Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 6, Number 2 (2013) pp. 111-119

https://dx.doi.org/10.37622/ACST/6.2.2013.111-119

Recognition of Handwritten Characters Using Cosine Measure

Sumit Saha [1] and Tanmoy Som [2]

Dept. of Mathematics, Assam University, Silchar -788011, INDIA[1]

Dept. of Applied Mathematics, Institute of Technology, Banaras Hindu University, Varanasi -221005, INDIA[2]

Abstract:

Most of the pattern recognition models are based on finding the statistical or geometrical properties (Hong et. al.,1998; Prasad et. al, 1993; Rekhter and Li,1995; Walters,2006) of substructures in the data. The two key concepts of describing geometry are angle and distances. In this paper we have tried to recognize handwritten characters depending on the geometrical angle. To find the geometrical angle we have taken the *cosine measure* as the tool.

Keywords: Cosine measure, Normalized vector, geometrical angle.

Introduction:

The cosine measure of two vectors has been found to be very effective in the comparison of two documents. The cosine measure of two vectors V_i and V_j can be computed as:

$$\cos(V_i, V_j) = \frac{\sum_{k=1}^{N} V_{i,k} * V_{j,k}}{\sqrt{\sum_{k=1}^{N} V_{i,k}^{\ 2} * \sum_{k=1}^{N} V_{j,k}^{\ 2}}} \quad \ldots\ldots\ldots \quad (1)$$

This is essentially the cosine of the angle between the two vectors. The *cosine measure* is nothing but the inner product of two vectors, after both V_i and V_j have been normalized to have unit length. As a result, the cosine measure reflects the relative distribution of the terms of the vectors and this measure has been found to be

very effective in the recognition of handwritten characters.

After normalizing the two vectors V_i and V_j the denominator of the right hand side of equation (1) becomes 1(one) and the equation takes the form

$$\cos(V_i, V_j) = \sum_{k=1}^{N} V_{i,k} * V_{j,k} \qquad \text{.....(2)}$$

This equation gives the smaller angle between the two vectors Vi and Vj. hence the value of $\cos(V_i, V_j)$ lies between 0 and 1. For almost similar vectors Vi and Vj the value of $\cos(V_i, V_j)$ is close to 1 and vice versa.

Steps used to calculate the angle between the characters

In this paper the vectors are the patterns of different handwritten characters. The patterns are 10 × 10 matrices and the method used to form the patterns is described below. In the preprocessing stage different characters are fitted within a rectangular frame that touches the characters at its left, right, top and bottom. This rectangular frame is then resized into a 100 × 100 square box. This 100 × 100 square box is divided into 10 × 10 grid structure. The number 1 (one) is assigned to the (i,j)th grid if the (i,j)th grid contains at least one black pixel and the number 0 (zero) is assigned to the (i,j)th grid if the (i,j)th grid contains no black pixel . These 10 × 10 matrices so formed are the initial pattern vectors for individual known character images.

Step1.

Pattern vectors of all the characters (in this case there are 26 characters) are normalized to unit length by dividing each element of the pattern vector by the length of that pattern. These matrices of unit length are nothing but the unit vectors in different direction.

For example of the pattern of the character ‘A’ when normalized to unit length is given below:

0 0 0 0 1 1 0 0 0 0
0 0 0 0 1 1 0 0 0 0
0 0 0 0 1 1 1 0 0 0
0 0 1 1 1 1 1 1 0 0
0 1 1 1 1 0 0 1 0 0
1 1 1 1 1 1 1 1 1 0
1 1 1 0 0 0 1 1 1 0
1 1 0 0 0 0 1 1 1 1
1 1 0 0 0 0 0 1 1 1
1 1 0 0 0 0 0 0 1 1

Pattern of ‘A’

Figure 1(a)

0	0	0	0	0.1443	0.1443	0	0	0	0
0	0	0	0	0.1443	0.1443	0	0	0	0
0	0	0	0	0.1443	0.1443	0.1443	0	0	0
0	0	0.1443	0.1443	0.1443	0.1443	0.1443	0.1443	0	0
0	0.1443	0.1443	0.1443	0.1443	0	0	0.1443	0	0
0.1443	0.1443	0.1443	0.1443	0.1443	0.1443	0.1443	0.1443	0.1443	0
0.1443	0.1443	0.1443	0	0	0	0.1443	0.1443	0.1443	0
0.1443	0.1443	0	0	0	0	0.1443	0.1443	0.1443	0.1443
0.1443	0.1443	0	0	0	0	0	0.1443	0.1443	0.1443
0.1443	0.1443	0	0	0	0	0	0	0.1443	0.1443

Pattern of 'A' when normalized to unit length

Figure 1(b)

In the above example the length of the pattern Vector 'A' is √50=7.071and this pattern is normalized to unit length by dividing each entry of the 10 × 10 pattern Vector 'A' by the length √50.

Step2. The pattern of the unknown character is then normalized to unit length.

Step3. The cosine of the angle between the directions of the unknown pattern and the known patterns are calculated with the help of the equation (2)
If the cosine of the angles between the two directions (*angle between the known pattern vector and the unknown pattern vector)* is close to 1 then the angle between those directions is small and vice versa.

When an unknown pattern is introduced to the system, the angular distance (i,e.cosine measure) between the unknown pattern and the known patterns are calculated using equation (2).

Thus we can say that the pair of the known and unknown pattern vectors which has the greatest cosine measure has the smallest angle between them i,e. they have the greatest degree of similarity between them, and the corresponding known character is considered as the recognized character.

Program Code used for recognition of unknown character by using Cosine measure (Using MATLAB 6.5)

```
A=imread(FILE NAME);       %read the image file
[row col]=size(A);       % row and column
          %numbers of the image
ABW=~im2bw(A);      %image A is converted
%into 0's and 1's
S=sum(ABW);
%vertical projection is
%taken to find the
%vhorizontal word
```

```
%boundary
chk=0;
%chk is used to change
%the selection mode
for i=1:col
%Finds the lift and right
%boundary of the %word
if S(1,i)~=0 && chk==0
left=i;
chk=1;
end
if S(1,i)==0 && chk==1
right=i;
chk=2;
end
end
RA=rot90(ABW);       %rotate the image to take
% the horizontal
%projection
SRA=sum(RA);
%horizontal projection is
%taken to find the left and
% right boundary of the
%word
chk=0;
for i=1: row     %Finds the upper and %lower boundary of the %word
if SRA(1,i)~=0 && chk==0
top=i;
chk=1;
end

if SRA(1,i)==0 && chk==1
bottom=i;
chk=2;
end
end

I=imcrop(ABW,[left top right-left bottom-
top]);
                        %capture the image within
%arectangular box
hgt=bottom-top;
bred=right-left;
hinc=(hgt/10);
binc=(bred/10);
```

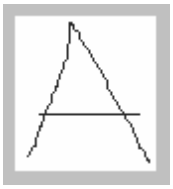
```
k=1;
l=1;
for i=1:10
for j=1:10
J=imcrop(I,[k l binc-1 hinc-1]);
if sum(J(:))==0
patt(i,j)=0;
else
patt(i,j)=1;
end
k=k+binc;
end
k=1;
l=l+hinc;
end
Test1=patt;
mod=1/(sqrt(sum(Test1(:))));
Test=mod*Test1;
%Test is the normalized
%unknown pattern
%vector of unit length
k=1;
for p=1:26
I=imcrop(unit,[k 1 9 9]);
%unit is a 10 ×260 matrix
%which stores the
%patterns of 26 character %and I capture the
% pattern of one character
%in each iteration
mod=1/sqrt(sum(I(:)));
J=mod*I;
% J is the normalized %pattern of I having  %unit length
S=0;
for i=1:10
for j=1:10
S=S+J(i,j)*Test(i,j);
% S is the dot product of
%two unit vectors Test
%and J
end
end

costheta(1,p)=S;
% costheta stores the
%cosine of different pair
```

```
%of vectors
k=k+10;
end
```

Few results are given below:
Results of recognition using cosine measure:

Result 1:



Unknown character

Figure 2(a)

A	B	C	D	E	F	G
0.6030	0.4458	0.2824	0.4222	0.4681	0.4129	0.3518
H	I	J	K	L	M	N
0.5311	0.2824	0.2227	0.4104	0.3030	0.4304	0.3790
O	P	Q	R	S	T	U
0.3633	0.3223	0.5303	0.5303	0.4523	0.2259	0.1846
V	W	X	Y	Z		
0.2388	0.1777	0.5264	0.3337	0.4458		

Figure 2(b)

Angular Distances(Cosine measure)between the unknown and the known patterns
Maximum value of the cosine measure corresponds to the character ‘A’
The character is recognized as ‘A’

Result 2:

Unknown character

Figure 3(a)

A	B	C	D	E	F	G
0.5567	0.6036	0.4171	0.5717	**0.7316**	0.6550	0.5567

H	I	J	K	L	M	N
0.5943	0.6257	0.5787	0.5349	0.4228	0.4710	0.5055

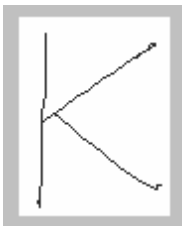
O	P	Q	R	S	T	U
0.4472	0.4409	0.5298	0.5487	0.6392	0.4867	0.2273

V	W	X	Y	Z
0.3675	0.3959	0.5939	0.5055	0.5304

Figure 3(b)

Angular Distances(Cosine measure)between the unknown and the known patterns
Maximum value of the cosine measure corresponds to the character ‘E’
The character is recognized as ‘E’

Result 3:



Unknown character

Figure 4(a)

A	B	C	D	E	F
0.4951	0.3294	0.2782	0.3536	0.3952	0.4339

G	H	I	J	K	L
0.3713	0.4281	0.3060	0.1828	**0.6897**	0.4180

M	N	O	P	Q	R
0.5418	0.5134	0.3579	0.2911	0.4089	0.4997

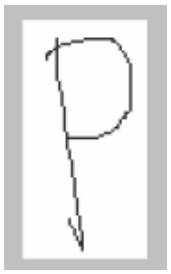
S	T	U	V	W	X
0.3218	0.3617	0.2729	0.4412	0.4503	0.4970

Y	Z
0.3793	0.3952

Figure 4(b)

Angular Distances(Cosine measure)between the unknown and the known patterns
Maximum value of the cosine measure corresponds to the character ‘K’
The character is recognized as ‘K’

Result 4 :



Unknown character

Figure 5(a)

A	B	C	D	E	F
0.2228	0.5270	0.2782	0.5199	0.4831	0.5423
G	H	I	J	K	L
0.3466	0.3805	0.3895	0.2194	0.3330	0.2985
M	N	O	**P**	Q	R
0.3769	0.4201	0.3068	**0.6351**	0.3180	0.3862
S	T	U	V	W	X
0.3466	0.5842	0.3941	0.5000	0.5503	0.4105
Y	Z				
0.4804	0.4392				

Figure 5(b)

Angular Distances(Cosine measure)between the unknown and the known patterns
Maximum value of the cosine measure corresponds to the character ‘P’
The character is recognized as ‘P’

Conclusion:

Although we have given only 4 results, the above method is tested with a huge number of handwritten characters of different variety. It has been found that the above method is very effective and the it gives a very high degree of recognition rate.

References:

[1] C.Y. Suen, C. Nadal, R. Legault, T.A. Mai, and L. Lam "Computer recognition of unconstrained handwritten numerals". *Special Issue of Proc. of the IEEE,* 7(80):1162–1180, 1992.

[2] Devijver P and Kittler J ,"Pattern Recognition : A Statistical approach" (Englewood Cliffs, NJ: Prentice-Hall) , 1982

[3] Fukunga K., Statistical Pattern Recognition (San Diego ,CA : Academy), 1991

[4] Hong C., Loudon G., Wu Y. and Zitserman R., Segmentation and recognition of continuous handwriting Chinese text," Int. J. Pattern Recognition and Artificial Intelligence 12, 2 (1998) 223-232.

[5] I. Guyon, R. Haralick, J. Hull, and I. Phillips *"Handbook of Character Recognition and Document Image"*. Database and benchmarking. In H. Bunke and P. Wand, editors, *Analysis*,chapter 30, pages 779–799. World Scientific, 1997.

[6] Nagendra Prasad M.V., Gupta A., and Feliberti V.,A new algorithm for correcting slant in handwritten numerals', in Discussion paper 215- 92, International Financial Services Research Center, Sloan school of management,1993.

[7] Rekhter Y. and Li T., "A Border Gateway Protocol 4 (BGP 4)." Internet Engineering Task Force: RFC 1771, March 1995.

[8] Sumit Saha and T.Som "Handwritten character recognition by using Neural-network and Euclidean distance metric" International Journal of Computer Science and Intelligent Computing Vol. 2, No. 1, November 2010

[9] T.Som and Sumit Saha " Handwritten Character Recognition Using Fuzzy Membership Function" International Journal of Emerging Technologies in Sciences and Engineering, Vol.5, No.2, Dec 2011

[10] Walters A., "Mitigating attacks against adaptation mechanisms in overlay networks," Master's thesis, Purdue University, May 2006.