From Informational Confidence to Informational Intelligence

Stefan Jaeger

From Informational Confidence to Informational Intelligence

Stefan Jaeger
Laboratory for Language and Media Processing
Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20742, USA
jaeger@umiacs.umd.edu

Abstract

This paper is a continuation of my previous work on informational confidence. The main idea of this technique is to normalize confidence values from different sources in such a way that they match their informational content determined by their performance in an application domain. This reduces classifier combination to a simple integration of information. The proposed method has shown good results in handwriting recognition and other applications involving classifier combination. In the present paper, I will focus more on the theoretical properties of my approach. I will show that informational confidence has the potential to serve as a theory for learning in general by showing that this approach naturally leads us to the famous Yin/Yang symbol of Chinese philosophy, a classic symbol describing two opposing forces. Furthermore, a closer inspection of the opposing forces and their interplay will reveal a new information-theoretical meaning of the golden ratio, which describes the points where both confidence and counter-confidence merge into one force, with performance matching expectation. Although this is mainly a theoretical paper, I will present some practical results for handwritten Japanese character recognition.

Keywords: Classifier Combination, Sensor Fusion, Information Theory, Machine Learning, Japanese Character Recognition

1. Introduction

Classifier combination has become a popular approach in pattern recognition. The prospect of achieving high recognition rates with a set of relatively simple classifiers instead of one single and hard to optimize classifier has attracted many researchers. The numerous publications reporting improvements in recognition performance show that multiple classifier systems are indeed powerful. Despite the progress in recent years, however, researchers are still struggling to find the optimal way of combining different classifiers and put classifier combination or sensor fusion on a solid theoretical basis. In my previous work, I proposed an information-theoretical solution to this problem. My idea is to normalize confidence values from different sources according to their performance in a given application domain. The normalized confidence values match their information actually conveyed. Once I have replaced each confidence value by its corresponding normalized value, classifier combination becomes a straightforward integration of information. I have shown the effectiveness of this approach for character recognition and other document processing applications [3, 4, 6]. In this paper, I am going to elaborate more on the theoretical aspects of informational confidence [5]. I will show that informational confidence offers an appealing theoretical framework that may serve as a more general model for learning processes, thus justifying the term "informational intelligence" as a name for this approach.

I structured my paper as follows: Section 2 introduces again the basic ideas of informational confidence. Section 3 shows how we can learn informational confidence from feedback provided by the application domain. The next section then provides practical experiments for combined on-line/off-line recognition of handwritten Japanese characters. In Section 5, I will present the actual theoretical contribution of this paper, making connections to Yin/Yang and the golden ratio.

2. Informational Confidence

For readers not familiar with informational confidence, this section describes shortly the main idea. More information can be found in the aforementioned papers [3, 4, 6].

On an abstract level, informational confidence is defined by the following fixed point equation, which defines a simple linear relationship between confidence and information:

$$K_i \quad = \quad E * I\left(\overline{p}(K_i)\right) + C \tag{1}$$

In this equation, K_i is the i-th confidence value of a finite set of discrete values that a classifier can output as its confidence in a recognition result. Integer values are not a restriction per se, but they make things a bit easier from an implementational point of view, as we will see later in Section 3 when I show how to learn informational confidence values. Parameter E is a multiplying scalar and C is an offset that I will simply set to zero in the following. The function $I()$ computes the information using the negative logarithm, as introduced by Shannon in [12]. Its argument $\overline{p}(K_i)$ is the complement of the performance

function $p(K_i)$ of K_i, which provides the performance of each K_i and is computed in a given application domain. We see that the information, and thus K_i, becomes larger for higher performances. Using $1 - p(K_i)$ as the complement of $p(K_i)$, Eq. 1 reads as follows:

$$K_i = -E * \ln(1 - p(K_i)) \quad (2)$$

The mathematical definition of the performance function follows directly by resolving Eq. 2 for $p(K_i)$:

$$\begin{aligned} & K_i = -E * \ln(1 - p(K_i)) \\ \iff & e^{-\frac{K_i}{E}} = 1 - p(K_i) \\ \iff & p(K_i) = 1 - e^{-\frac{K_i}{E}} \end{aligned} \quad (3)$$

This result shows that the performance function $p(K_i)$ is a distribution of exponentially distributed confidence values, as I will explain in the following text. I can therefore consider confidence as a random variable with exponential density and parameter $\lambda = \frac{1}{E}$. The general definition of an exponential density function $e_\lambda(x)$ with parameter λ is:

$$e_\lambda(x) = \begin{cases} \lambda * e^{-\lambda x} & : \quad x \geq 0 \\ 0 & : \quad x < 0 \end{cases} \qquad \lambda > 0 \quad (4)$$

Parameter λ influences the steepness of the exponential density curve: The higher λ, the steeper the corresponding exponential density function. Based on the density function, a distribution describes the probability that the random variable assumes values lower than or equal to a given value k. For a random variable with exponential density $e_\lambda(x)$, we can compute the corresponding distribution $E_\lambda(k)$ as follows:

$$\begin{aligned} E_\lambda(k) &= \int_{-\infty}^{k} e_\lambda(x)\, dx \\ &= \int_{0}^{k} \lambda * e^{-\lambda x}\, dx \\ &= \left[-e^{-\lambda x}\right]_0^k \\ &= 1 - e^{-\lambda k} \end{aligned} \quad (5)$$

Consequently, a random variable with exponential density is also called "exponentially distributed random variable." Again, parameter λ influences the steepness of the distribution function: The higher λ, the steeper the distribution function. Moreover, there is a direct relation between λ and the expectation value E of the exponentially distributed random variable. Both are in inverse proportion to each other, i.e. $E = \frac{1}{\lambda}$.

We see that the only difference between the performance specification in (3) and the exponential distribution in (5) lies in the exponent of the exponential function. In fact, we can make performance and exponential distribution the same by simply setting λ to $\frac{1}{E}$. This relationship between performance and distribution now also sheds light on the parameter E. As mentioned above, the expectation value of an exponentially distributed random variable with parameter λ is $\frac{1}{\lambda}$. The parameter E therefore denotes the specific expectation value for classifier C. I summarize this important result in the Performance Theorem:

Performance Theorem:
A classifier C with performance $p(K)$ provides informational confidence $K = -E * \ln(1 - p(K))$ if, and only if, $p(K)$ is an exponential distribution with expectation E.

The next section shows how I learn informational confidence values.

3. Learning

In practice, classifiers typically violate the fixed point equation in the performance theorem and thus do not provide informational confidence values. Nevertheless, I can estimate the informational values from the performance of the raw values in the application domain. In addition to the native training algorithm of each classifier, I apply a second training process that learns confidence values that satisfy the requirement stated in the performance theorem. Motivated by the fact that the performance function describes an exponential distribution, I first estimate the performance for each confidence value K_i on an evaluation set, and then compute the corresponding informational confidence values by inserting these estimates into the fixed point equation of the performance theorem. Mathematically, my performance estimate is based on accumulated partial frequencies and is defined as follows [5, 6, 13]:

$$\hat{p}(\hat{K}_i) = \frac{\sum_{k=0}^{i} n_{correct}(K_k)}{N} \quad (6)$$

In this equation, K_i is again the ith confidence value and N is the number of all patterns tested. The help function $n_{correct}(K_k)$ returns the number of patterns correctly classified with confidence K_k. The idea of this performance estimate is to describe the different areas delimited by the confidence values under their common density function. For the estimate $\hat{E}$ of parameter E, I use basically the overall recognition rate of the classifier. For more information on the estimate of parameter E, I refer readers to the references [3, 4, 6].

Insertion of the estimates $\hat{p}(K_i)$ and $\hat{E}$ into Eq. 2 provides the estimated informational confidence values $\hat{K}_i$:

$$\hat{K}_i = -\hat{E} * \ln(1 - \hat{p}(K_i)) \quad (7)$$

The adjusted confidence values $\hat{K}_i$ replace the old values whenever I integrate these values with those of other classifiers, which adjust their values in exactly the same way. Since $\hat{p}(\hat{K}_i)$ is a monotonously increasing function over K_i, the new confidence values $\hat{K}_i$ will also increase monotonously and will thus not change the relative order of the original confidence values. Hence, informational confidence values will have no affect on the recognition rate of a single classifier system, except for ties introduced by mapping two different confidence values to the same

Table 1. Single n-best rates for handwritten Japanese character recognition.

Japanese	offline	online	AND	OR
1-best	**89.94**	81.04	75.41	95.56
2-best	94.54	85.64	82.62	97.55
3-best	95.75	87.30	84.99	98.06

Table 2. Combined recognition rates for handwritten Japanese character recognition.

Japanese (**89.94**)	Raw Confidence	Inf. Confidence
Sum-rule	93.25	**93.78**
Max-rule	91.30	91.14
Product-rule	92.98	65.16

informational confidence value. I also experimented with other performance estimates, and refer the reader again to the references for more information [3, 4, 6].

The next section will present practical evaluations for a multiple classifier system recognizing on-line Japanese characters.

4. Practical Experiments

Handwriting recognition is a very promising application field for classifier combination. The duality of handwriting recognition, with its two branches off-line recognition and on-line recognition, makes it suitable for multiple classifier systems. Compared to the time-independent pictorial representations used by off-line classifiers, on-line classifiers suffer from stroke-order and stroke-number variations inherent in human handwriting and thus in on-line data. On the other hand, dynamic on-line data provides valuable information that helps on-line classifiers to discriminate between classes with higher accuracy. Off-line and on-line classifiers thus complement each other, and their combination can overcome the problem of stroke-order and stroke-number variations.

The experiments I am going to present here use slightly different data sets compared to my earlier work reported in [4]. In particular, I use the performance estimate defined in Eq. 6 and normalize it to one bit, as explained in [4]. I will not go into a more detailed discussion about the advantages of different performance estimates. The jury is still out as to what performance estimate generally provides the best overall performance. The performance estimate in Eq. 6 has provided consistent improvement over a range of applications and displays the important characteristic of not changing the order of the original confidence values [3, 4, 6].

Table 1 lists the individual recognition rates for an off-line and on-line classifier [7, 10, 11]. Both classifiers were trained with more than one million handwritten Japanese characters. While the on-line classifier was trained with genuine on-line data, the off-line classifier was trained with the corresponding off-line images generated by a sophisticated painting method [7, 14]. The test and evaluation set contains $54,775$ handwritten characters. From this set, I took about two third of the samples to estimate the performances and about one third to do the final evaluation of the recognition rate. The off-line recognition rates are much higher than the corresponding on-line rates. Clearly, stroke-order and stroke-number variations are largely responsible for this performance difference. They complicate considerably the classification task for the on-line classifier. The last two columns of Table 1 show the percentage of test patterns for which the correct class label occurs either twice (AND) or at least once (OR) in the n-best lists of both classifiers. The relatively large gap between the off-line recognition rates and the numbers in the OR-column suggests that on-line information is indeed complementary and useful for classifier combination.

Table 2 shows the recognition rates for combined off-line/on-line recognition, using sum-rule, max-rule, and product-rule as combination schemes. Sum-rule adds the confidence values provided by each classifier for the same class, while product-rule multiplies the confidence values. Max-rule simply takes the maximum confidence without any further operation. The class with the maximum overall confidence will then be chosen as the most likely class for the given test pattern. Note that the sum-rule is the mathematically appropriate combination scheme for integrating information from different sources [12]. Also, the sum-rule is very robust against noise, as was shown in [8]. The upper left cell of Table 2 lists again the best single recognition rate from Table 1, achieved by the off-line recognizer. The second column contains the combined recognition rates for the raw confidence values as provided directly by the classifiers, while the third column lists the recognition rates for informational confidence values computed according to Eq. 7. Compared to the individual rates, the combined recognition rates in Table 2 are clear improvements. The sum-rule on raw confidence values already accounts for an improvement of almost 3.5%. The best combined recognition rate achieved with normalized informational confidence is 93.78%. It outperforms the off-line classifier, which is the best individual classifier, by almost 4.0%. Sum-rule performs better than max-rule and product-rule, a fact in accordance with the results in [8].

5. Informational Intelligence

In this section, I show that informational confidence is not only a concept for classifier combination, but also for learning and natural processes in general [5]. In other words, informational confidence can be considered a concept not confined merely to pattern recognition problems. I therefore choose to use the term “informational intelligence” to convey this broader meaning of informational confidence.

My basic idea is that decision making is based on two opposing forces, one supporting a certain outcome and

one acting against it. In the following, I propose a mathematical formalization of these two opposing forces based on Eq. 2. In fact, I postulate that the first force, Force A, is already defined by Eq. 2, with

$$K = -E * \ln\left(1 - p(K)\right) \tag{8}$$

As shown above, the performance function $p(K)$ follows immediately as $p(K) = 1 - e^{-\frac{K}{E}}$. If the performance in the logarithmic expression on the right-hand side of Eq. 8 is 1, and given a positive expectation, then A-Force becomes infinity. On the other hand, if the performance is zero, then the logarithm becomes zero and there is no A-force at all.

The second force, Force B, is defined similarly but performs complementary to Force A. The difference lies in the interpretation of the performance function:

$$K = -E * \ln\left(p(K)\right) \tag{9}$$

Again, the performance function $p(K)$ follows immediately by a straightforward transformation:

$$\begin{aligned} & K = -E * \ln\left(p(K)\right) \\ \iff \quad & p(K) = e^{-\frac{K}{E}} \end{aligned} \tag{10}$$

We see that the performance function of Force B is similar to the performance of Force A. However, it looks at the problem from a different side. Instead of describing the area delimited by K under its exponential density curve, it describes the remaining area not delimited. Parameter E is again a statistical expectation value. Unlike Force A, Force B becomes infinity for a performance equal to zero, and becomes zero itself whenever the performance is perfect, i.e. $p(K) = 1$. While Force A defines informational confidence values, Force B can be considered as defining informational counter-confidence values.

Having defined both Force A and Force B, I postulate that all processes are the result of the interplay between these two forces. What we can actually experience is the dominance that one of these forces has achieved over its counterpart. Mathematically, I describe the net effect of both forces using the difference between the fixed point equations defining both forces:

$$K = -E * \ln\left(\frac{1 - p(K)}{p(K)}\right) \tag{11}$$

This equation is a fixed point equation itself. It describes the net force, which is the result of both forces acting simultaneously. Naturally, the net force becomes zero when Force A equals Force B. This is the case when either the expectation value is zero or the performance $p(K)$ is 0.5.

Let us now assume that the expectation always matches the performance, i.e. $E = p(K)$. In this case, the definition of the net force in Eq. 11 reads as

$$K = -p(K) * \ln\left(\frac{1 - p(K)}{p(K)}\right) \tag{12}$$

Figure 1 graphically shows the net force of Eq. 12 depending on the performance $p(K)$. As I said before, the net force becomes zero for $p(K) = 0$ and $p(K) = 0.5$. For performances higher than 0.5, the net force goes into infinity. Figure 1 also shows a mirrored variant of the net force, namely

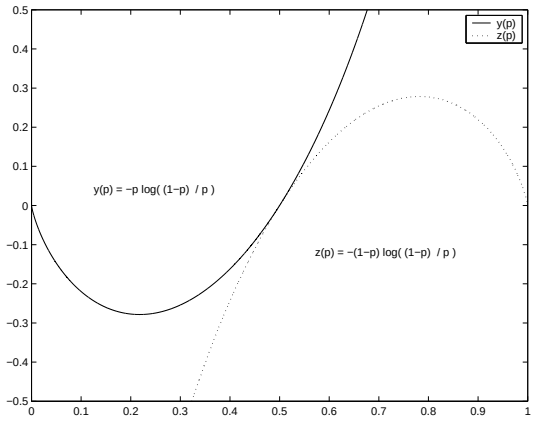

Figure 1. Net Force.

$$K = -(1 - p(K)) * \ln\left(\frac{1 - p(K)}{p(K)}\right) \tag{13}$$

Eq. 13 follows directly from Eq. 12 when we change the sign of the equation and replace $p(K)$ with $1 - p(K)$. The net force and its mirrored variant both meet at $p(K) = 0.5$. We can actually consider $p(K) = 0.5$ as a transition point where the net force transforms into the mirrored variant. After the transition, we are still looking at the same problem. However, our point of view and evaluation has changed, reflected now by the mirror net force. This will become important in the next section.

5.1. Yin and Yang

I now relate the above theoretical results with one of the oldest philosophical world views, namely the principle of Yin and Yang, which is the archetype of opposing forces. In particular, I dare to advance the hypothesis that Yin and Yang are the underlying forces of all learning processes, with Yin and Yang being defined by the fixed point equations given above. If this can indeed be confirmed by further observations, this ancient philosophical concept could play an important role in machine learning.

5.1.1. Philosophy

The concept of Yin and Yang is deeply rooted in Chinese philosophy. Its origin dates back at least 2500 years, probably much earlier, playing a crucial role in the oldest Chinese philosophical texts. Yin and Yang stand for two principles that are opposites of each other, and which are constantly trying to gain the upper hand over each other. However, neither one will ever succeed in doing so, though one principle may temporarily dominate the other one. Both forces define a perennial alternating cycle of Yin or Yang dominance. The central tenet of Chinese philosophy is that Yin and Yang need to be in harmony because only the equilibrium between Yin and Yang is able

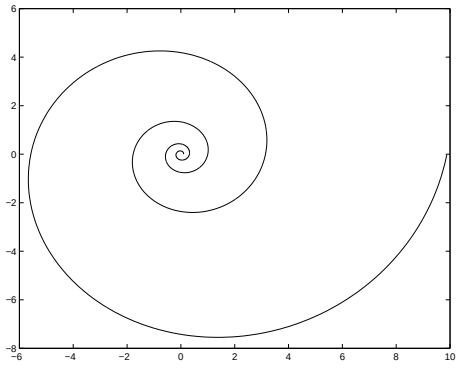

Figure 2. Logarithmic spiral.

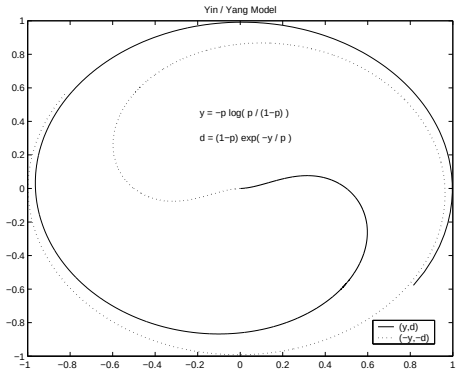

Figure 3. Yin-Yang Spirals.

to overcome this cycle. According to Chinese philosophy, the principle of Yin and Yang is the foundation of the entire universe. They flow through, and thus affect, every being. Any imbalance of economical, biological, physical, or chemical systems can be directly put down to a distorted equilibrium between Yin and Yang. According to the principle of Yin and Yang outlined above, neither Yin nor Yang can be observed directly. Both Yin and Yang are intertwined forces always occurring in pairs, rather than being isolated forces independent from each other.

5.1.2. Yin and Yang Spirals

Yin and Yang are often depicted as spirals. In this subsection, I demonstrate that the fixed point equations given above, and in particular the fixed point equation of the net force, are spirals too. In order to do so, I will first introduce the general definition of the logarithmic spiral before I then illustrate the similarity to the famous Yin/Yang symbol.

A logarithmic spiral is a spiral curve that plays an important role in nature, where it occurs in all different kinds of objects and processes, such as mollusc shells, hurricanes, galaxies, and many more [1]. In polar coordinates (r, θ), a logarithmic spiral is defined as

$$r = ae^{b\theta} \tag{14}$$

Parameter a is a scale factor determining the size of the spiral, while parameter b controls the direction and tightness of the wrapping. The distances between the turnings of a logarithmic spiral increase. This distinguishes the logarithmic spiral from the Archimedian spiral, which features constant distances between turnings. Figure 2 shows a typical example of a logarithmic spiral. Resolving Eq. 14 for θ provides the class of logarithmic spirals looking similar to the fixed point equations given above:

$$\theta = \frac{1}{b}\, ln\left(\frac{r}{a}\right) \tag{15}$$

Using this form of logarithmic spirals, the polar coordinates (r, θ) follow directly from Eq.12, which defines the net force in Figure 1. The polar coordinates of its mirrored variant follow accordingly. Figure 3 shows the spirals for both forces plotted in a Cartesian coordinate system. For the sake of easier illustration, it actually shows the negative forces. We can see that both spirals are, of course, symmetrical and their turnings approach the unit circle. Figure 4 depicts the well-known black and white symbol of Yin and Yang. The dots of different color in the area of each spiral symbolize the fact that each force bears the seed of its counterpart within itself. A comparison of the Yin/Yang symbol of Figure 4 with the spirals in Figure 3 shows the strong similarities between both figures. A simple mirror operation would transform the spirals in Figure 3 into the Yin/Yang symbol.

Figure 4. Yin and Yang.

The next section is going to present another interesting discovery, namely that the net force equals one of its constituent forces in points determined by the golden ratio. This gives the golden ratio a new information-theoretical meaning.

5.2. Golden Ratio

The golden ratio is an irrational number, or rather a pair of two numbers, describing the proportion of two quantities. Expressed in words, two quantities are in the golden ratio to each other, if the whole is to the larger part as the larger part is to the smaller part. The whole in this case is simply the sum of both parts. Figure 5 shows an example of a line divided into two segments that are in the golden ratio. Historically, the golden ratio was already studied by ancient mathematicians. It plays an important role in different fields like geometry, biology, physics, and others. Many artists and designers deliberately or un-

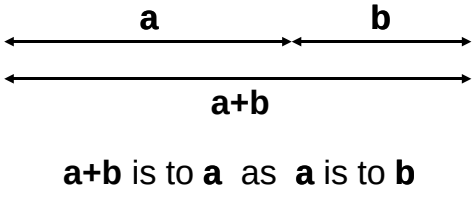

Figure 5. Golden Ratio.

consciously make use of it because it seems that artwork based on the golden ratio has an esthetic appeal, and features some kind of natural symmetry. Despite the fact that the golden mean is of paramount importance to so many fields, I think it is fair to say that we still do not have a full, or rather correct, understanding of its true meaning in science. The reader interested in the golden mean can find more information in [2, 9].

Mathematically, the golden mean can be derived from the following equation, which describes the colloquial description given above in mathematical terms.

$$\frac{a+b}{a} = \frac{a}{b} \tag{16}$$

Accordingly, the golden mean, which is typically denoted by the Greek letter φ, is then given by the ratio of a and b, i.e. $\varphi = \frac{a}{b}$. Using the relationship in Eq. 16, the golden ratio φ can be resolved into two possible values:

$$\varphi = \frac{1+\sqrt{5}}{2} \vee \frac{1-\sqrt{5}}{2} \tag{17}$$

$$\Longrightarrow \quad \varphi \approx 1.618 \vee -0.618 \tag{18}$$

Usually, the positive value (≈ 1.618) is identified with φ.

Interestingly, the golden mean also plays a role in the information-theoretical approach outline above. If we take another look at the net force in Eq. 12 and the definition of the counter-confidence (Force B) in Eq. 9, we see that both become equal when $E = p(K)$ and the performance $p(K)$ satisfies the following relationship:

$$p(K) = \frac{1-p(K)}{p(K)}$$

$$\Longleftrightarrow \quad p(K) \approx -1.618 \vee 0.618 \tag{19}$$

The two possible performance values satisfying this equation are exactly the negative values of the golden ratio given above. Confining ourselves to the positive value, we can say that counter-confidence and net force are the same for a performance of about 0.618.

6. Summary

I presented an information-theoretical approach for the normalization of confidence values. Starting out from the postulate that confidence is basically information, I derived a fixed point equation that normalizes confidence values according to their performance in a given application domain. The performance function turned out to be an exponential distribution following directly from the fixed point equation. My idea is to learn the informational confidence values by estimating their performance and inserting the estimates into the fixed point equation. A closer look at this equation revealed that it has a counterpart playing the role of a counter-confidence. The net effect of both confidences is a spiral looking similar to the well-known Yin/Yang symbol. Furthermore, I showed that the golden ratio defines points where the compound effect of confidence and counter-confidence is determined by a single force.

References

[1] T. Cook, *The Curves of Life*, Dover Publications, 1979.

[2] H. Huntley, *The Divine Proportion*, Dover Publications, 1970.

[3] S. Jaeger, "Informational Classifier Fusion", *Proc. of the 17th Int. Conf. on Pattern Recognition*, 2004, pp 216–219, Cambridge, UK.

[4] S. Jaeger, "Using Informational Confidence Values for Classifier Combination: An Experiment with Combined On-Line/Off-Line Japanese Character Recognition", *Proc. of the 9th Int. Workshop on Frontiers in Handwriting Recognition*, 2004, pp 87–92, Tokyo, Japan.

[5] S. Jaeger, "Entropy, Perception, and Relativity", Technical Report LAMP-TR-131/CS-TR-4799/UMIACS-TR-2006-20/CAR-TR-1012, University of Maryland, College Park, April 2006.

[6] S. Jaeger, H. Ma and D. Doermann, "Identifying Script on Word-Level with Informational Confidence", *Int. Conf. on Document Analysis and Recognition (ICDAR)*, 2005, pp 416–420, Seoul, Korea.

[7] S. Jaeger and M. Nakagawa, "Two On-Line Japanese Character Databases in Unipen Format", *6th International Conference on Document Analysis and Recognition (ICDAR)*, 2001, pp 566–570, Seattle.

[8] J. Kittler, M. Hatef, R. Duin and J. Matas, "On Combining Classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[9] M. Livio, *The Golden Ratio*, Random House, Inc., 2002.

[10] M. Nakagawa, K. Akiyama, L. Tu, A. Homma and T. Higashiyama, "Robust and Highly Customizable Recognition of On-Line Handwritten Japanese Characters", *Proc. of the 13th International Conference on Pattern Recognition*, 1996, volume III, pp 269–273, Vienna, Austria.

[11] M. Nakagawa, T. Higashiyama, Y. Yamanaka, S. Sawada, L. Higashigawa and K. Akiyama, "On-Line Handwritten Character Pattern Database Sampled in a Sequence of Sentences without Any Writing Instructions", *Fourth International Conference on Document Analysis and Recognition (ICDAR)*, 1997, pp 376–381, Ulm, Germany.

[12] C. E. Shannon, "A Mathematical Theory of Communication", *Bell System Tech. J.*, 27(623-656):379–423, 1948.

[13] O. Velek, S. Jaeger and M. Nakagawa, "Accumulated-Recognition-Rate Normalization for Combining Multiple On/Off-Line Japanese Character Classifiers Tested on a Large Database", *4th International Workshop on Multiple Classifier Systems (MCS)*, 2003, pp 196–205, Guildford, UK. Lecture Notes in Computer Science, Springer-Verlag.

[14] O. Velek, C.-L. Liu, S. Jaeger and M. Nakagawa, "An Improved Approach to Generating Realistic Kanji Character Images from On-Line Characters and its Benefit to Off-Line Recognition Performance", *16th International Conference on Pattern Recognition (ICPR)*, 2002, volume 1, pp 588–591, Quebec.