Character Recognition based on DTW–Radon

Santosh K.C.

INRIA Nancy-Grand Est, LORIA - Campus Scientifique
BP 239 - 54506 Vandoeuvre-lés-Nancy Cedex, France
Email: Santosh.KC@inria.fr

Abstract—The paper presents a method for isolated off-line character recognition using radon features. The key characteristic of the method is to use DTW algorithm to match corresponding pairs of radon histograms at every projecting angle. Thanks to DTW, it avoids compressing feature matrix into a single vector which may miss information. Comparison has been made with the state–of–the–art of shape descriptors over several different character as well as numeral datasets from different scripts.

Keywords-Character Recognition; Radon Transform; DTW.

I. INTRODUCTION

With the advent of handwriting recognition technology since a few decades [20], [2], applications are challenging. For example, OCR is becoming an integral part of document scanners, and is used in many applications such as postal processing, script recognition, banking, security (signature verification, for instance) and language identification. In handwriting recognition, feature selection has been an important issue [28]. Both structural and statistical features as well as their combination have been widely used [13], [10]. Structural features tend to vary since characters' shapes vary widely. As a consequence, local structural properties like intersection of lines, number of holes, concave arcs, end points and junctions change time to time. In those situations, statistical approaches [15] tolerate to a few extent, with no substantial change in the global signatures. Therefore, the present work is inspired to use statistical feature to represent off-line isolated characters.

Owing to large variation in shapes and writing styles, OCR systems (for unconstrained scripts in particular) do not provide optimal accuracy. Despite those circumstances, it is desirable to develop shape representation such that it can be used for machine recognition [23], [7]. To represent shape of the character, contour-based shape descriptors have been widely used in comparison to region-based ones. Fourier processing of the image is an example of using contour information [12], [14]. Besides, curvature approaches have been used [18], [16] where shape is described in the scale space by the maximum number of curvatures. Further, the *Shape Context* (SC) [3], [25] is robust to small perturbations while it does not guarantee scale-invariance. On the other side, common methods on region-based shape descriptors are based on moment theory [9] including *geometric*, *Legendre*, *Zernike* [17], and *Pseudo-Zernike* moments. Contour-based descriptors on the whole, are appropriate for silhouette shapes since they cannot capture the interior content as well as disconnected shapes or shapes with holes. To represent internal structure of the shape, unlike in *Fourier Descriptors* (FD), Zhang and Lu [26] have proposed a region-based *Generic Fourier Descriptor* (GFD) that avoids the problem of rotation in the Fourier spectra. Recently, use of polar shape descriptors has been mentioned in [11]. Overall, from almost all global signal based descriptors [27], the use of normalisation in order to satisfy common geometric invariance properties introduces errors as well as they are sensitive to noise, eventually affecting the whole recognition process.

In document analysis and recognition (character recognition in particular), a single and integrated open source tool-set will be a global choice since it can be re-used as plug and play [14] as a generic model. Another common implementation problem is the inability to assume the shape distribution of the patterns in the feature space. Consequently, non-parametric methods are much more practical. Moreover, those methods in general, use linear functions to describe classifiers. Besides, feature selection must be sufficiently enriched with important information. To represent the pattern, global shape representation is a premier choice due to its simplicity as well as it sometimes does not require extra pre-processing and segmentation process as in local pattern representation. To accomplish recognition, matching is another concern. Feature selection corresponds to the matching techniques. For instance, fixed size pattern representation as in global signal based shape descriptors [3], [26], [17], [24], provide immediate matching while DTW for instance, has been popularly used for non-linear sequences having potentially different lengths.

In those respects, the paper presents a simple idea to represent isolated character via radon transform [8] and recognition is made with the help of DTW [19]. Thanks to DTW, it avoids compressing pattern representation into a single vector – as in [24], for instance – which may miss information. The concept of the work is originally derived for off-line signature verification [6], [21]. In this paper, the method has been tested over several different numeral and character datasets from different scripts such as Roman,

Devanagari, Bangla, Oriaya and Japanese–katakana.

The remaining of the paper is organised as follows. The proposed method is explained in Section II, which mainly includes character representation and matching. Section III provides a series of tests. In Section IV, results are analysed thoroughly. The paper is concluded in Section V.

II. METHOD

A. Pre-processing

The preliminary task is to do pre-processing since characters tend to be highly degraded as they are taken from newspaper, postal cards *etc.* under varying different lighting conditions, for instance. It mainly considers stroke synthesis, thresholding, gray-scale to binary conversion; noise removal, foreground textual information extraction by removing background [1]. However, this introduces many ad-hoc techniques. In this work, isolated character images are simply converted to binary. While converting, an average *grayscale* pixel intensity values is used to make a threshold. Fig. 1 shows a few examples of it. In this illustration, binary conversion is followed by contour detection and thinning.

(a) straight-forward conversion

original sample
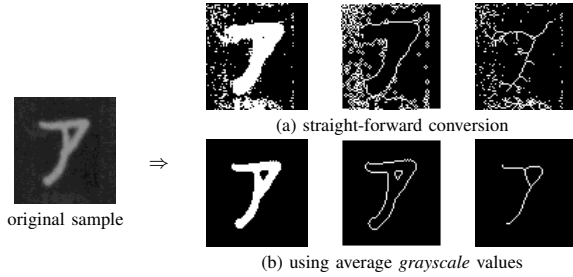
(b) using average *grayscale* values

Figure 1. An example of binarisation.

B. Feature Selection

Besides pre-processing, recognition system is based on how characters are represented [28]. In this work, radon transform is used to represent patterns [8].

As shown in Fig. 2, a collection of projections of the pattern at different angles refers to radon transform [8]. In other words, radon transform for any pattern $f(x, y)$ and for a given set of angles can be thought of as computing the projection of all non-zero points. The resulting projection is the sum of the non-zero points for any pattern $f(x, y)$ in each direction, which eventually form a matrix. Therefore the integral of f over a line $L(\rho, \theta)$ defined by $\rho = x \cos \theta + y \sin \theta$ can formally be expressed as,

$$R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy$$

where $\delta(.)$ is the Dirac delta function, $\delta(x) = 1$, if $x = 0$ and 0 otherwise. Also, $\theta \in [0, \pi[$ and $\rho \in]-\infty, \infty[$. For radon transform, L_i be in normal form (ρ_i, θ_i). Following Fig. 2 (c), for all θ_i, the Radon transform now can be described

as the length of intersections of all lines L_i. Note that the range of ρ *i.e.*, $-\rho_{min} < \rho \le \rho_{max}$ is entirely based on the size of pattern.
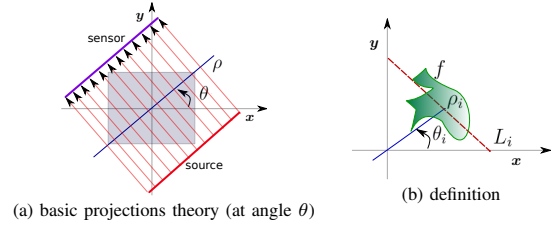

(a) basic projections theory (at angle θ)   (b) definition

Figure 2. Radon transform.

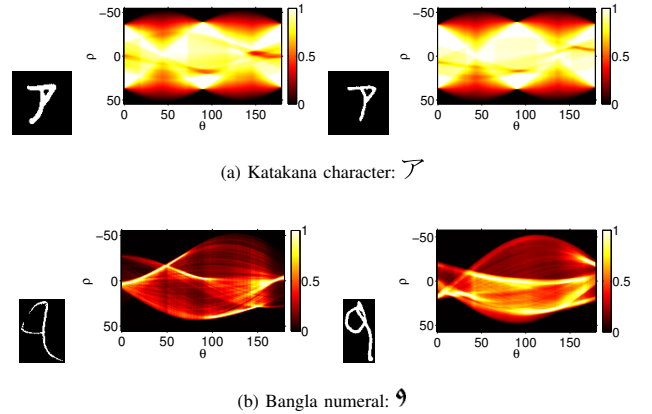(a) Katakana character: ア

(b) Bangla numeral: ৭

Figure 3. Radon features for corresponding samples.

Radon transform itself does not satisfy invariance properties. In this work, it is assumed that the characters are vertically aligned. While, translation and scaling are considered. In case of translation, image centroid (x_c, y_c) is used such that translation vector is $\vec{u} = (x_c, y_c)$: $R(\rho - x_c \cos \theta - y_c \sin \theta, \theta)$. Therefore, translation of f results in the shift of its transform in ρ by a distance equal to the projection of translation vector of the line L (see Fig. 2 (b)). For scaling, features are normalised into $[0, 1]$ at every projecting angle.

C. Feature Matching

Given a query pattern $\mathcal{P}^q$ and database $\mathcal{P}^d$, matching has been done between corresponding histograms of $R^q(\rho, \theta)$ and $R^d(\rho, \theta)$. In radon matrix, for each projecting angle θ_i, column (vertical line in Fig. 4) refers to as histogram of radon features. Let us say histograms $\{\mathcal{H}_{\theta_i}\}_{i=0,...,\Theta-1}$. Radon transform generates different ρ size according to the image size. Keeping images into their original sizes, DTW algorithm [19] is proposed to match corresponding radon histograms.

Let us consider two column vector sequences from $R(\rho, \theta)$, representing histograms $\mathcal{H}^q$ and $\mathcal{H}^d$ of length K and L, respectively $\mathcal{H}^q = \{h^q_k\}_{k=1,...,K}$ and $\mathcal{H}^d = \{h^d_l\}_{l=1,...,L}$. At first, a matrix $\mathcal{M}$ of size $K \times L$ is constructed. Then for each element in matrix $\mathcal{M}$, local distance
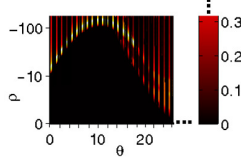
Figure 4. Radon histogram at every projecting angle θ_i.

metric $\delta(k,l)$ between the events e_k and e_l is computed. $\delta(k,l)$ can be expressed as, $\delta(k,l) = (e_k - e_l)^2$, where $e_k = h_k^q$ and $e_l = h_l^d$. Let $D(k,l)$ be the global distance up to (k,l),

$$D(k,l) = \delta(k,l) + \min \begin{bmatrix} D(k-1,l-1) \\ D(k-1,l) \\ D(k,l-1) \end{bmatrix}$$

with an initial condition $D(1,1) = \delta(1,1)$ such that it allows warping path going diagonally from starting node $(1,1)$ to end (K,L). The main aim is to find the path for which the least cost is associated. The warping path therefore provides the difference cost between the compared sequences. Formally, the warping path is, $\mathcal{W} = \{w_t\}_{t=1\ldots T}$ where $max(k,l) \leq T < k + l - 1$ and t^{th} element of $\mathcal{W}$ is $w(k,l)_t \in [1:K] \times [1:L]$ for $t \in [1:T]$. The optimised warping path $\mathcal{W}$ satisfies the following three conditions.

c1. boundary condition: $w_1 = (1,1)$ and $w_T = (K,L)$
c2. monotonicity condition: $k_1 \leq k_2 \leq \cdots \leq k_K$ and $l_1 \leq l_2 \leq \cdots \leq l_L$
c3. continuity condition: $w_{t+1} - w_t \in \{(1,1)(0,1),(1,0)\}$ for $t \in [1:T-1]$

Now the global distance between $\mathcal{H}_q$ and $\mathcal{H}_d$ can be defined as,

$$\Delta\left(\mathcal{H}^q, \mathcal{H}^d\right) = \frac{D(K,L)}{T},$$

where T is the number of discrete warping steps along diagonal DTW-matrix. Fig. 5 shows an example of it.
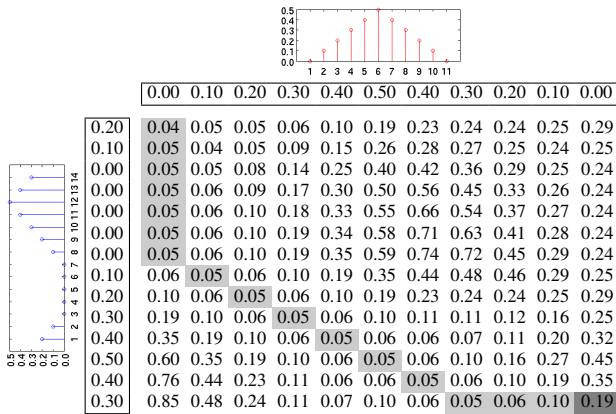

Figure 5. Computing DTW distance between two non-linear sequences.

Aggregating distances between histograms in all corresponding projecting angles θ_i between $\mathcal{P}^q$ and $\mathcal{P}^d$ yields a global pattern-matching score,

$$Dist(\mathcal{P}^q, \mathcal{P}^d) = \sum_{i=0}^{\Theta-1} \Delta(\mathcal{H}_{\theta_i}^q, \mathcal{H}_{\theta_i}^d).$$

Scores are normalised in $[0,1]$ by, $\frac{Dist(.) - Dist^{min\cdot}(.)}{Dist^{max\cdot}(.) - Dist^{min\cdot}(.)}$.

III. EXPERIMENTS

A. Datasets

Several different datasets from different scripts[1] are tested. They are categorised into two groups: Numeral and Character. For Numeral category, Roman, Bangla, Devanagari and Oriya [5], [4] are used, while Roman, Bangla as well as Katakana (Japanese) character are used in character category.

Figure 6. 2 numeral samples from 4 different known scripts for 10 $(0-9)$ classes.

B. Comparison World – Benchmarking Methods

A set of major global signal based shape descriptors such as $\mathcal{R}-$signature [24], GFD [26], SC [3] and *Zernike* [17] have been implemented.

For those descriptors, it is important to fit the best parameters. For GFD, the parameters are tuned as, radial $(4:12)$ and angular $(6:20)$ frequencies to get the best combinations. For SC, the test follows [3] *i.e.*, 100 number of sample points are used. In case of *Zernike*, 36 *Zernike* functions of order less than or equal to 7 has been used. For radon, number of projecting angle is considered *i.e.*, $[0, \pi[$ by default.

C. Evaluation protocol

While experimenting, every test sample is matched with training candidates and the closest one is reported. The closest candidate corresponds to the labelled class, which we call recognition.

---

[1]ISI character datasets for Indian scripts, CVPR unit, India
ETL3 and ETL5 datasets for Roman and Katakana scripts, AIST, Japan

Figure 7. A sample per class from 3 different known scripts: 27 classes for Roman, 50 classes for Bangla and 47 classes for Katakana (Japanese).

To evaluate the performance of the method, $\mathcal{X}$−fold Cross-Validation (CV) has been implemented unlike traditional dichotomous classification. In $\mathcal{X}$−fold CV, the original sample for every class is randomly partitioned into $\mathcal{X}$ sub-samples. Of the $\mathcal{X}$ sub-samples, a single sub-sample is used for validation, and the remaining $\mathcal{X}-1$ sub-samples are used for training. This process is then repeated for $\mathcal{X}$ folds, with each of the $\mathcal{X}$ sub-samples used exactly once. Eventually, a single value results from averaging all. The opposite process holds for inverse $\mathcal{X}$−fold CV. While experimenting, tests move from normal $\mathcal{X}$−fold CV to inverse.

D. Experimental Results

Table I shows the average recognition rates for all datasets using $\mathcal{X}$−fold CV. In all datasets, $\mathcal{D}$−Radon supersedes all methods, while providing a marginal difference with SC. GFD is lagging behind the first two best performers. *Zernike* provides substantial difference in recognition rate against GFD, SC and $\mathcal{D}$−Radon. While, $\mathcal{R}$−signature performs far behind the all with the difference fairly in large amount.

Considering such datasets, we have observed that SC performs close to $\mathcal{D}$−Radon, except the two character datasets: Bangla and Katakana. GFD and SC almost go equally.

IV. EXPERIMENTAL RESULTS ANALYSIS

Considering the datasets (*cf.* Section III-A) used, methods (*cf.* Section III-B) are analysed in response to the evaluation protocol (*cf.* Section III-C) based on experimental results

Table I
AVERAGE RECOGNITION RATE IN % USING $\mathcal{X}$−FOLD CROSS-VALIDATION ($\mathcal{X} = 5$)

	Training	$\mathcal{R}$−sign.	Zernike	GFD	SC	$\mathcal{D}$-Radon
1. Numeral Datasets						
Roman 10×100	$\mathcal{X}-1$	78	83	97	98	**100**
	$\mathcal{X}-2$	75	78	95	97	**100**
	$\mathcal{X}-3$	70	74	94	96	**98**
	$\mathcal{X}-4$	66	67	91	96	**98**
Oriya 10×100	$\mathcal{X}-1$	58	44	**98**	98	98
	$\mathcal{X}-2$	52	38	96	95	**98**
	$\mathcal{X}-3$	46	43	85	93	**97**
	$\mathcal{X}-4$	43	32	72	92	**95**
Devanagari 10×300	$\mathcal{X}-1$	55	40	86	96	**97**
	$\mathcal{X}-2$	54	40	84	94	**96**
	$\mathcal{X}-3$	50	38	81	93	**96**
	$\mathcal{X}-4$	46	34	69	87	**95**
Bangla 10×400	$\mathcal{X}-1$	48	47	73	**95**	95
	$\mathcal{X}-2$	48	44	69	**94**	94
	$\mathcal{X}-3$	46	43	68	91	**93**
	$\mathcal{X}-4$	44	40	64	89	**92**
2. Character Datasets						
Roman 27×100	$\mathcal{X}-1$	80	77	96	99	**100**
	$\mathcal{X}-2$	77	74	94	**98**	98
	$\mathcal{X}-3$	75	67	93	97	**98**
	$\mathcal{X}-4$	69	58	91	**98**	97
Bangla 50×20	$\mathcal{X}-1$	22	11	23	55	**77**
	$\mathcal{X}-2$	18	09	22	49	**74**
	$\mathcal{X}-3$	14	06	18	46	**72**
	$\mathcal{X}-4$	11	06	16	43	**71**
Katakana 47×200	$\mathcal{X}-1$	69	42	88	91	**96**
	$\mathcal{X}-2$	66	37	85	87	**95**
	$\mathcal{X}-3$	63	34	84	84	**94**
	$\mathcal{X}-4$	59	28	78	81	**91**

in Table I (*cf.* Section III-D). In addition, a few common difficulties in character recognition will be highlighted.

(a) Multi-class similarity classes of characters.

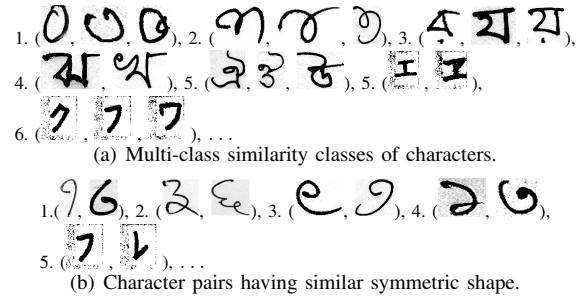(b) Character pairs having similar symmetric shape.

Figure 8. Difficulties in character recognition – a few examples.

Multi-class similarity is one of the major problems in character recognition as shown in Fig. 8 (a). In addition, shape descriptors with rotation invariance properties have been affected from those samples as shown in Fig. 8 (b). These are the major reasons affecting those existing methods. Missing parts as well as long ascender and/or descender part of the stroke affects SC severely compared to GFD. Due to these circumstances, existing methods require more training samples in order to get varieties of writing. This

has been proved in experimental results (see Table I) where substantial decrement in recognition rate has been observed from $\mathcal{X}-1$ to $\mathcal{X}-4$, where $\mathcal{X} = 5$. Despite, $\mathcal{D}-$Radon provides fairly better in those situations. Besides, scale invariance does not affect much as it employs DTW for radon features alignment.

Another important issue is computational complexity. In $\mathcal{D}-$Radon, running time complexity is high since it uses DTW for matching, however it is largely depend on how big the image is. As far as concern to computational cost, the observed average running time for all methods is provided in Table II. All tests have been made using MATLAB 7.8.0 in Linux platform.

Table II
AVERAGE RUNNING TIME FOR FEATURE SELECTION AND MATCHING FOR A SINGLE PAIR

	$\mathcal{R}-$sign.	GFD	*Zernike*	SC	$\mathcal{D}-$Radon
Time (sec.)	1	12	16	34	52

V. CONCLUSION AND FUTURE PERSPECTIVES

A method for character recognition based on the combination of radon features for representing pattern and dynamic programming for matching, has been presented. It supersedes the state–of–the–art of shape descriptors for isolated off-line character recognition over several different publicly available datasets.

Computing radon transform is quite immediate. But the running time complexity lies in matching. However, it could be substantially reduced by using optimised DTW [22].

REFERENCES

[1] Y. Alginahi. Preprocessing techniques in character recognition. *Character Recognition*, 2010.
[2] N. Arica and F. Yarman-Vural. An overview of character recognition focused on off-line handwriting. *IEEE SMC Part C: Applications and Reviews*, 31(2):216 –233, 2001.
[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE PAMI*, 24(4):509–522, 2002.
[4] U. Bhattacharya and B. B. Chaudhuri. Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals. *IEEE PAMI*, 31(3):444–457, 2009.
[5] T. K. Bhowmik, S. K. Parui, U. Bhattacharya, and B. Shaw. An hmm based recognition scheme for handwritten oriya numerals. *ICIT*, 0:105–110, 2006.
[6] J. Coetzer. *Off-line Signature Verification*. PhD thesis, University of Stellenbosch, 2005.
[7] T. E. de Campos, B. R. Babu, and M. Varma. Character recognition in natural images. In *VISAPP*, pages 273–280, 2009.
[8] S. R. Deans. *Applications of the Radon Transform*. Wiley Interscience Publications, New York, 1983.
[9] J. Flusser. On the independence of rotation moment invariants. *PR*, 33(9):1405 – 1410, 2000.
[10] P. Foggia, C. Sansone, F. Tortorella, and M. Vento. Combining statistical and structural approaches for handwritten character description. *IVC*, 17(9):701–711, 1999.
[11] D. Frejlichowski. Analysis of four polar shape descriptors properties in an exemplary application. In *ICCVG*, pages 376–383, 2010.
[12] G. H. Granlund. Fourier Preprocessing for Hand Print Character Recognition. *IEEE Trans. on Computers*, C–21(2):195–201, 1972.
[13] L. Heutte, T. Paquet, J.-V. Moreau, Y. Lecourtier, and C. Olivier. A structural/statistical feature based vector for handwritten character recognition. *PRL*, 19(7):629–641, 1998.
[14] J. Hopkins and T. L. Andersen. A fourier-descriptor-based character recognition engine implemented under the gamera open-source document-processing framework. In *DRR*, pages 111–118, 2005.
[15] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE PAMI*, 22(1):4–37, 2000.
[16] M. Khoddami and A. Behrad. Farsi and latin script identification using curvature scale space features. In *10th Symposium on NEUREL*, pages 213 –217, 2010.
[17] W.-Y. Kim and Y.-S. Kim. A region-based shape descriptor using zernike moments. *Signal Processing: Image Communication*, 16(1-2):95 – 102, 2000.
[18] S. Kopf, T. Haenselmann, and W. Effelsberg. Enhancing curvature scale space features for robust shape classification. In *ICME*, pages 478–481, 2005.
[19] J. B. Kruskall and M. Liberman. The symmetric time warping algorithm: From continuous to discrete. In *Time Warps, String Edits and Macromolecules: The Theory and Practice of String Comparison*, pages 125–161. Addison-Wesley, 1983.
[20] R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE PAMI*, 22:63–84, January 2000.
[21] J. R., S. R. Kolhe, and P. M. Patil. Dynamic time warping based static hand printed signature verification. *PRR*, 4(1):52–65, 2009.
[22] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11:561–580, October 2007.
[23] R. Shinghal and C. Suen. A method for selecting constrained hand-printed character shapes for machine recognition. *IEEE PAMI*, 4(1):74–78, 1982.
[24] S. Tabbone, L. Wendling, and J.-P. Salmon. A new shape descriptor defined on the radon transform. *CVIU*, 102(1):42–51, 2006.
[25] M. Tepper, D. Acevedo, N. A. Goussies, J. C. Jacobo, and M. Mejail. A decision step for shape context matching. In *ICIP*, pages 409–412, 2009.
[26] D. Zhang and G. Lu. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17:825–848, 2002.
[27] D. Zhang and G. Lu. Review of shape representation and description techniques. *PR*, 37(1):1–19, 2004.
[28] ivind Due Trier, A. K. Jain, and T. Taxt. Feature extraction methods for character recognition-a survey. *PR*, 29(4):641 – 662, 1996.