on shapes. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries.

To add robustness to the system, the horizontal and vertical character strokes along with the boundary lines have been used to create the final set of hough lines. For perspective correction using Hough lines, two cases are considered and these are:

- Extreme most Vertical or Horizontal extended lines that do not intersect when extended substantially.
- Extreme most Vertical or Horizontal extended lines that have multiple intersections when extended substantially.

Segmentation is the process of dividing the digital image into different segments. For vanishing point detection in document analysis, most researchers directly voted vanishing points with intersected lines or other information on the image plane. In graphical perspective, a vanishing point is a point in the picture plane that is the intersection of the projections (or drawings) of a set of parallel lines in space on to the picture plane. Firstly, we clustered line intersections to vanishing points on the image plane.

In the above two cases the following observations are made:

- The two extreme most Hough Lines are parallel to each other, therefore these lines only coincide at infinity.
- The extreme-most horizontal lines intersect, hence the intersection points of the horizontal lines are clustered using k-means clustering algorithm.

K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
In this case, hough line intersection points are clustered into 3 groups using k-means clustering algorithm, to find the mean point of the cluster that has the maximum number of intersection points.
In a horizontal document, each text row has a clue horizontal direction. Our method uses document boundaries and text rows to detect the horizontal vanishing point. However, in the vertical direction of a horizontal document, there will be no text columns for vertical clues. In our case, two situations are considered. These are:

- The extreme- most parallel vertical parallel lines are drawn.
- The horizontal lines joining the rightmost corner points of the image with the vanishing point detected on the right of the image are extended across the breadth of the image.

To build a parallel projection of the image from the perspectively distorted image, the plane homography is used. The homography can be computed by knowing the relative positions of the four points on the perspectively distorted image and the positions of the transformed image to be constructed. The four corner points are required to implement the perspective transformation. The corner method detects all the available corner points. Perspective distortion occurs due to the perspective projection of 3D scene on a 2D surface. Correcting the distortion of a single image without losing any desired information is one of the challenging task in the field of Computer Vision. The following figure 5 illustrates this concept.

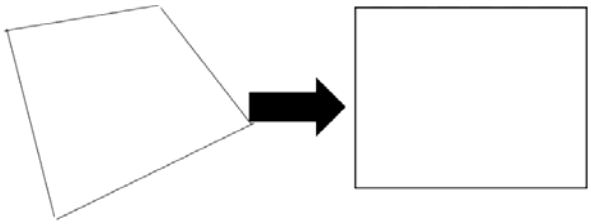
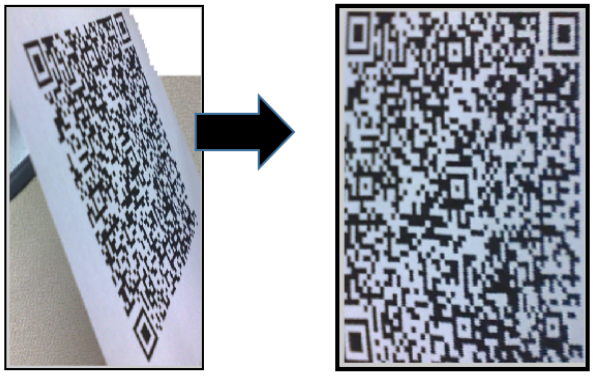


Fig 5: 2D Homography

Planar Homography can be used to correct camera perspective to obtain clear and straight image of the number plate. Perspective rectification is performed by determining position of 4 corner points of the object in the image and finding the homography matrix that defines the movement of those corner points at new location.
The transformation can be summarized as follows:

a) Input the segmented image and the four corner points.
b) The projective transformation is applied on the coordinates of the input image.
c) The resultant image contains the transformed view volume of the rectified image.

The distorted and the rectified images are given in the Figure 6 and Figure 7.



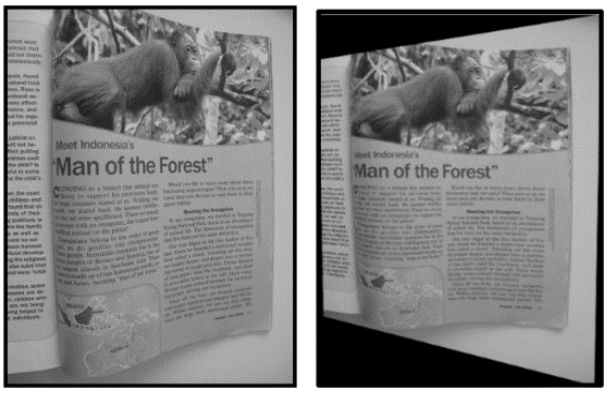


Fig 6: Perspective Corrected Image

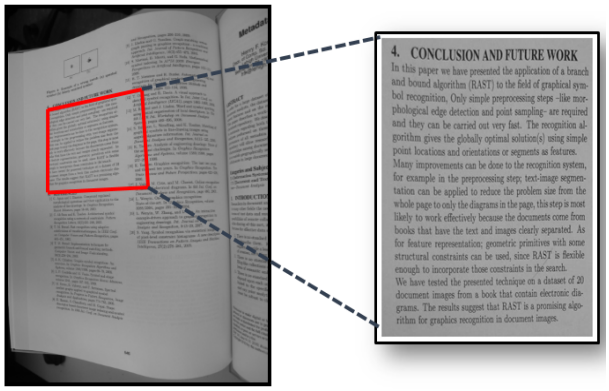
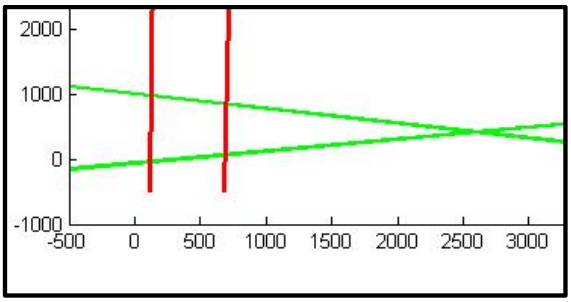


Fig 7: Perspective distortion correction of other similar images using the same methodology

The research presented within this paper reflects the use of computer vision techniques to extract important information from 2D images of roads. This work also detailed the use of a secondary Hough Transform to detect parameter space ridges within Hough Transform parameter space. These secondary HT parameter space maxima allowed the collection of multiple parameter space lines into sets for common analysis. This is explained in figure 8.



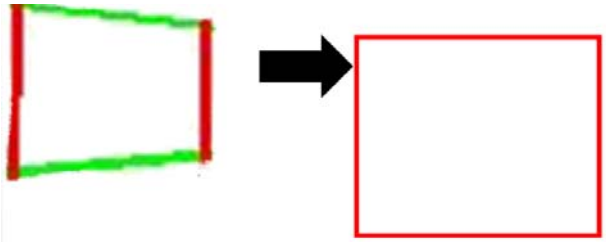


Fig 8: 2 D Homography

If the document image contains only text then it requires more time than the document containing only pictures. If the document image containing both text and pictures such as figures, tables, images then it requires more time to recognize it and it will also affect the

performance of recognition. The result accuracy is also depends upon the source. If the document is machine printed images it will give better result than the scanned document image.

Conclusions

Document analysis is one of the most challenging and interesting research area in the field of document image processing and pattern recognition and has a more demand in the OCR (Optical Character Recognition) systems. We consider the problem of estimating perspective distortion from a single still image of an unstructured environment and to make perspective correction which is both quantitatively accurate as well as visually pleasing. Corners are detected based on the orientation of the image. A method based on plane homography and transformation is used to make perspective correction.

References

1. Tian Jipeng G, Hemantha Kumar, Chethan HK. Skew correction for Chinese character using Hough Transform, International Journal of Advanced Computer Science and Applications, Special Issue on Image Processing and Analysis. 2011; 22:33-36.
2. Mandip Kaur, Simpel Jindal. An Integrated Skew Detection and Correction Using Fast Fourier Transform and DCT, International Journal of scientific & technical research. 2013; 2:164-171.
3. Prakash Aithal K, Rajesh G, Dinesh Acharya U. A Fast and Novel Skew Estimation Approach using Radon Transform", International Journal of Computer Information Systems and Industrial Management Applications. 2013; 5:337-344,
4. Shutao Li, Qinghua Shen, Jun Sun. Skew Detection Using Wavelet Decomposition and Projection Profile Analysis, Pattern Recognition Letter 2007; 28:555-562.
5. Yue Lu, Chew Lim Tan. Improved Nearest Neighbor Based Approach to Accurate Document Skew Estimation International Conference on Document Analysis and Recognition 2003; 1:503-507.
6. Yang Cao, Shuhua. Skew Detection and Correction in Document Images Based on Straight-Line Fitting, Pattern Recognition Letters 2003; 24:1871-1879.
7. Atallah Mahmoud, Al-Shatnawi, Khairuddin. Skew Detection and Correction Technique for Arabic Document Images Based on Centre of Gravity, Journal of Computer Science. 2009; 5:363-368,
8. Narasimha Rao AV, Ramanjaneyulu M, Devaraju M. Skewed Character Recognition using Permutation Invariant RAS Transform, International Journal of Computer Applications. 2012; 3:37-41.
9. Shafait F, Breuel TM. Document image dewarping contest, in 2nd Int. Workshop on Camera-Based Document Analysis and Recognition (CBDAR), 2007, 181-188.
10. Bukhari SS, Shafait F, Breuel TM. The IUPR dataset of camera-captured document images, in Camera-Based Document Analysis and Recognition 2011, ser. Lecture Notes in Computer Science, M. Iwamura and F. Shafait, Eds. Springer Berlin Heidelberg 2012; 7139:164-171.