

## Aberystwyth University

### *Handwritten Chinese character recognition using fuzzy image alignment*

Li, Fangyi; Shen, Qiang; MacParthalain, Neil; Li, Ying

*Published in:*  
Soft Computing

*DOI:*  
[10.1007/s00500-015-1923-y](https://doi.org/10.1007/s00500-015-1923-y)

*Publication date:*  
2016

*Citation for published version (APA):*

Li, F., Shen, Q., MacParthalain, N., & Li, Y. (2016). Handwritten Chinese character recognition using fuzzy image alignment. *Soft Computing*, 20(8), 2939–2949. <https://doi.org/10.1007/s00500-015-1923-y>

#### **Document License** CC BY

#### **General rights**

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)

# Handwritten Chinese character recognition using fuzzy image alignment

Fangyi Li<sup>1</sup> · Qiang Shen<sup>2</sup> · Ying Li<sup>3</sup> · Neil Mac Parthaláin<sup>2</sup>

© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** The task of handwritten Chinese character recognition is one of the most challenging areas of human handwriting classification. The main reason for this is related to the writing system itself which encompasses thousands of characters, coupled with high levels of diversity in personal writing styles and attributes. Much of the existing work for both online and off-line handwritten Chinese character recognition has focused on methods which employ feature extraction and segmentation steps. The preprocessed data from these steps form the basis for the subsequent classification and recognition phases. This paper proposes an approach for handwritten Chinese character recognition and classification using only an image alignment technique and does not require the aforementioned steps. Rather than extracting features from the image, which often means building models from very large training data, the proposed method instead uses the mean image transformations as a basis for model building. The use of an image-only model means that

no subjective tuning of the feature extraction is required. In addition by employing a fuzzy-entropy-based metric, the work also entails improved ability to model different types of uncertainty. The classifier is a simple distance-based nearest neighbour classification system based on template matching. The approach is applied to a publicly available real-world database of handwritten Chinese characters and demonstrates that it can achieve high classification accuracy and is robust in the presence of noise.

**Keywords** Handwritten Chinese character recognition · Fuzzy-entropy · Group-wise image alignment

## 1 Introduction

Handwritten Chinese character recognition (HCCR) has been the topic of much interest, in both research circles and also for practical applications. However, the task of HCCR where the domain is non-constrained (in which the uncertainty is not limited to particular training examples of handwritten characters) is still a problem that poses significant challenges (Du et al. 2014). This is mainly due to the high levels of diversity of both handwriting styles and large character sets. These two aspects make the task of recognition and classification for handwritten Chinese characters much more complex than that of languages which are based upon the Roman or Latin alphabet writing systems (Hildebrandt et al. 1993; Suganthan and Yan 1998).

Chinese characters are used for daily communication by more than 25 % of the world's population, mainly in Asia. In basic terms, there are three main character sets: traditional Chinese characters, simplified Chinese characters, and Japanese Kanji. Indeed, literacy in written Chinese requires a knowledge of thousands of characters, many of which have

---

Communicated by D. Neagu.

✉ Neil Mac Parthaláin  
ncm@aber.ac.uk

Fangyi Li  
lifangyi@mail.nwpu.edu.cn

Qiang Shen  
qqs@aber.ac.uk

Ying Li  
lybyp@nwpu.edu.cn

<sup>1</sup> School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

<sup>2</sup> Department of Computer Science, Institute of Maths, Physics and Computer Science, Aberystwyth University, Aberystwyth, Ceredigion, Wales, UK

<sup>3</sup> Northwestern Polytechnical University, Xi'an 710072, China

their own unique shape and structure, but certain distinct characters may actually be visually very similar. In particular, Chinese characters are based on stroke structures and a great many groups of characters are constructed and share the same components. Such characters are easily confused even by humans. The most common Chinese character structures can be divided into five different categories: (1) characters which have a single structure, (2) characters which have a left and right structure, (3) characters which have an upper and a lower structure, (4) characters which have a ‘surrounded by’ structure, and (5) characters which have a framework structure (Zhang et al. 2009). Given this level of complexity, differences in writing styles (regular, fluent, and cursive) (Liu et al. 2004), and personal attributes of handwriting, it is easy to appreciate the complexity of the task at hand.

In most existing approaches to HCCR, a series of steps aimed at extracting features or feature information from images is performed (Du et al. 2014; Liu et al. 2004). In addition, such approaches typically also rely on very large training datasets (in the order of thousands of training examples) (Liu et al. 2004; Shao et al. 2014), resulting in models of considerable size and complexity. In this paper, an alternative approach is presented which does not require any feature extraction step. Instead, it relies upon an image alignment technique to bring various examples of characters into correspondence. This means that the resulting classification models are entirely data driven, and overly complex feature extraction steps are avoided.

In addition, metrics based upon fuzzy-entropy are employed to assess the degree of image alignment, offering much flexibility in capturing of the relations between image characteristics, and the ability to model both probabilistic and possibilistic uncertainty.

The image alignment process adopted here is termed congealing (Miller et al. 2000). It is a group-wise image alignment approach for a set of images and is not constrained to subjectively defined ‘good’ training examples. This approach allows the consideration of a set of images (or more generally, a set of arrays) and translates them with respect to a continuous set of given transformations to make the images more ‘similar’, according to a particular measure of assessment.

The remaining sections of this paper are organised as follows. Section 2 outlines the existing state-of-the-art work in HCCR and summarises the theoretical basis and concepts of image congealing. Section 3 presents the modified fuzzy image congealing algorithm. Section 4 describes the novel handwritten Chinese character recognition approach using fuzzy image congealing. Section 5 reports on the experiments carried out for fuzzy image congealing for handwritten Chinese characters and shows recognition results. Finally Sect. 6, concludes the paper and briefly discusses future work and possible extensions.

## 2 Related work

This section includes two parts: the background work relating to HCCR, and that of image congealing.

### 2.1 Handwritten Chinese character recognition

The recognition and classification of handwritten Chinese characters pose a significant challenge for automated methods (Du et al. 2014; Liu et al. 2004; Shao et al. 2014). Indeed the sheer number of characters, intricate complexity of such characters, and variations in writing styles mean that the task can be difficult even for humans (Tai 1992; Plamondon and Srihari 2000).

In mainland China, there are typically two character sets: simplified containing 3755 characters, and traditional containing 6763 characters. These have been formally recognised as an international standard, and the first set is a subset of the second (Tai 1992). However, in Taiwan there are 5401 traditional characters and these are defined in a single standard set. In both the traditional and simplified Chinese sets, around 5000 characters are regularly used. Japan on the other hand has 2965 Kanji characters (many of which are similar to Chinese characters), which are once again formalised in an international standard. There is also a second non-overlapping set of 3390 Kanji characters.

Chinese characters are ideographs and are composed primarily of a series of straight lines or multiple-line strokes. Quite a few characters contain fairly independent substructures or components, known as *radicals*. However, very different characters can share common radical components, for example, the character 备 (bei) and character 奋 (fen) share the same radical in the bottom half of the character, whilst the character 王 (wang) differs only from the character 玉 (yu) only by the short stroke in the right-hand corner. Other examples include the following three characters 辨 (bian), 辩 (bian) and 辨 (bian). They can only be distinguished from one another by their middle part. This property can pose a problem in recognition, often leading to confusion between distinct characters, particularly where feature extraction/recognition/classification maps similar features in different characters to the same decision. Handwriting further obscures the recognisability of such characters. This makes the task very challenging and has prompted much work in the area (Hildebrandt et al. 1993; Suganthan and Yan 1998).

The task of HCCR has attracted much interest and a number of different methods have been proposed for tackling this problem. Generally, a typical HCCR system implements a pre-processing, segmentation step initially to establish character patterns (Liu et al. 2004), followed by feature extraction to assess the structure type and representation. Since the structure may be represented differently, a variety of recognition approaches can be utilised to classify a given handwritten Chinese character to a particular class.

In the existing literature, almost all of such approaches rely on a certain form of feature extraction and underlying representation (Liu et al. 2004), with various approaches adopted for Chinese characters in this regard (Hildebrandt et al. 1993; Liu and Kim 2001; Umeda 1996). Basically, the approaches can be divided into two different types: those based upon the examining the structural characteristics of the characters (Okamoto and Yamamoto 1999) and those based upon the statistical features of the handwritten characters (Tai 1992). The former focuses upon writing stroke analysis, while statistical methods tend to focus upon features extracted from the shape information. The extraction of features and its subprocesses has a direct impact upon the recognition performance as those features extracted must offer a form of unified representation to build models in the steps that follow (classification/recognition/etc.).

To obtain reliable recognition performance, a divide-and-conquer classification approach is common amongst HCCR approaches. This is usually implemented using so-called ‘rough classification’ and ‘fine classification’ to decide which types of character structure the approach is dealing with Kato et al. (1999), Umeda (1996). Most existing techniques are variations on this theme of general approaches, although more specific methods (Du et al. 2014; Suganthan and Yan 1998; Wong Wong and Chan 1999) are often used for the recognition and classification step.

### 2.2 Image congealing

The goal of image congealing is to align or reduce the variation within a set of images. Such a technique is simple and relies on iteratively transforming all of the images by a small amount so that they are better aligned with respect to each other (Miller et al. 2000). As a result, all of the training images (and later testing images) can be brought into correspondence. To achieve this aim, a metric or objective function is employed. By optimising the objective function, the images are aligned with respect to a set of allowable transformations. There are two important concepts in this process: (a) the pixel-stack entropy and (b) the transformation matrices.

Given a ‘stack’,  $\mathbb{I}$ , of  $n$  images of size  $m$  pixels, a single pixel value in this stack is denoted  $x_i^j$  where  $i \in [1, m]$  and  $j \in [1, n]$  as illustrated in Fig. 1. The ‘stack’ of pixels  $\{x_i^1, x_i^2, \dots, x_i^n\}$  is denoted  $x_i$ . Each image,  $\mathbf{I} \in \mathbb{I}$ , is independently transformed by an affine transformation  $\mathbf{U}$ . The transformation for the  $j$ th image is denoted  $\mathbf{U}^j$  and the transformed pixel stack is denoted  $x_i^j$ .

The image congealing algorithm seeks to iteratively minimise the entropy across the stack of images by transforming each image by a small amount with respect to a set of possible affine transforms. The transform that gives rise to the greatest decrease in entropy is discovered via a hill-climbing approach. At each iteration of the algorithm, the goal is to

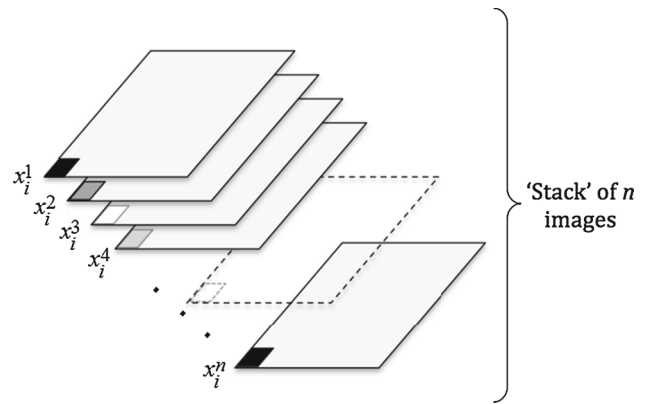


Fig. 1 Image congealing—a set of pixels drawn from the same location in the set of  $n$  images

minimise the cost function which corresponds to the sum of entropy across all images in the image ‘stack’ (as shown in Fig. 1):

$$f(\mathbb{I}) = \sum_{i=1}^n \left( - \sum_m \left( \frac{1}{n} \sum_j x_i^{j'}(k) \log_2 \frac{1}{n} \sum_j x_i^{j'}(k) \right) \right) \quad (1)$$

where  $x_i^j(k)$  is the probability of the  $k$ th element of the multinomial distribution in  $x_i^j$ . This process corresponds to minimising the total joint entropy of pixels across the stack of images and can be reformulated simply as

$$f(\mathbb{I}) = \sum_{i=1}^n H(D_i) \quad (2)$$

where  $H(D_i)$  is the entropy across the distribution field of probabilities for the  $i$ th pixel (Learned-Miller 2006). The minimisation can be performed by finding the transformation matrix  $\mathbf{U}$ , for each image, that maximises the log-likelihood of the image with respect to the distribution of pixels across the stack.

The transformation for the  $j$ th image, as denoted by  $\mathbf{U}^j$  is composed of component transforms:  $x$ -translation ( $t_x$ ),  $y$ -translation ( $t_y$ ), rotation ( $\theta$ ),  $x$ -scale ( $s_x$ ),  $y$ -scale ( $s_y$ ),  $x$ -shear ( $h_x$ ), and  $y$ -shear ( $h_y$ ). The affine transformation matrix  $\mathbf{U}^j$  can then be built such that

$$\begin{aligned} \mathbf{U} &= F(t_x, t_y, \theta, s_x, s_y, h_x, h_y) \\ &= \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} e^{s_x} & 0 & 0 \\ 0 & e^{s_y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & h_x & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ h_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

After each iteration the scale of all the transforms,  $U_1, \dots, U_n$ , are readjusted by the same amount so that the mean log-determinant of the transforms is 0; that is, the set of all transforms has zero-mean. The rationale for this step is to prevent the algorithm from succumbing to a situation where all of the images are shrunk to a point where they are no longer representative, as a result of an ever decreasing value of entropy.

Once the congealing algorithm has converged, or has reached a predefined number of iterations, the congealed versions of the input images can be revealed by multiplying the original images by their associated transformation matrix.

### 3 Fuzzy-entropy-based image congealing

Image congealing has been employed for a variety of different tasks and has been the subject of a number of different modifications (Cox et al. 2009; Learned-Miller et al. 2005; Learned-Miller 2006). An extension to the image congealing approach that uses fuzzy-entropy (FE) as an objective function is described in Mac Parthlain and Strange (2013) and uses a definition of FE that is derived from the work in Hu et al. (2006) and Kosko (1986). Several different definitions for FE have been proposed (Al-Sharhan et al. 2001), but this work utilises an approach based on similarity relations, which is described below.

#### 3.1 Similarity relation-based fuzzy-entropy

As the name suggests this definition of fuzzy-entropy is based on fuzzy-similarity relations. As such it collapses to classical information entropy (Shannon 1951), when the similarities between objects (image pixels in this case) are crisp. This ensures that the approach subsumes the original Boolean algorithm as well as offering the ability to model fuzzy uncertainty. Also, by adopting such an approach, no additional subjective thresholding information or ‘fuzziness’ parameter is required; only the information contained within the data is utilised. Formally, for a non-empty finite set  $X$ ,  $R$  is a binary relation defined on  $X$ , denoted by a relation matrix  $M(R)$ :

$$M(R) = \begin{pmatrix} r_{11} & r_{21} & \dots & r_{n1} \\ r_{12} & r_{22} & \dots & r_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1n} & r_{2n} & \dots & r_{nn} \end{pmatrix}$$

where  $r_{jk} \in [0, 1]$  is the (similarity) relation value of  $x_i^j$  and  $x_i^k$  in a particular pixel stack  $x_i$ . Note that a crisp equivalence relation will generate a crisp partition of the universe, whereas a fuzzy equivalence relation induces a fuzzy partition which subsumes the crisp condition as a specific case.

This property is important as it ensures that when all relations are crisp that the resulting partition is also crisp. In other words, crisp or discrete object values (non real-valued) will result in a crisp/discrete similarity matrix, populated only by absolute similarity values of 0 or 1.

For a finite set  $U$ ,  $A$  is a fuzzy or real-valued attribute set, which generates a fuzzy equivalence relation  $R$  on  $U$ . Given a fuzzy-similarity relation matrix  $M(R)$ , as previously defined, the fuzzy equivalence class generated by  $[x_i^j]$  where  $i \in [1, \dots, m]$  and  $j \in [1, \dots, n]$   $R$  can be defined by:

$$[x_i^j]_R = \frac{r_{j1}}{x_i^1} + \frac{r_{j2}}{x_i^2} + \dots + \frac{r_{jn}}{x_i^n} \tag{3}$$

Note that Eq. (3) is a conventional short-hand representation of a discrete fuzzy set, indicating that the membership value of the  $i$ th element of the  $j$ th image is  $r_{ji}$  where  $i \in [1, \dots, m]$  and  $j \in [1, \dots, n]$ . A number of different similarity relations can be used to induce such a matrix, but three that are particularly useful for this work and defined in Eqs. (7), (8), and (9). The cardinality of  $[x_i^j]_R$  is then represented by  $|[x_i^j]_R|$ . The uncertainty measure of the information embedded in the data or fuzzy-entropy of the fuzzy equivalence relation is then defined as:

$$H(x_i) = -\frac{1}{n} \sum_{j=1}^n \log \lambda_j. \tag{4}$$

where

$$\lambda_j = \frac{|[x_i^j]_R|}{n} \tag{5}$$

This definition of fuzzy-entropy directly offers a fuzzy information metric that may be used for the image congealing problem. Recall the concept of ‘image stack’ introduced earlier. By taking all of the pixel grey-level values across the stack, a metric based upon the classical entropy is used as an assessment of alignment. By modifying this metric, the fuzzy-entropy metric is now employed in its place.

The first step in formulating the FE metric is to induce a fuzzy-similarity relation matrix. This is also achieved using the pixel grey-level values in this particular case, although there are other image characteristics that can be used (Learned-Miller 2006). The similarity relation matrix is essentially a symmetric matrix which allows the comparison of all of the objects to each other. The matrix can be constructed by employing a fuzzy-similarity relation to determine the pairwise similarity of objects.

Let  $R_P$  be the fuzzy-similarity relation induced by a pixel  $P$ :

$$\mu_{R_P}(x, y) = T_{a \in P} \{ \mu_{R_a}(x, y) \} \tag{6}$$

$\mu_{R_a}(x, y)$  is the degree to which objects (image pixels in the ‘stack’)  $x$  and  $y$  are similar for the indexed pixel  $a$ . As mentioned previously this may be defined in many ways, thus allowing much flexibility in how the matrix is constructed. The fuzzy-similarity relations defined in (7), (8) and (9) are typical of those that may be used.

$$\mu_{R_a}(x, y) = 1 - \frac{|a(x) - a(y)|}{|a_{\max} - a_{\min}|} \quad (7)$$

$$\mu_{R_a}(x, y) = \exp\left(-\frac{(a(x) - a(y))^2}{2\sigma_a^2}\right) \quad (8)$$

$$\mu_{R_a}(x, y) = \max\left(\min\left(\frac{(a(y) - (a(x) - \sigma_a))}{(a(x) - (a(x) - \sigma_a))}, \frac{(a(x) + \sigma_a) - (a(y))}{(a(x) + \sigma_a) - (a(x))}\right), 0\right) \quad (9)$$

### 3.2 Fuzzy-entropy for image congealing

What image congealing seeks to do is to iteratively minimise the fuzzy-entropy  $E$  across all of the stacks of pixels for the whole of each image. This is done by transforming it with respect to a set of possible affine transforms. The aggregated measurement is essentially the summation of the fuzzy entropies amongst all  $m$  pixel stacks:

$$E = \sum_{i=1}^m H(x_i') \quad (10)$$

where  $x_i'$  is the transformed pixel stack as defined previously. One possible way of doing this can be described as follows. Firstly, a transformation matrix is generated for each image in the training set. Following that each of the parameters related to the transforms are changed iteratively. The fuzzy-

entropy metric is then used to assess whether the parameter modification has resulted in a decrease in the entropy. If that is the case, the transformation is retained and the matrix is modified in that direction accordingly. If, however, it results in an increase in entropy, then the transformation is reversed and the next transform is applied. The congealing process will stop when the fuzzy-entropy value no longer continues to fall or the maximum number of iterations is reached.

## 4 Fuzzy image congealing for handwritten Chinese character recognition

In this section, the novel approach for handwritten Chinese character recognition using fuzzy image congealing is described. The fuzzy image congealing process is employed not only in the training phase to construct character models, but also when a test character is presented for classification. The approach is illustrated in Figs. 2 and 3.

During the training phase, it is desirable to be able to bring all of the instances of each particular character into correspondence. This is done by aligning each character to all of the other training instances of the same class; that is, a set of training character images are aligned using the fuzzy-entropy-based congealing algorithm for that particular class. This process is then repeated for all training classes, which results in image models for each handwritten character. Both the training and application of such models are described below followed by an algorithmic complexity analysis

### 4.1 Model building

Compared with the feature extraction-based techniques (Liu et al. 2004) or other techniques such as those in Liu and Kim

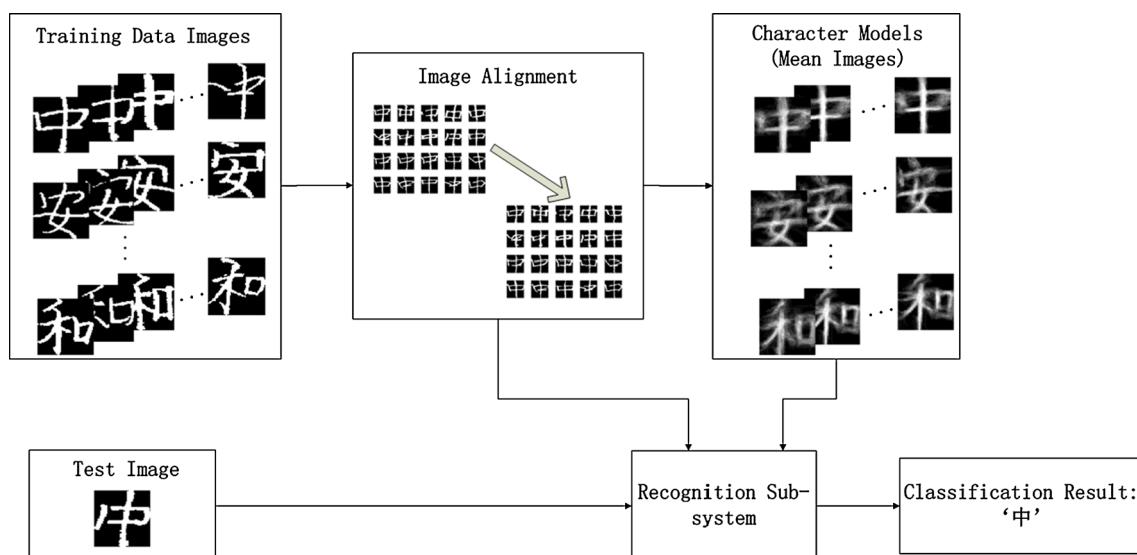


Fig. 2 Fuzzy congealing-based HCCR

(2001) and Suganthan and Yan (1998) for building classification models, the mean character image method employed here is much more straightforward and does not require any subjective judgement or decision with respect to the images under consideration. Moreover, since the character instances of a particular class have been aligned to be more similar to each other, once the alignment has taken place, a ‘mean image’ can be calculated from *all* of the congealed images. Such a mean image is, therefore, a wholly data-driven model.

For the model building step, mean images of each congealing *iteration* are collected for use as matched templates. Such mean images from different iterations rather than a *single* mean image obtained at the end of training congealing process are necessary because by aligning all of the images and then using the final mean image output, much of the variation that is useful for classifying test images can be retained. This is due to the fact that the alignment process when used to generate only a single mean image tends to result in a regularisation of all of the transformations and, therefore, removes any useful variation.

## 4.2 Recognition

The recognition sub-system can be seen in Fig. 3. As mentioned previously, a further fuzzy image congealing process is carried out when a new character is presented for classification.

The reason this new congealing step is performed during the recognition phase is that in general the test image will not initially be in correspondence with the training images. Since the same character written by different people can have quite different characteristics (size, shape, style and so on) it is necessary to align it so that it can be consistently classified. Once this has been completed, the similarity between the congealed images can be computed. Note that since the training images are already in correspondence with each other, only a small number of congealing iterations are required to bring the test image into correspondence with these.

To calculate the similarity between a test character and the classification models, a series of steps are necessary. These steps are shown in Fig. 3. During the congealing process for the training phase,  $k$  mean images are saved for each character after every iteration. Then, the test image is congealed with the mean image models of a particular character. Finally, a nearest neighbour classifier using a Euclidean distance metric is used to classify the test image (though other distance metrics may also be used for this purpose):

$$\frac{1}{k} \sum_{i=1}^k \|I_{\text{congealed\_test}} - I_{\text{congealed\_mean}(i)}\| \quad (11)$$

where  $\|\cdot\|$  represents the absolute distance.

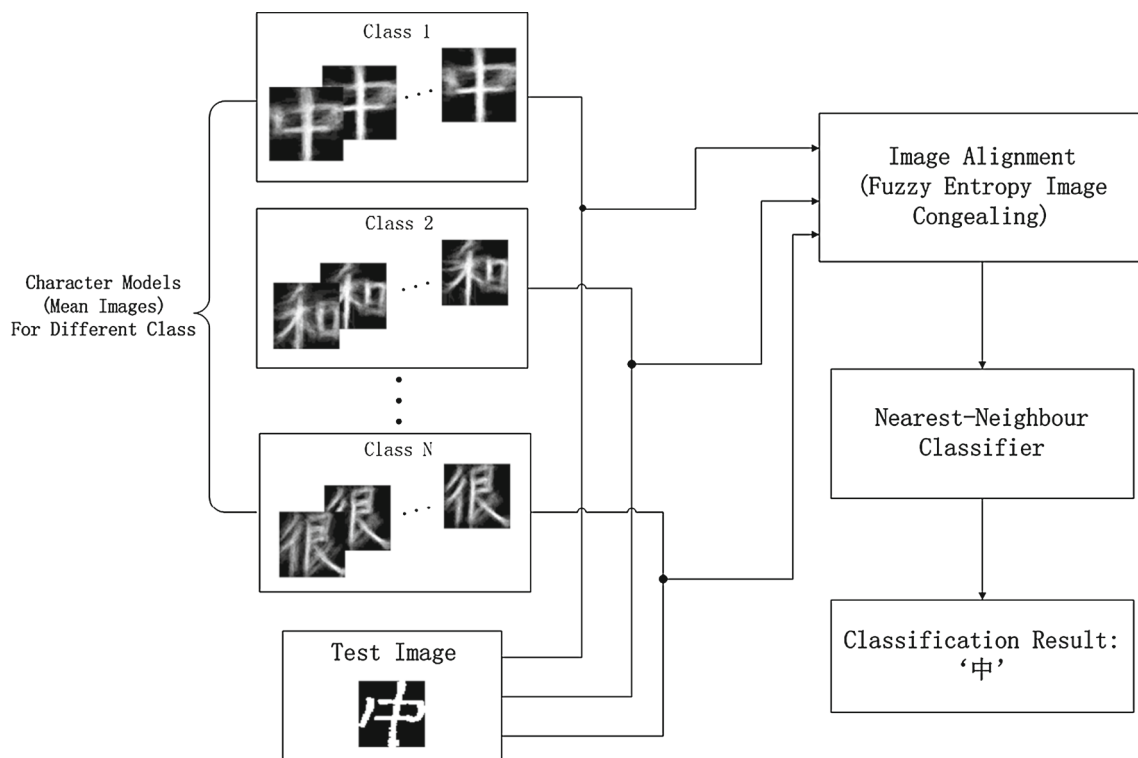


Fig. 3 Recognition sub-system in Fig. 2

This process is then repeated for each model character, respectively. The model with the smallest distance to the test character is then used for classification.

### 4.3 Complexity

The complexity of the original congealing algorithm based upon information entropy (Shannon 1951) is linear with respect to the number of parameters (related to the number of image transformations). It is clear that the incremental update involved happens for  $m - 1$  images and the update to each image is performed in constant time with respect to the (constant) number of parameters or transforms. This can be summarised as  $O(n - 1 \times m \times t)$  where  $n$ ,  $m$ , and  $t$  relate to the number of images, pixels, and transforms, respectively.

Unfortunately, as noted previously, the fuzzy-entropy metric requires the additional  $n \times n$  calculation for the similarity relation per pixel stack, over all image pixels. This results in added complexity of  $O(n - 1 \times (m - 1)^2 \times t)$ , meaning that the approach is quadratic with respect to the number of pixels. The additional overhead means that the fuzzy-entropy-based classifier system proposed here may not perform as quickly as the techniques based upon feature extraction methods or indeed the version that is based upon traditional information entropy. Initial thoughts on addressing this issue are discussed briefly in the conclusion.

## 5 Experimental evaluation

This section details the experiments conducted and the results obtained for the novel HCCR approach using fuzzy-entropy congealing algorithm. The handwritten Chinese characters used in a series of experiments are drawn from the HCL2000 dataset (Zhang et al. 2009).

### 5.1 Data and experimental setup

The HCL2000 dataset (Zhang et al. 2009) is a large-scale offline database which contains 3755 frequently used simplified Chinese characters handwritten by 1000 different subjects. The 3755 characters are formalised as the Chinese National Standard GB2312-80. Each character image is  $64 \times 64$  pixels and is binary in format, although congealed models contain grey values in the range [0, 255].

A large number of characters are commonly used in Chinese writing systems. Moreover, there are several types of structure for Chinese characters. For the experimentation detailed here, 80 different characters (or character classes as each character may be written in a variety of forms) are used. These are representative of the aforementioned types of structure and are selected from the list of most commonly

**Table 1** Chinese character data

Structure type	Character Examples	No.
Single	了,大,不	9
Left-and-right	阿,的,和	26
Upper-and-lower	宝,安,名	24
Surrounded-by	边,国,同	12
Framework	为,我,中	9
<b>Total</b>		<b>80</b>

used characters in modern Chinese. A selection of these can be seen in Table 1.

For the training and model construction phase, 20 different handwritten Chinese character instances of the same class are randomly chosen from the dataset to perform image alignment using fuzzy-entropy-based congealing. The fuzzy-similarity relation which was used for training is that shown in Eq. (8), where  $a$  is defined as a particular pixel stack. For each character model built, the total number of iterations was set to a value of 15. After the fuzzy congealing step has been completed, it results in 15 ‘mean’ images for each character of the training data.

For the recognition process, 15 mean character training images of a particular class are then congealed with the test image. The testing data consisted of 50 characters (5 instances of each class) which were randomly selected from the same dataset. Note that the number of congealing iterations in this second alignment process is only 3 since most of the 15 training images are mean samples of the training alignment process.

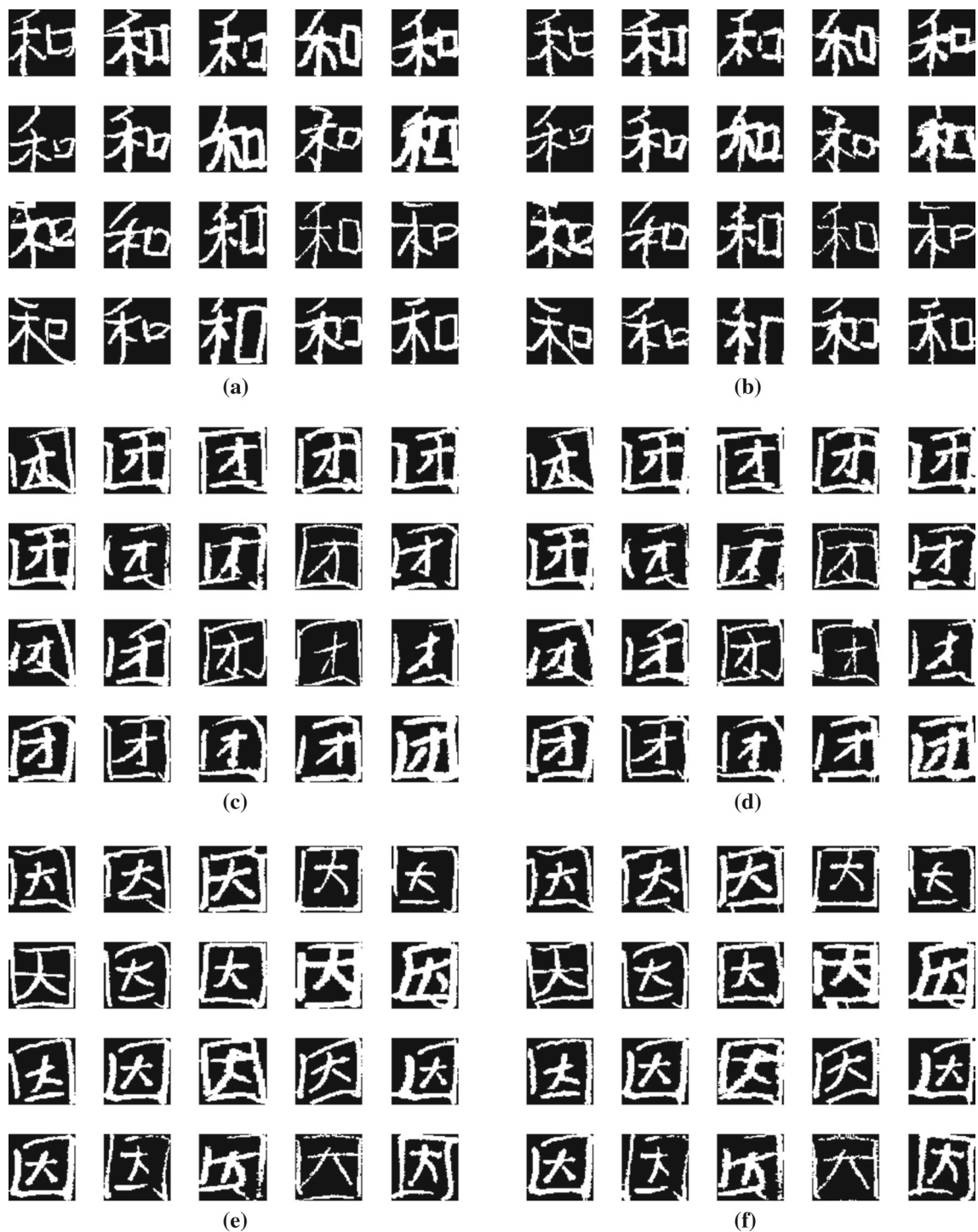
Whilst carrying out the testing phase, a more strict similarity relation is used (Eq. 7). The reason for doing so ensures that the classification has more harsh boundaries and that only very similar images can be classified into the correct class. For the fuzzy-similarity relation Eq. (7) is used.

To test the robustness of the approach, in the presence of noise, a separate series of experiments was also carried out. Ten character classes of three test samples per character (30 test characters in total) had Gaussian white noise artificially added to the images. The characters used for this are: 的 (de), 中 (zhong), 国 (guo), 和 (he), 安 (an), 我 (wo), 名 (ming), 同 (tong), 不 (bu), 大 (da). The classification accuracies are then compared for different variational intensities of noise.

### 5.2 Results

Since the generation of mean images is a key step in the proposed approach, and plays an important role as the classification template for the recognition process, examples of training samples for 3 characters: (1) 和 (he), (2) 团 (tuan), (3) 因 (yin), before and after fuzzy-entropy-based



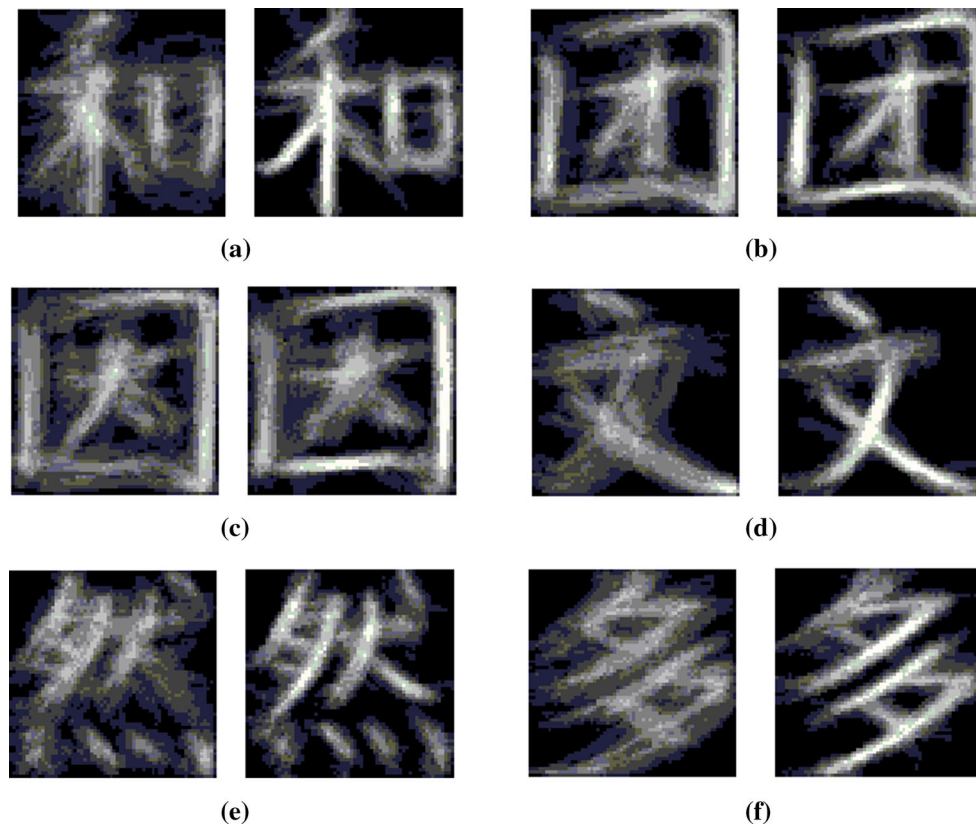


**Fig. 4** Handwritten characters 和 (he), 团 (tuan), 因 (yin), before congealing (a, c, e) and after congealing (b, d, f)

congealing are shown in Fig. 4. The mean images computed after all images in a particular class before and after 15 iterations of congealing are shown in Fig. 5. It is clear that the alignment has a significant effect on the models of the mean images and the variation contained therein.

### 5.2.1 Classification accuracy

Although a simple nearest neighbour classifier is employed for this work, it proves quite effective as the 90.50% average classification accuracy (weighted by the number of instances of each class) shown in Table 2 demonstrates. This is con-



**Fig. 5** Mean images of handwritten Chinese characters before and after fuzzy image congealing: *left-hand images* are the mean images prior to alignment and *right-hand images* are the mean images after congealing

sistent with the state of the art, with current unconstrained approaches reporting similar accuracies—92.39% (Du et al. 2014; Liu et al. 2004). Indeed, the *framework* and *single structure* characters, prove easiest to classify with a rate of 100%. The *upper and lower* and *surrounded-by* characters are a little more challenging, however, with accuracies of 85 and 83.33%, respectively. Most of the confusion in the classification and recognition phase relating to the individual characters of these two types of structure happens between characters which are very similar to one another. This indicates that additional effort needs to be focused on the removal or at least the significant reduction of such confusion in future work.

An excerpt from the full confusion matrix for those characters which are most problematic for the approach is shown in Table 3. It can be seen that 国 (guo) is most often confused with the characters 团 (tuan), 困 (yin) and 同 (tong). This misclassification can be explained by the similarity of these characters in terms of their structure types. The same is true for the character 奋 (fen), which is confused with 备 (bei). Indeed there seems to be a level of symmetry between these two characters in terms of confusion but again the structures are similar. Rather less intuitive is the confusion between the character 伴

**Table 2** Classification accuracies for types of Chinese characters

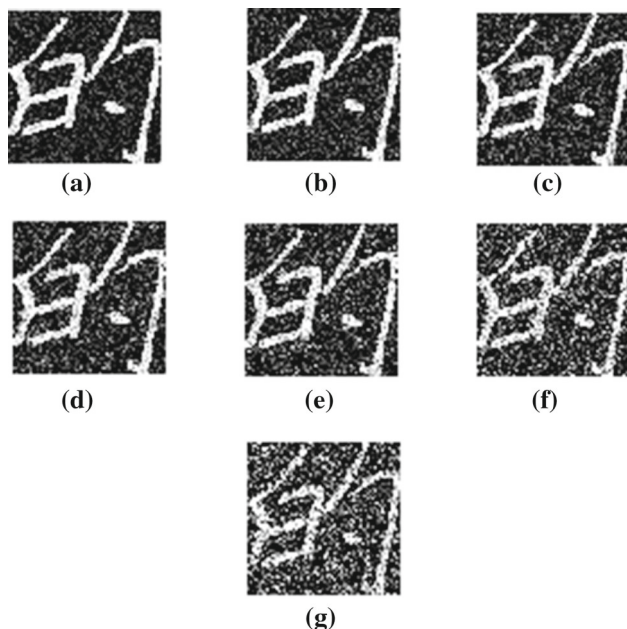
Character structure	Classification accuracy (%)
Single	100
Left-and-right	92.31
Upper-and-lower	85.00
Surrounded-by	83.33
Framework	100

(ban) and 备 (bei) as these are drawn from different structures (left–right and upper and lower, respectively). An in-depth investigation into this remains an important further research.

As can be seen in Table 3, those characters which belong to the same character structure are easily confused and can be classified incorrectly. For example, the character 国 (guo), 同 (tong), 团 (tuan) and 困 (yin) belong to the ‘Surrounded-by’ type structure, since the character 国 (guo) looks similar to the other 3 characters. The classification accuracy of the characters of the ‘Surrounded-by’ type structure is indeed the lowest in Table 2 and once again emphasises the challenge facing HCCR systems.

**Table 3** Excerpt from confusion matrix for difficult cases

Actual Char.	Classified as							
	国	同	伴	体	奋	备	团	因
国	1	1	0	0	0	0	2	1
同	0	5	0	0	0	0	0	0
伴	0	0	3	0	0	2	0	0
体	0	0	0	5	0	0	0	0
奋	0	0	0	0	3	2	0	0
备	0	0	0	0	1	4	0	0

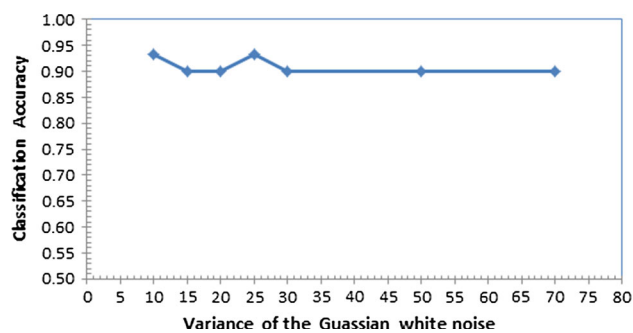
**Fig. 6** Character 的 (de) with increasing levels of added noise: a–g show the intensities 10–70, respectively

### 5.2.2 Robustness

A subset of the original data is used for this series of experiments as mentioned previously. To illustrate the effects of noise, an example of the character 的 (de) is shown in Fig. 6 with different intensities of artificially added noise. Despite an initial fall in accuracy, it is encouraging to note that even with extreme levels of added noise, the classification accuracy remains stable, as shown in Fig. 7. This robustness may stem from the use of mean image templates as the action of transforming all images until they are aligned in some ways mitigates this.

## 6 Conclusion

This paper has proposed an image-only model for performing handwritten Chinese character recognition. The motivation for such an approach stems from the desire to employ a purely data-driven approach, and therefore avoid any associated

**Fig. 7** Classification accuracy in terms of different levels of noise

subjective decision-making that requires feature extraction. As such it also helps to avoid generating large and complex models which require enormous amounts of training data. Since handwritten Chinese characters are much more complex than those of Latin/Roman writing systems, such a step is important in ensuring robust and accurate classification whilst simultaneously avoiding the overhead in dealing with large numbers of data samples.

Although the approach detailed in this paper offers important advantages over existing methods, there are a number of areas which require further attention. The complexity of the approach (due to the computation of fuzzy-similarity relations) needs to be addressed. To tackle this, more computationally efficient ways of calculating the fuzzy-entropy for the image stack may prove useful. The definition of fuzzy-entropy used here is based on similarity relations, but an alternative distance-based interpretation could also be implemented. Also, it would be interesting to investigate all current similarity measures to discover if there is a particular definition which may offer similar performance but with reduced computation. Another aspect of the approach which would benefit from further attention is the classification phase. Currently classification is carried out using a simple but effective template matching nearest neighbour approach. However, the density of the transforms may provide more useful information as noted in Miller et al. (2000).

Whilst the application of the proposed approach here is limited to handwritten Chinese characters, it could equally be applicable for other types of imaging (e.g. medical imaging Chen et al. 2014, remote sensing Li et al. 2014, and planetary imaging Shang and Barnes 2013 etc.). Indeed, there is no reason why it could not also be adapted for the handwriting of other writing systems such as Bangla, Thai script, Arabic or others.

**Acknowledgments** This work was carried out when the first author was with Aberystwyth University as a visiting student. The project team is grateful to the China Scholarship Council and the Royal Academy of Engineering for their support in this research.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

#### Compliance with ethical standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Al-Sharhan S, Karray F, Gueaieb W, Basir O (2001) Fuzzy entropy: a brief survey. In: Proceedings of the 10th IEEE international conference on fuzzy systems, pp 1135–1139
- Chen Z, Strange H, Denton E, Zwiggelaar R (2014) Analysis of mammographic microcalcification clusters using topological features. In: LNCS breast imaging, pp 690–697
- Cox M et al (2009) Least-squares congealing for large numbers of images. In: 2009 IEEE 12th international conference on computer vision. IEEE, New York
- Du J, Hu JS, Zhu B, Wei S, Dai LR (2014) A study of designing compact classifiers using deep neural networks for online handwritten Chinese character recognition. In: 2014 22nd international conference on pattern recognition (ICPR). IEEE, New York
- Hildebrandt TH, Liu W (1993) Optical recognition of handwritten Chinese characters: advances since 1980. *Pattern Recognit* 26(2):205–225
- Hu Q, Yu D, Xie Z (2006) Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recognit Lett* 27(5):414–423
- Kato N, Suzuki M, Omachi S, Aso H, Nemoto Y (1999) A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance. *IEEE Trans Pattern Anal Mach Intell* 21(3):258–262
- Kosko B (1986) Fuzzy entropy and conditioning. *Inf Sci* 40(2):165–174
- Learned-Miller EG (2006) Data driven image models through continuous joint alignment. *IEEE Trans Pattern Anal Mach Intell* 28(2):236–250
- Learned-Miller EG, Jain V (2005) Many heads are better than one: jointly removing bias from multiple MRIs using nonparametric maximum likelihood. Information processing in medical imagingSpringer, Berlin
- Li Y, Li P, Shen Q (2014) Real-time infrared target tracking based on  $l_1$  minimization and compressive features. *Appl Opt* 53:6518–6526
- Liu C-L, Jaeger S, Nakagawa M (2004) Online recognition of Chinese characters: the state-of-the-art. *IEEE Trans Pattern Anal Mach Intell* 26(2):198–213
- Liu C-T, Kim I-J, Kim JH (2001) Model-based stroke extraction and matching for handwritten Chinese character recognition. *Pattern Recognit* 34(12):2339–2352
- Mac Parthlain N, Strange H (2013) Fuzzy-entropy based image congealing. In: IEEE international conference on fuzzy systems (FUZZ-IEEE)
- Miller EG, Matsakis NE, Viola PA (2000) Learning from one example through shared densities on transforms. In: Proceedings of 2000 IEEE international conference computer vision and pattern recognition
- Okamoto M, Yamamoto K (1999) On-line handwritten character recognition method using directional features and clockwise/counterclockwise direction-change features. In: Proceedings of the fifth international conference on document analysis and recognition, pp 491–494
- Plamondon R, Srihari SN (2000) Online and off-line handwriting recognition: a comprehensive survey. *IEEE Trans Pattern Anal Mach Intell* 22(1):63–84
- Shang C, Barnes D (2013) Fuzzy-rough feature selection aided support vector machines for mars image classification. *Comput Vis Image Underst* 117(3):202–213
- Shannon CE (1951) Prediction and entropy of printed English. *Bell Syst Tech J* 30:50–64
- Shao Y, Wang C, Xiao B (2014) A character image restoration method for unconstrained handwritten Chinese character recognition. *Int J Doc Anal Recognit (IJ DAR)* 18(1):1–14
- Suganthan PN, Yan H (1998) Recognition of handprinted Chinese characters by constrained graph matching. *Image Vis Comput* 16(3):191–201
- Tai W (1992) Some research achievements on Chinese character recognition in China. In: Wang PSP (ed) Character and handwriting recognition: expanding frontiers. World Scientific, Singapore, pp 199–206
- Umeda M (1996) Advances in recognition methods for handwritten Kanji characters. *IEICE Trans Inf Syst* 79(5):401–410
- Wong P-W, Chan C (1999) Postprocessing statistical language models for handwritten Chinese character recognizer. *IEEE Trans Syst Man Cybern Part B Cybern* 29(2):286–291
- Zhang H, Jun G, Chen G, Li C (2009) HCL2000—a large-scale handwritten Chinese character database for handwritten character recognition. In: Proceeding of the 10th IEEE international conference on document analysis and recognition ICDAR'09. IEEE, New York, pp 286–290