

A NOVEL SCHEME FOR IMAGE ROTATION FOR DOCUMENT PROCESSING

Kaushik Mahata and A. G. Ramakrishnan, Senior Member, IEEE

Dept. of Electrical Engineering, Indian Institute of Science, Bangalore 560012, INDIA

ABSTRACT

Rotation of the binary image of a document page for correcting the skew in the case of OCR or signature verification systems entails disfigurement of the shape of the characters, which brings down the performance of the pattern classifier. An elegant algorithm is presented in this paper for distortion-free rotation. Multirate signal processing principles are applied to solve this problem, which was hitherto handled using only heuristics. The resolution of the image is increased during rotation. The rotated image is decimated and thresholded after low pass filtering. Polyphase implementation of the filter is performed for efficient computation. The results show that it solves the problem of mutilations and occasional breaks obtained in simple rotation schemes and further, performs better than the scheme involving interpolation in the original or the rotated domains. Consequent on the better retention of their original shape, the recognition accuracy of originally skewed documents improves by more than 5%.

1. INTRODUCTION

In OCR and signature verification systems, the binary image is to be rotated for correcting the skew or the orientation of the principal axis of signature. This involves rotation of the image in the discrete spatial plane through an angle $-\theta$ with respect to the horizontal line, if the detected angle (of skew or the axis of signature) is θ . Conventionally, this is achieved by transforming the co-ordinates of each foreground pixel by the following transformation matrix :

$$Q = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

If H is the height of the image and W , the width, we select the origin at $(H/2, W/2)$. For every foreground pixel at (x, y) , we find out the new co-ordinates (x_0, y_0) in the rotated image as follows:

$$\mathbf{u}_0 = Q\mathbf{u}, \quad (1)$$

where, $\mathbf{u}_0 = (x_0, y_0)^t$ and $\mathbf{u} = (x, y)^t$.

Practical implementation of above has a serious drawback. This is because of the rectangular tiling and the

finite resolution of the image. In the above equation, x and y take only integral values. However, the computed values of x_0 and y_0 are real numbers. The rounding operation, needed to convert these non-integral pair of coordinates into integral values, introduces a shift in their position. The effect is most noticeable while rotating straight edges (of which characters are made in most languages). This is so because, if a square with its sides parallel to the axes of co-ordinates is rotated, it will no longer have its sides parallel to the axes. But we are forced to keep its sides parallel to the axes according to our predefined tiling.

These two type of errors are sometimes severe enough to distort the characters in such a way, that the recognition performance is degraded. This problem is extremely prominent in case the character is thin. Figure 1(b) shows the rotated version of the skewed document in Fig. 1(a). The distortions in the former are obvious. In the case of thick characters, one gets voids (white dots amidst black ones), which entail distortion in the morphology of the character after skeletonisation. Fig. 2 shows examples of this phenomenon. A scheme which can partially solve this problem is to use bilinear interpolation. Here, we go from the rotated image space to the original space to obtain the gray level of each pixel in the rotated image. We map the coordinates of each pixel to the original space and obtain the intensity at the point with nonintegral pair of coordinates using the values of the four integral nearest neighbours. The final image is obtained by appropriate thresholding. However, for certain angles, this scheme still introduces distortion, as seen from Fig. 1(c).

2. METHODS

We propose a rotation algorithm that reduces the amount of distortion. The key step is to increase the resolution of tiling in order to reduce the rounding error. If we increase the resolution l times, the error due to rounding is reduced by a factor of $1/l$. At the same time, (1) gets modified as,

$$\mathbf{u}_1 = L\mathbf{u}_0 \quad (2)$$

where $L = \text{diag}(l, l)$

The image in the higher resolution ($l = 4$) corresponding to the image in Fig. 3 (a) is shown in Fig. 3 (e). It is to be noted that we have been able to preserve the shape of the original character. At the same time, we have expanded the image. According to the concepts of multirate signal processing, the expanded image $I_e(x, y)$ is required to be passed through a low pass filter (LPF) and

then decimated [1] in order to obtain the image with the original resolution. The pass band of the LPF is from $-\pi/l$ to $+\pi/l$ in digital frequency domain [1], and the decimation factor is 1.

Since the signal does not impose any restrictions on us from the point of view of causality, we are free to use ideal filters [1], expressed in the frequency domain by $G(\exp(j\omega)) = 1$ for $|\omega| \leq \pi/l$ (3)
 0 for $\pi/l < |\omega| \leq \pi$

It is sufficient to define it in $-\pi \leq \omega \leq \pi$. In the time domain,

$$g[n] = \sin((\pi/l)n)/\pi n, \quad -\infty < n < +\infty \quad (4)$$

being the inverse Fourier transform of (3).

Decimation results in an attenuation of the signal by factor of l , and hence we scale the coefficients of the filter by a factor of $1/l$. $g[n]$, being an infinite sequence, needs to be truncated for practical implementation. For all practical purposes, it suffices to consider $g[n]$ for $n \in [-l, l]$, i.e., only the central lobe of $lg[n]$.

Computation complexity of the filtering procedure is $O(l^2)$. However, since we are going to decimate the filtered signal by a factor of l , it is required to compute only one out of each l consecutive samples of the filter output, and this helps to reduce the computation. For two-dimensional images, this entails a gain by a further factor of l^2 .

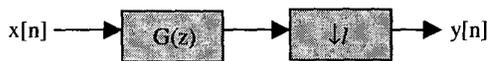


Fig. 4. LPF followed by decimator.

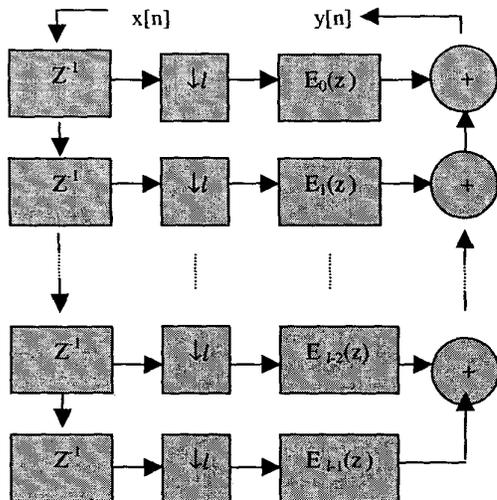


Fig. 5. Polyphase decomposition of the LPF.

The basic building block is shown in Fig.4 as a block diagram. This is computationally better realized by polyphase decomposition of the filter, $G(z)$. We decompose $G(z)$, the z-transform [2] of $g[n]$, into its polyphase components as follows.

$$G(z) = E_0(z^l) + z^{-1} E_1(z^l) + \dots + z^{-(l-1)} E_{l-1}(z^l),$$

where, $E_0(z), \dots, E_{l-1}(z)$ are the polyphase components of $G(z)$.

The polyphase implementation of the filter is illustrated in Fig. 5, together with decimation [2]. It is to be noted that we still perform the same amount of computation in obtaining each filtered sample. However, due to the decimation before filtering, the number of samples calculated is only $1/l$ times the previous case. The same operation is performed on each row of the interpolated image, and subsequently, on each column. This is shown as a block diagram in Fig. 6.

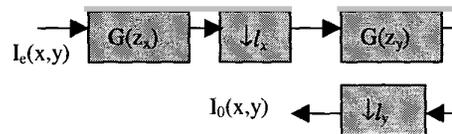


Fig. 6. 2-D filtering achieved through 1-D steps.

The filtered and decimated outputs are, in general, real numbers. It is required to threshold them properly to obtain a binary image $I_0(x,y)$.

3. RESULTS

Fig. 1 (d) shows the resulting image after rotation by the proposed scheme. The absence of serious disfigurement is obvious. In actual studies involving Tamil and Kannada scripts, an improvement of a minimum of 5% is obtained in recognition accuracy with the above scheme, compared to the performance of traditional schemes.

4. REFERENCES

- [1] V. Oppenheim and R. N. Schaffer, *Discrete Time Signal Processing*. Englewood Cliffs, NJ; Prentice Hall, 1989.
- [2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ; Prentice Hall, 1993.

NGOs are call
from various go'
and law enforce
what they say a
dowry **deaths ar**
Bangalore.

About 5 per ce

Fig. 1(a) Original Document
with a skew angle of 1°.

NGOs are call
from various go'
and law enforce
what they say a
dowry **deaths ar**
Bangalore.

About 5 per ce

Fig. 1. (c) Document Rotated
using bilinear interpolation.

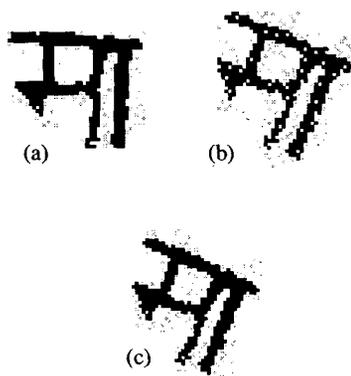


Fig. 2. a) Original Character.
b) Character rotated by conventional
scheme. c) Character rotated by
proposed scheme.

NGOs are call
from various go'
and law enforce
what they say a
dowry **deaths ar**
Bangalore.

A b u t 5 per ce

Fig. 1(b) Document Rotated
by conventional scheme.

NGOs are call
from various go'
and law enforce
what they say a
dowry **deaths ar**
Bangalore.

A b u t 5 per ce

Fig. 1. (d) Document rotated by
the proposed scheme.

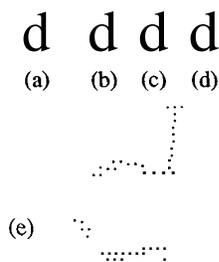


Fig. 3. a) Original character.
b) Character rotated by conventional
scheme through 5°.
c) Character rotated by proposed
scheme. d) Character rotated by
bilinear interpolation e) Expanded and
rotated Character.