Rectification of Perspective Text Images on Rectangular Planes

Huy Phat Le , Kavitha Madhubalan, GueeSang Lee*
Department of Electronics and Computer Engineering
Chonnam National University, Gwangju, 500-757, Korea

ABSTRACT

Natural images often contain useful information about the scene such as text or company logos placed on a rectangular shaped plane. The 2D images captured from such objects by a camera are often distorted, because of the effects of the perspective projection camera model. This distortion makes the acquisition of the text information difficult. In this study, we detect the rectangular object on which the text is written, then the image is restored by removing the perspective distortion. The Hough transform is used to detect the boundary lines of the rectangular object and a bilinear transformation is applied to restore the original image.

1. INTRODUCTION

Images taken in a natural scene often contain structural distortions or deformations. An obvious example of a nonlinear distortion can be found in a photograph taken by a camera at an oblique angle. If the direction of the camera is perpendicular to the object plane, a rectangular digitized image without perspective distortion is produced. However, this condition is not always satisfied and it is common for nonlinear distorted images to be generated.

Distortion in an image can result in a loss of image information such as text, which can be important for many applications. Therefore, we need to restore those images with perspective distortions. The restoration phase can be considered as pre-processing steps for the detection and recognition of the image information.

Several researchers have developed algorithms for the restoration of perspective images. G.K. Myers *et al.* [1] proposed a method of extracting each text line, which is then rectified independently. C.R Dance *et al.* [2] estimates the vanishing points by sets of parallel lines which are text lines and formatted column boundaries. These vanishing points are then used for perspective correction.

S.J. Lu *et al.* [3] used the character stroke boundaries and tip points to rectify document images when the document boundary and paragraph formatting information are not available.

Lin Liu *et al.* [4] proposed a method of detecting the distortion of a vehicle license plate by categorizing the distortion into two main types, which are the horizontal slant and vertical slant. For the horizontal slant, the connected component on the binarized image is used to select the digits in the license plate and then a technique called straight-line fitting is employed to rotate the original license plate image. For vertical slant correction, this method is based on rotation experiments with various angles. Self-organizing feature maps (SOFMs), which are a type of Neural Network, along with the Coons surface are employed to estimate the boundary of the distorted image [5]. Then, a nonlinear transformation is used to remove the perspective projection in the images. The study of the nonlinear shape restoration by Y.Y. Tang *et al.* [6] based on the shape transformation is one of the recent and successful approaches. The key idea of Tang's method is to find an appropriate analytical shape transformation to model the nonlinear distortion (distortion function) which transforms the original image to the distorted one. Once we know the distortion shape function, we can infer the inverse distortion function, which transforms the distorted image to the original one. However, to find the distortion function, some a priori information, such as the shape of the original object or the corresponding points between the original distorted object and the restored object, must be known. Finding the exact distortion function for an arbitrarily distorted image is nearly impossible. Therefore, some techniques require the shape of the object to be specified in advance, e.g., a rectangular shape. When the shape function is identified, the next step is to compute the reference points, defining the shape of the object, in the distorted image. The more exact the reference points are, the more correctly the distorted image is restored.

In this paper, we propose an algorithm to detect and restore the perspective rectangular object in the image. The flow of the proposed algorithm is illustrated in Fig. 1.

---
* *Corresponding author. E-mail : gslee@jnu.ac.kr*

pre-processing step
+ Removing non line-like edge

Detect perspective rectangular image
+ Detecting line candidates
+ Detecting four lines

Resorting by bilinear transformation
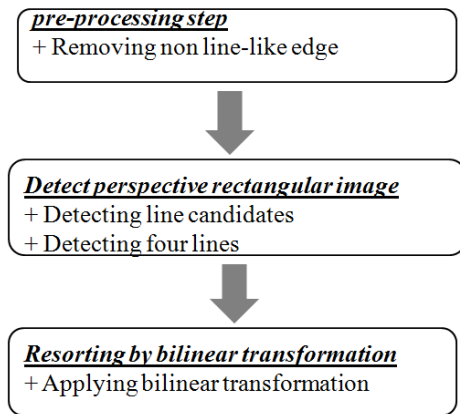+ Applying bilinear transformation

Fig. 1. Flow of detection and restoration

The remainder of this paper is organized as follows. In section 2, we analyze the problem of the perspective projection camera based model. An overview of nonlinear shape distortion is given in section 3. In section 4, we describe the way to remove noise in the edge image. The detection and restoration of the perspective rectangular image are presented in sections 5 and 6, respectively. Some experimental results are discussed in section 7.

2. THE PROBLEM OF PERSPECTIVE PROJECTION

In this section, we analyze the cause of the existing problem of the perspective projection in which digital camera results.

There are many models of imaging devices that are created to capture 3D scenes. The perspective projection model is a simple one that projects a 3D scene toward a single point called a pinhole camera or centre of projection. The light travelling toward the scene object passes through the pinhole and creates an image on the image plane.

The coordinate system (O, i, j, k) attached to the pinhole camera, whose origin O coincides with the pinhole, and vectors i and j form a basis. The basis vector k is orthogonal to the image plane. When vector k is not orthogonal to the scene object, a perspective effect will occur, in which the images of the parallel lines intersect with each other in the image plane. Fig. 2 illustrates the case where the upper and lower lines of a rectangle in the image plane are not parallel to each other.
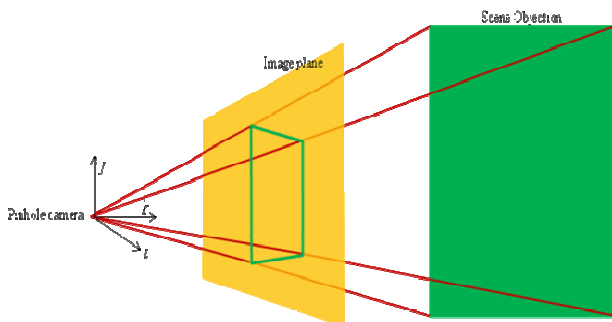

Fig. 2. The effect of perspective projection

In the process of capturing a scene image using a digital camera, we assume that the position of the camera is a pinhole. The direction of the camera toward the scene object is the optical axis, which is regarded as vector k. When this direction is not orthogonal to the scene object, the above perspective effect occurs. Therefore, if the scene object has a rectangular shape, the received image in the image plane does not have the same shape. In general, the shape of the received image in the image plane is changed from its original shape. In this paper, we regard this problem as shape distortion.

3. APPROACH FOR RECTIFICATION OF PERSPECTIVE PROJECTION IMAGE

As mentioned in section 2, the shape distortion problem always occurs in the data acquisition phase, due to perspective projection. Shape distortion can be divided into two categories: linear and nonlinear. Linear shape distortion such as rotation, scaling, translation and reflection was successfully solved in [7]. However, the correction of nonlinear shape distortion remains a challenging topic [6]. Many approaches to the correction of nonlinear shape distortion have been proposed. In this section, we outline some of them.

3.1 Rectification by the Orientation Angles

Assuming the text is detected, the orientation or the perspective angles of the text is analyzed by recognizing text that lies on a planar surface in 3-D space [1]. The authors modeled the orientation angle of such text relative to the camera in terms of three angles, as illustrated in Fig. 3(b). These angles are

+ θ, the rotation in the plane perpendicular to the camera's optical axis, and

+ φ and γ, the horizontal (azimuth) and vertical (elevation) components, respectively, of the angles formed by the normal to the text plane and the optical axis.

These three angles represent the extent of rotation that the text plane must change relative to the camera in each of its three axes to yield a frontal, horizontal view of the plane in the camera's field of view. However this method assumes the detection of text before the rectification, so its rectification performance is heavily depends on the performance of text detection.
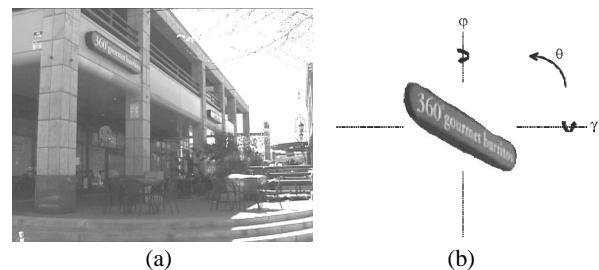

(a)   (b)

Fig. 3. Café scene[1] : (a) captured image.
(b) Orientation angles of text

3.2 Vanishing Point based Approach

Studying the vanishing points provides a strong clue for inferring some information about the 3D structure of a scene. With a perfect perspective projection of the imaging device, e.g. with a pinhole, the parallel lines in the 3D scene are projected onto lines in the image plate that meet at a common point. This point of intersection is called the vanishing point. The blue point in Fig. 4 is an example of a vanishing point when a box is projected onto an image plate.
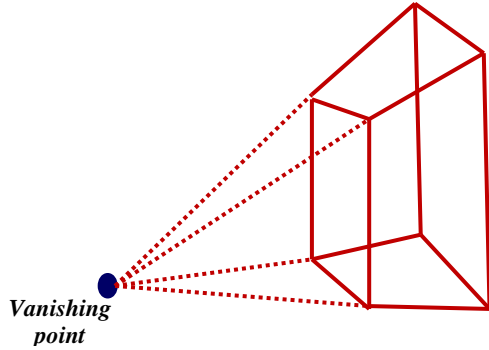

Vanishing point

Fig. 4. Vanishing point

There are many studies on establishing the 3D structure based on the vanishing point. A. Criminisi *et al.* [9] employed texture information to find the vanishing points of a plane. C. Rother *et al.* [10] detected the vanishing points by finding three mutual orthogonal directions. For document images, a method of locating a horizontal vanishing point based on a 2D projection profile was proposed [11]. However, the primary target of this method is the document image which contains multiple text lines of similar orientation. So this method cannot be used for the rectification of text lines in natural scene images.

3.3 Nonlinear Shape Restoration Approach

Y.Y Tang *et al.* [6] defined various models for perspective correction such as the bilinear, biquadratic and bicubic transformation models. In this paper, the bilinear transformation model is used to correct the perspective problem. Therefore, some of terminologies are also employed.

A shape function can represent the shape of an object such as a rectangle or quadrangle. The shape transformation serves to transform the shape function of a perspective distortion to the original shape function.

Restoration based on shape transformation is a successful approach to perspective distortion. Some studies on this topic can be found in [12]. The main idea of restoration based on shape transformation can be described as follows.

- First, the type of shape function of the origin, which is bilinear, biquadratic, or bicubic, is identified.
- Then, the corresponding inverse transformation function, which is used to transform the distorted shape function to the original shape function defined in the previous step, is applied on distorted shape function.

In this paper, we also follow the above process. We assume that the nonlinear shape function is a rectangle. Therefore, the corresponding transformation function is a bilinear transformation.

4. REMOVING NOISE IN THE EDGE IMAGE

Because of the complex background, a lot of noise, which does not constitute the four sides of the rectangle, is produced. This step aims to remove this noise in the edge image by using some morphological operations. First, the edge map is obtained by applying a particular edge detector to a grayscale input image. This edge map is the combination of the useful sides of the rectangle and noise, which are represented by the white pixels. This type of image can be formulated by the following equation.

$$g(x, y) = f(x, y) + \eta(x, y) \qquad (1)$$

Where $g(x, y)$ is the function of the edge map. $f(x, y)$ and $\eta(x, y)$ are functions of the rectangular sides and noise, respectively. $f(x, y)$ is generated from the sides of the rectangular object and $\eta(x, y)$ is generated by the surrounding small objects. This kind of noise is more likely to be of the salt-and-pepper type.

In the morphological approach, there are four important algorithms: dilation, erosion, opening, and closing. Dilation is a kind of white-pixel extension, while erosion is white-pixel reduction. The opening and closing algorithm is a combination of the dilation and erosion algorithms. The structural element is used as the input of the morphological operation. For example, the erosion of image A by a structural element R is denoted by the following equation.

$$I = A \ominus R \qquad (2)$$

The result of the erosion operation is heavily based on the design of the structural elements. In this paper, we design the structural elements for erosion operations to eliminate the noise, whose length of connected edge pixels is less than a particular number. The extension of the erosion algorithm for an image A is formulated as follows.

$$I = (A \ominus R_1) \mid (A \ominus R_2) \mid (A \ominus R_3) \mid (A \ominus R_4) \mid \qquad (3)$$
$$(A \ominus R_5) \mid (A \ominus R_6)$$

Where R_1, R_2, R_3, R_4, R_5, and R_6 are designed as shown in Fig. 5. The size of each structural element is a 5x5 matrix. The gray cells represent the 1's and the white cells represent the 0's of the matrix. The purpose of using such structural elements is to remove noises while preserving lines which could be part of the boundary of rectangular object.

There could be numerous other methods for removing noises, but in this particular application the erosion with structuring elements in Fig. 5 are appropriate. For example, median or Gaussian filtering are not proper in this problem.

The experimental results demonstrate that the proposed method for eliminating noise is effective. Fig. 7(a) shows the edge map including noise, and the result obtained after applying erosion is shown in Fig. 7(b).
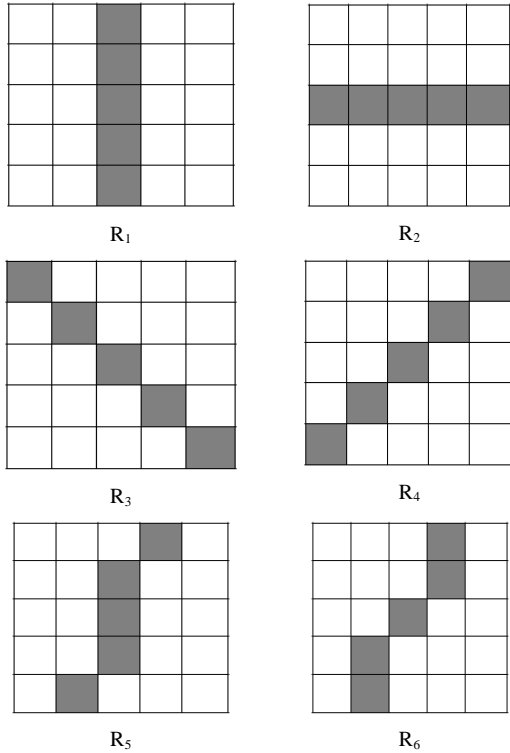
R₁   R₂

R₃   R₄

R₅   R₆

Fig. 5. Some basic structuring elements

5. DETECTING THE LINE CANDIDATES OF A DISTORTED RECTANGULAR OBJECT

(a)   (b)

Fig. 6. (a) color image   (b) grayscale image

In this section, we describe how to detect the line candidates which can be the sides of a rectangular object. First, the input color image is converted into grayscale image as shown in Fig 6. Then, a Sobel edge detector is applied on grayscale image to get edge image in the Fig 7 (a). The noise removal is applied on edge image by a method described in section 4. Finally, Hough transform is applied to detect the line candidates.
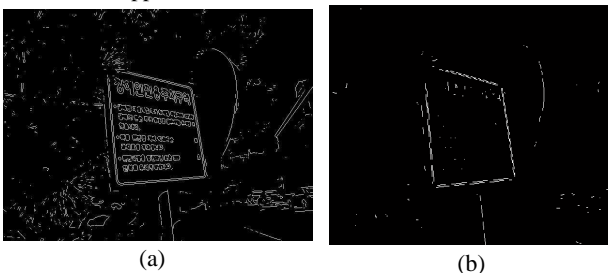
(a)   (b)

Fig. 7.   (a) Edge map   (b) After noise removal

5.1 Hough Transform

The Hough Transform plays an important role in image processing. Generally, the Hough Transform is used to transform an image from Cartesian coordinates to Polar coordinates. Duda and Hart *et al.* [13] analyze the Hough space to detect lines and curves. As shown in Fig. 8(a), a line in Cartesian coordinates is transformed to the Hough space in Fig. 8(b). In the new space, there are many curves which intersect at one point (ρ, θ). ρ represents the normal distance from the origin to the line and θ is the normal angle between the x-direction and the line in the xy coordinates.

In this paper, we use the Hough Transform to detect the four sides of a rectangular object, from which we obtain the four reference points for the bilinear transformation algorithm. To detect the four sides of the perspective rectangular object, first, we detect all of the straight lines existing in the images. Then, we select four sides by analyzing the geometrical relationship among them.
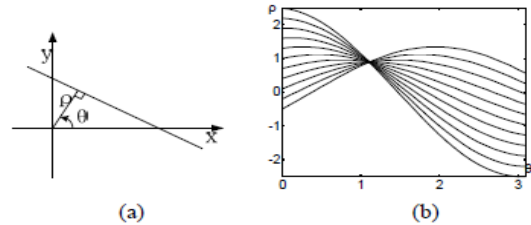
(a)   (b)

Fig. 8. (a) Line in Cartesian coordinates (b) Corresponding Hough space in Polar coordinates

5.2 Detection of Straight Lines

This step is used to detect line segments in the binarized image after removing the noise in section 4. Let us denote $C(\rho, \theta)$ as the number of white pixels satisfying the linear equation $\rho = x * \cos\theta + y * \sin\theta$. We extract all of the peaks where $C(\rho, \theta)$ is larger than a particular threshold, TH_1. We assume that the length of each side of the detected object is larger than or equal to one-third of the width or height of the image. Therefore, we can choose the value of the TH_1 value to be about one-third of the size of the width or height of the image. Fig. 9 shows some segments of detected lines which are marked in red color.
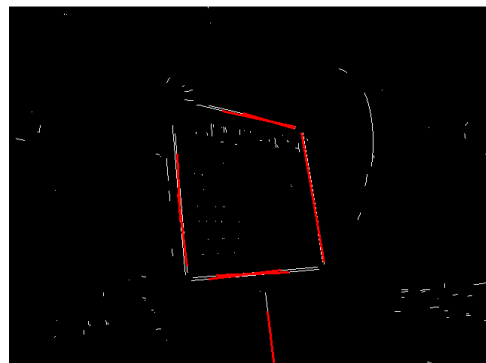
Fig. 9. Line segments detected by Hough Transform

5.3 Detecting Pairs of Line

The lines detected in section 5.2 can include straight lines which are not the sides of a rectangular object. In this section,

we describe a method which allows us to partly remove these lines. Let us consider a rectangle with four vertices $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, $P_3 = (x_3, y_3)$ and $P_4 = (x_4, y_4)$ with centre point M. These points have the following geometrical conditions.

1. The two opposite sides are parallel. For instance, the sides P_1P_2 and P_3P_4 are parallel, as are P_2P_3 and P_4P_1.
2. Side P_1P_2 is opposite to side P_3P_4 via the centre point *M*. Similarly, P_2P_3 is opposite to side P_4P_1 via the centre point *M*.
3. Let us denote d_1 and d_2 as the distances from the centre point *M* to any two opposite sides. Then, the value $(d_1 - d_2)$ is zero.
4. The difference in the angles between any two adjacent sides is 90 degrees.
5. The position of the intersection between any two adjacent sides is in the range of the image size.

Let L_i and L_j be any two lines detected in section 5.2. θ_i and θ_j which are also computed in the previous sub-section are the angles of lines L_i and L_j, respectively. We remove the line L_i if we can't find any line L_j satisfying conditions 1, 2, and 3. However, the rectangular object is not an ideal rectangle which conforms exactly to conditions 1 and 3. Therefore, we must make conditions 1 and 3 more practical by using the following inequality (by adding a thresholding value).

$$| \theta_1 - \theta_2| < TH_2$$
$$| d_1 - d_2| < TH_3 \qquad (4)$$

where θ_1 and θ_2 are the angles of two opposite lines. In our experimental result, we set TH_2 to 25, while TH_3 is based on the width and height of the given image. The sample image with boundary detection is given in Fig. 10.

Fig. 10. Detected boundary

6. RESTORATION BY BILINEAR SHAPE TRANSFORMATION

Once the shape (or boundary) of perspective image is known, bilinear transformation can be used to restore it to normal image as shown in Fig. 11. In this section, we describe how to detect the boundary of perspective image.

By now, we have obtained the candidates of pairs of lines which can be the two opposite sides of a rectangular object. In this section, we describe a way to choose the four sides (two

pairs of lines) of a rectangle. Then, the reference points are inferred from these sides. The selection of the two pairs of lines is based on conditions 4 and 5. Let α_i and α_j be the average angles of any two pairs of lines p_i and p_j, respectively. To realize condition 4 in the perspective rectangular object, we find any two pairs, namely p_i and p_j, satisfying the condition that the difference in their angles is within the range of $90 - TH_4$ to $90 + TH_4$. Fig. 11 shows an example of the detected perspective rectangular object bounded with blue color.

6.1 Bilinear Transformation

A bilinear transformation is one kind of nonlinear transformation. Eq. (5) shows the formula of the bilinear transform from a distorted image to the original one.

$$\begin{bmatrix} X \\ y \end{bmatrix} = \begin{bmatrix} [1 - \varepsilon & \varepsilon] & 0 \\ 0 & [1 - \varepsilon & \varepsilon] \end{bmatrix} \begin{bmatrix} x_{p1} & x_{p4} \\ x_{p2} & x_{p3} \\ x_{y1} & x_{y4} \\ x_{y2} & x_{y3} \end{bmatrix} \begin{bmatrix} 1 - \eta \\ \eta \end{bmatrix} \qquad (5)$$

where (ε, η) is the coordinate representing the normal image, (X, Y) is the coordinate representing the distorted image and P_1, P_2, P_3, and P_4 are the four vertices of the distorted rectangle. Fig. 11 illustrates the relation between the bilinear shape distortion and the original shape.
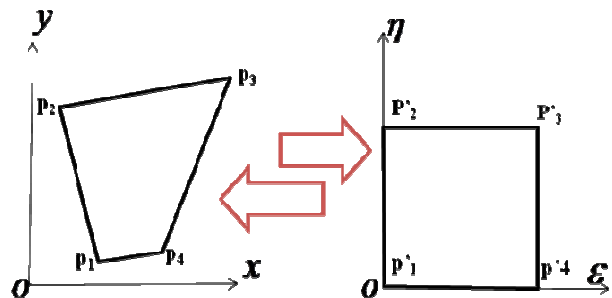

Fig. 11. Some basic structuring elements

7. EXPERIMENTAL RESULTS

We tested our algorithm with input images taken under various camera directions. Fig. 12 shows the results of the restoration operation, which demonstrate the performance of the proposed methods. The comparison with existing methods may be viable because either they are mostly proposed for document images[2,3] or they detect text first and then rectification is performed to the shape of the those images[1].

(a)   (b)

Fig. 12. (a) Original image with bounding box (b) The
restored rectangular object

8. CONCLUSIONS

In this paper, we proposed an algorithm to detect and restore perspective rectangular objects. The boundary of the text rectangle has been detected by Hough transform and the preprocessing algorithm for the removal of noises is given with morphological operations. The proposed algorithm has been applied to different sample images and proved to be effective in such images. The proposed method is based on assumptions that the position of the object is in the centre of the image and the size of the object is within a specific range. For future research,

the algorithm must be improved to work on a wider range of input images.

REFERENCES

[1] G.K. Myers, R.C. Bolles, Quang-Tuan Luong, James A.Herson and Hrishikesh B. Aradhye, "Rectification and Recognition of Text in 3-D Scenes," International Journal on Document Analysis and Recognition, vol.7, no.2, 2004, pp. 147-158.

[2] C.R. Dance, "Perspective Estimation for Document Images," proceedings of the SPIE Conference on Document Recognition and Retrieval IX, 2002, pp. 244-254.

[3] S.J. Lu, B.M. Chen, C.C. Ko, "Perspective Rectification of Document Images Using Fuzzy Set and Morphological Operations," Imnage and Vision Computing, vol.23, 2005, pp.541-553.

[4] Lin Liu, "Slant Correction of Vehicle License Plate Image," International Conference on Computer Science and Software Engineering, vol.3617, 2008, pp. 237-244.

[5] D.H. Han, H.K. Sung, K.T. Park, Y.H. Cho and H.M. Choi, "Neural network approach to the nonlinear shape restorations", IEEE International Conference on Systems, Man, and Cybernetics, vol.1, 1996, pp.504-509.

[6] Y.Y. Tang and C.Y. Suen, "Image transformation approach to nonlinear shape restoration," IEEE Trans. Syst. Man Cybernet, Jan. 1993, pp. 155–171.

[7] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall, 1999.

[8] Huaigu Cao; Xiaoqing Ding; Changsong Liu; , "A cylindrical surface model to rectify the bound document image," Ninth IEEE International Conference on Computer Vision, vol.1, 2003, pp.228-233.

[9] A. Criminisi, A. Zisserman, "Shape from texture: homogeneity revisited," In Proc.11th British Machine Vision Conference, 2000, pp. 82–91.

[10] C. Rother, "A new approach for vanishing point detection in architectural environments," In Proc.11th British Machine Vision Conference, 2000, pp. 382–391.

[11] Paul Clark, Majid Mirmehdi, "Rectifying perspective views of text in 3D scenes using vanishing points," Pattern Recognition, vol.36, no.11, Nov. 2003, pp.2673-2686.

[12] Li, T.D. Bui, C.Y. Suen, Y.Y. Tang and Q.L. Gu, "Splitting-integrating method for normalizing images by

inverse transformations," IEEE Trans on Pattern Anal. Mach. Intell, vol.14, no.6, 1992, pp. 678–686.

[13] R. Duda and P. Hart, "Use of the Hough transform to detect lines and curves in pictures," Communications of the ACM, 1972, vol.15, No.1, pp.11–15.

Huy Phat Le

He received the B.S degree in Information Technology from University of Science, Ho Chi Minh City, Vietnam in 2006. He is currently a MS student in Computer Science department of Chonnam National University, Korea. His interesting researches are Image processing, especially document analysis, and computer vision.

Kavitha Madhubalan

She received the B.S degree in Electronic and Computer Engineering from AVINASHILINGAM UNIVERSITY FOR WOMEN, in INDIA, 2008. She is currently a MS student in Computer Science department of Chonnam National University, Korea. Her interests include issues in image processing and computer vision.

Gueesang Lee

He received the B.S. degree in Electrical Engineering and the M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received the Ph.D. degree in Computer Science from Pennsylvania State University in 1991. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video coding.