18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Correlation Coefficient Based Template Matching: Accounting for Uncertainty in Selecting the Winner

Nicholas J. Napoli
Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22904, USA
Email: njn5fg@virginia.edu

Laura E. Barnes
Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22904, USA
Email: lbarnes@virginia.edu

Kamal Premaratne
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33146, USA
Email: kamal@miami.edu

***Abstract*—The problem of selecting a template that matches a given candidate signal is applicable across a wide variety of domains. Using the correlation coefficient as the avenue for selecting the winning template is perhaps the most common technique. The challenge lies in selecting the winning template when there is no clear separation between the correlation coefficient values of the winning template and the others. In this paper, we present a simple Dempster-Shafer (DS) theoretic model that enables one to capture the uncertainty regarding the winner selection in correlation coefficient based template matching. The DS theoretic framework provides an avenue to develop the model with few resources and little to no prior knowledge. We validate the model using several numerical examples and a numerical character recognition application where the evidence provided by several sets of templates are combined using a DS theoretic fusion strategy to arrive at a better decision.**

I. Introduction

The correlation coefficient is widely used as a measure that captures the strength of the linear relationship between two variables. When applied to random variables, the correlation coefficient between the random variables $\mathbf{X}$ and $\mathbf{Y}$ usually takes the form [1]

$$r_{\mathbf{X},\mathbf{Y}} = \frac{\text{cov}(\mathbf{X},\mathbf{Y})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} = \frac{E[(\mathbf{X}-\overline{\mathbf{X}})(\mathbf{Y}-\overline{\mathbf{Y}})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}, \qquad (1)$$

where $E[\boldsymbol{.}]$ denotes the expectation, $\overline{\mathbf{X}} = E[\mathbf{X}]$ and $\overline{\mathbf{Y}} = E[\mathbf{Y}]$ denote the means, and $\sigma_{\mathbf{X}} = E[\mathbf{X}^2] - E^2[\mathbf{X}]$ and $\sigma_{\mathbf{Y}} = E[\mathbf{X}^2] - E^2[\mathbf{X}]$ denote the standard deviations. The correlation coefficient takes values in $[-1,+1]$: values closer to $+1$ indicate a strong positive relationship; values closer to -1 indicate a strong negative relationship; and values closer to 0 indicate weak a or non-existent relationship.

The correlation coefficient is also widely employed to gauge the 'closeness' or similarity between two data vectors, in which case it usually takes the form [2]

$$r_{\mathbf{x},\mathbf{y}} = \frac{\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})(y_i-\overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{\mathbf{x}})^2\sum_{j=1}^{n}(y_j-\overline{\mathbf{y}})^2}}, \qquad (2)$$

where $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}x_i$ and $\overline{\mathbf{y}} = \frac{1}{n}\sum_{j=1}^{n}y_j$ denote the 'sample' means of the real-valued time series data vectors $\mathbf{x} = [x_1,\ldots,x_N]^T$ and $\mathbf{y} = [y_1,\ldots,y_N]^T$, respectively.

Prior Work. In practice, one is often interested in determining how well a *template* signal vector belonging to a set of templates matches a given *candidate* signal. The template which yields the highest correlation coefficient with the candidate can be considered the 'winner' thus implying that the candidate indicates the presence of the winning template [3]–[7]. Over the years, this simple and popular method based on correlation coefficients has also been employed when there is a *set* of candidate signals to be considered [8]–[11].

Challenges. However, this technique of selecting the winning template may not be satisfactory when multiple templates yield high correlation coefficients with no significant separation of values between the correlation coefficient of one template and the others.

For instance, two templates which yield correlation coefficient values that are close to each other creates an uncertainty regarding which template should be declared the winner. Approaches that attempt to deal with such situations tend to employ various weighting and voting strategies to select the winning candidate [8], [11], [12]. Another source of difficulty in selecting the winner is the presence of noise. For example, within the context of reconstructing physiological signals, the work in [12] employs additional leads to mitigate problems caused by sources corrupted from noise. Fuzzy reasoning has been suggested as a way to improve the detection process in such situations [9].

These methods however do not provide a satisfactory solution for capturing the uncertainty associated with assigning a template to the candidate signal. To account for noisy signals, multiple templates, and multiple candidates, researchers have also embraced the use of more elaborate Kalman filtering and other statistical approaches [13].

Contributions. How should we handle such uncertainties in selecting the winning template? While probability theory is perhaps the most widely used approach for representing and handling imperfect information, probabilistic methods usually call for a priori assumptions regarding the underlying distributions and priors for handling data uncertainties. On the other hand, alternate imprecise probability formalisms, such as the Demspter-Shafer (DS) belief theoretic approach, provide ways to represent and deal with data uncertainties while requiring only little more information than voting and set intersection techniques [14]–[17]. DS theoretic (DST) methods can represent a wider variety of data imperfections in a more intuitive manner [15], [16]; they are more robust to modeling errors [18], [19], and when compared to what alternate frameworks provide, the DST belief and plausibility measures enable a decision to be made with a better understanding of the associated uncertainties [20]. When there is no uncertainty regarding the underlying distribution, these DST measures equal probability thus allowing seamless integration of DST methods with probabilistic methods [14]–[16], [20], [21]. This is a unique feature of the DST framework.

In this paper, we present a simple DST model which can be used to capture the uncertainty associated with assigning a template to the candidate signal. This model, which is only slightly more complicated than a probabilistic model, is based on the correlation coefficients between the candidate and the templates, and it allows us to view the candidate as an evidence source which indicates the 'presence' or 'absence' of each template. One may then utilize a DST evidence fusion strategy to combine multiple evidence sources, where each source provides evidence towards the 'presence' or 'absence' of each template, to make a decision regarding the winning template. Moreover, the availability of DST measures of belief and plausibility provide valuable information regarding the confidence one can place on this decision.

II. Basic Notions of DS Theory

The *frame of discernment (FoD)* Θ refers to the set of mutually exclusive and exhaustive propositions of interest. We take the FoD to be $\Theta = \{\theta_1, \ldots, \theta_n\}$, i.e., Θ is finite and composed of n *singleton* propositions. The power set of Θ is denoted by 2^{Θ}. A *basic probability assignment (BPA),* otherwise referred to as a *mass function,* is a function $m : 2^{\Theta} \to [0, 1]$ such that

$$m(\emptyset) = 0; \quad \sum_{A \subseteq \Theta} m(A) = 1. \tag{3}$$

Contextual considerations (e.g., accuracy, source reliability, source conflicts, etc.) all play a role in determining the mass to be allocated to a given proposition [22]. A proposition which has been allocated a non-zero mass is referred to as a *focal element.* The *core* $\mathfrak{F}$ refers to the set of focal elements and the *body of evidence (BoE)* $\mathcal{E}$ refers to the triplet $\{\Theta, \mathfrak{F}, m(\centerdot)\}$.

By allowing the allocation of masses directly to non-singleton or *composite* propositions, DS theory provides an avenue to capture *ignorance* and *uncertainty.* While $m(A)$ measures the support that is directly assigned to proposition $A \subseteq \Theta$ *only,* the *belief* $Bl(A)$ represents the total support that can move into A without any ambiguity; *plausibility* $Pl(A)$ represents the extent to which one finds A plausible. So,

$$Bl(A) = \sum_{B \subseteq A} m(B); \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{4}$$

These DST belief and plausibility measures are closely related to the inner and outer measures of a non-measurable event $A \subseteq \Theta$ with respect to probability mass functions (p.m.f.s) defined on Θ. Furthermore, when focal elements are constituted of singletons only, the mass, belief and plausibility all reduce to a p.m.f. The *uncertainty interval* $Un(A) = [Bl(A), Pl(A)]$ provides information regarding the support for $A \subseteq \Theta$.

Dempster's combination rule (DCR) allows one to combine or fuse evidence represented as DST models [14]:

$$m(A) = \frac{\displaystyle\sum_{B \cap C = A} m_1(B)\, m_2(C)}{\displaystyle\sum_{B \cap C = \emptyset} m_1(B)\, m_2(C)}, \tag{5}$$

where $\mathcal{E}_1 = \{\Theta, \mathfrak{F}_1, m_1(\centerdot)\}$ and $\mathcal{E}_2 = \{\Theta, \mathfrak{F}_2, m_2(\centerdot)\}$ are the BoE being fused to generate the fused BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\centerdot)\}$. The fused mass and BoE generated by the DCR are usually denoted by $m = m_1 \oplus m_2$ and $\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2$, respectively. The DCR is commutative and associative, thus allowing one to fuse multiple sources of evidence with ease.

III. Proposed DST Model

Consider the correlation vector corresponding to a single candidate vector and a set of N templates $\{T_1, \ldots, T_N\}$:

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_N \end{bmatrix}^T, \tag{6}$$

where $v_i \geq 0$, $i = 1, \ldots, N$, denotes the positive correlation coefficient between the candidate vector and the template T_i. In what appears below, we will deal with only non-negative values of correlation coefficients. In most applications, all the information regarding the identity of the template is captured by the absolute value of the correlation coefficient. In some application contexts, a negative correlation coefficient usually implies the absence of the corresponding template and treating such coefficients as having a value of zero would force the corresponding DST mass to be zero as well. On the other hand, if negative correlation coefficients provide additional information regarding the identity of the template, one can easily modify the algorithm we present below to account for negative coefficients values.

In constructing a DST model to represent the correlation coefficient vector $\mathbf{v}$ in (6), we must ensure that the model captures the potential conflicts among templates that are 'competing' for a match with the candidate signal reasonably well. For example, a correlation vector with multiple entries having a value of $+1$ would indicate that multiple templates are perfect matches for the candidate.

Weighting Matrix. First, consider the following $(N \times N)$-sized matrix $\Delta\mathbf{V} = \{\Delta V_{ij}\}$:

$$\Delta\mathbf{V} = \{\Delta V_{ij}\}, \text{ where } \Delta V_{ij} = v_i - v_j. \tag{7}$$

Note that $\Delta V_{ij} \in [-1, +1]$, $\forall i, j = 1, \ldots, N$, can be considered the 'distance' between the correlation coefficients for templates T_i and T_j; the sum of the entries in the j-th column of $\Delta\mathbf{V}$ is the 'sum of distances' $\sum_{i=1}^{N} \Delta V_{ij} = \sum_{i=1}^{N} (v_i - v_j)$ from the correlation coefficient of template T_j.

However, the entries in the $\Delta\mathbf{V}$ matrix still lack information about the 'strength' of the correlations. For example, take

$\mathbf{v} = [0.4,\ 0.3,\ 0.7,\ 0.6]^T$. This yields $(v_1 - v_2) = (v_3 - v_4) = 0.1$, irrespective of the fact that the values v_3 and v_4 are significantly higher (thus more correlated with the corresponding two templates) than the values v_1 and v_2 (which are less correlated). To account for this, we utilize the following weighting strategy:

$$\Delta\mathbf{W} = \{\Delta W_{ij}\} = \Delta\mathbf{v} \cdot \mathbf{D_V}, \qquad (8)$$

where $\mathbf{D_v} = \text{diag}\{v_1, \ldots, v_N\}$ denotes the diagonal matrix with the diagonal entries $\{v_1, \ldots, v_N\}$. Column j of $\Delta\mathbf{W}$ is simply column j of $\Delta\mathbf{V}$ weighted by the correlation coefficient v_j corresponding to template T_j. Taking the same example $\mathbf{v} = [0.4,\ 0.3,\ 0.7,\ 0.6]^T$, we get $v_1\Delta V_{12} = (0.4)(0.4 - 0.3) = 0.04$ and $v_3\Delta V_{34} = (0.7)(0.7 - 0.6) = 0.07$, which now accounts for the strength of the correlation coefficients.

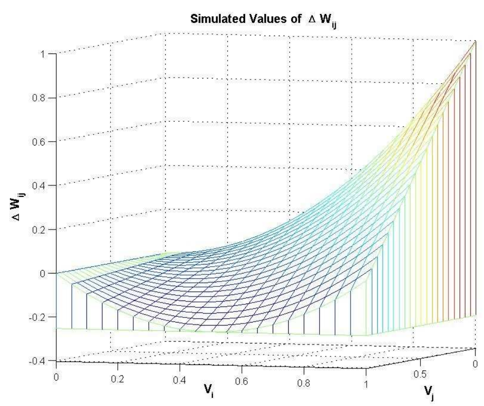

Fig. 1. Plot of $\Delta W_{ij} = v_i\,\Delta V_{ij}$ versus (v_i, v_j), for $v_i, v_j \in [0, 1]$.

Note that each column of $\Delta\mathbf{W}$ evaluates the correlation coefficient corresponding to a specific template against the correlation coefficients of the entire set of templates. For example, column j of $\Delta\mathbf{W}$ captures the strength of the correlation coefficient v_j and the distance between v_j and all other correlation coefficients $(v_i - v_j)$. In essence, the matrix $\Delta\mathbf{W} = \{\Delta W_{ij}\}$ accounts for both the *strength* of the correlation to each template and also the *distance* between the correlations of pairs of templates. When one or both aspects are low, the corresponding entries in $\Delta\mathbf{W}$ will take on lower values. This feature of $\Delta\mathbf{W}$ is important when attempting to fit our DST model. Fig. 1 demonstrates the mapping $\Delta W_{ij} = v_i\,\Delta V_{ij}$ versus (v_i, v_j). Note that the maximum and minimum achievable values of ΔW_{ij} are $+1$ and $-1/4$, respectively; and these values are achieved when $(v_i, v_j) = (1, 0)$ and $(v_i, v_j) = (1/2, 1)$, respectively. Note how ΔW_{ij} rewards higher strength of correlation coefficients and larger distances of pairs of templates.

Column Weights Vector. The sum of column j of $\Delta\mathbf{W}$ informs us about how well the template T_j matches the candidate and how different the correlation coefficient associated with T_j is when compared to the other correlation coefficients. We refer to the vector created by the sum of the entries in each column of the weighting matrix $\Delta\mathbf{W}$ as its *column weight vector* $\mathbf{c}$*:*

$$\mathbf{c} = \begin{bmatrix} c_1 & \cdots & c_N \end{bmatrix} = \mathbf{1}_{1\times N}\Delta\mathbf{W}, \qquad (9)$$

where $\mathbf{1}_{1\times N} = [1,\ 1, \ldots, 1]$, i.e., the $(1 \times N)$-sized row vector with all entries taking the value 1. Note that the diagonal entries of $\Delta\mathbf{W}$ are always 0. Hence, $c_i \in [-(N-1)/4, (N-1)]$, $\forall\ i \in \overline{1, N}$. These column weights allow us to identify the rival templates that are competing to be a match for the candidate. Let us take some examples to demonstrate this.

Mass Measure Vector. Non-positive entries of $\mathbf{c}$ indicate templates that cannot compete for being a match to the candidate vector. So, such entries are replaced with values of 0 to get the *mass measure vector* $\mathbf{h}$*:*

$$\mathbf{h} = [h_1,\ h_2 \cdots,\ h_N], \text{ where } h_i = \frac{c_i + |c_i|}{2}. \qquad (10)$$

Clearly, $h_i \in [0, (N-1)], \forall i \in \overline{1, N}$. Note that, $\mathbf{h}$ is identical to $\mathbf{c}$, except that it substitutes 0 for all the non-positive elements of $\mathbf{c}$. We will later use the entries of $\mathbf{h}$ to generate the DST masses associated with the templates T_j. For convenience, we use the notation

$$S_\mathbf{v} = \sum_{i=1}^{N} v_i; \quad S_\mathbf{h} = \sum_{i=1}^{N} h_i. \qquad (11)$$

As it turns out, $S_\mathbf{h}$ denotes the total of DST masses that will be assigned to the singletons $\{T_i\}$.

TABLE I
SOME EXAMPLES

	$\mathbf{v}$	$\mathbf{c}$	$\mathbf{h}$
(1)	$[1,\ 0,\ 0,\ 0]^T$	$[3,\ 0,\ 0,\ 0]$	$[3,\ 0,\ 0,\ 0]$
(2)	$[1,\ 1,\ 0,\ 0]^T$	$[2,\ 2,\ 0,\ 0]$	$[2,\ 2,\ 0,\ 0]$
(3)	$[1,\ 1,\ 0.9,\ 0]^T$	$[1.10,\ 1.10,\ 0.63,\ 0]$	$[1.10,\ 1.10,\ 0.63,\ 0]$
(4)	$[1,\ 1,\ 0.9,\ 0.1]^T$	$[1,\ 1,\ 0.54,\ -0.26]$	$[1,\ 1,\ 0.54,\ 0]$

Examples. Table I show some examples of how $\mathbf{h}$ can be used to capture those templates that can be a match for the candidate. Note the following:

Example (1): Template T_1 is the only match, and $\mathbf{h}$ puts the maximum weight on T_1.

Example (2): Templates T_1 and T_2 are both competing for being a match, and $\mathbf{h}$ puts equal weights to T_1 and T_2. This weight is however less than what T_1 gets in Example (1) because of the presence of potentially two matching templates.

Example (3): Templates T_1-T_3 are competing for being a match, with template T_3 being the slightly weaker, and $\mathbf{h}$ distributes its weights among T_1-T_3.

Example (4): All templates have positive correlation coefficients, but only T_1-T_3 are legitimate potential matches. Note how $\mathbf{h}$ discards T_4, but puts less values for T_1-T_3 because $c_4 < 0$. ■

'Extreme' Case. Examples 1-2 above are instances of an 'extreme' situation when the correlation coefficients of all the templates take on values 1 or 0 only, i.e., $v_i \in \{0, 1\}$, $\forall i \in \overline{1, N}$. We refer to this case as the *extreme case.* For this case,

without loss of generality, we may assume that the templates yielding a perfect match appear as the top elements of $\mathbf{v}$, i.e.,

$$\mathbf{v} = [\overbrace{1, \ldots, 1}^{P \text{ of } 1\text{s}}, \overbrace{0, \ldots, 0}^{(N-P) \text{ of } 0\text{s}}]^T. \tag{12}$$

This yields

$$\Delta\mathbf{W} = \begin{bmatrix} \emptyset_{P\times P} & \emptyset_{P\times(N-P)} \\ \mathbf{1}_{(N-P)\times P} & \emptyset_{(N-P)\times(N-P)} \end{bmatrix}$$
$$\mathbf{c} = \mathbf{h} = [\overbrace{(N-P), \ldots, (N-P)}^{P \text{ of } (N-P)\text{s}}, 0, \ldots, 0]. \tag{13}$$

Here $\emptyset_{K\times L}$ and $\mathbf{1}_{K\times L}$ denotes the $(K \times L)$-sized matrices with entries 0 and 1, respectively. Note that, $S_{\mathbf{h}} = P(N-P)$.

Focal Elements. The FoD of the proposed DST model is the set of N templates, i.e., $\Theta = \{T_1, \ldots, T_N\}$. We use the column weights vector $\mathbf{c}$ to identify the focal elements of our DST model. First, we make the following observations:

(a) When $v_i \to 1$ and $v_j \to 0, \forall j \neq i$, the column weights $c_i \to (N-1)$ and $c_j \to 0, \forall j \neq i$, i.e., the uncertainty regarding the template T_i being a match decreases (see Example 1 above). Conversely, the higher the difference between $(N-1)$ and a column weight, the less likely the corresponding template is the correct match. When the column weight is zero or negative, we become more certain that the corresponding template is not a match.

(b) When a template is unable to compete against the other templates, the corresponding column weight becomes zero or negative (see Example 4 above). So, in our model, we neglect the templates corresponding to non-positive values of c_i, thus preventing them from becoming focal elements in our DST model. This strategy restricts the domain of candidate templates that are potential focal elements thus avoiding having to assign DST masses to 'weak' candidates.

The number of non-zero entries of $\mathbf{h}$ (or equivalently, the number of positive entries of $\mathbf{c}$) determine the number P of singleton focal elements of our DST model:

$$P = \sum_{i=1}^{N} t_i, \text{ where } t_i = \begin{cases} 1, & \textit{if } h_i > 0; \\ 0, & \textit{if } h_i = 0. \end{cases} \tag{14}$$

DST Mass Allocation. We will develop the model first for the extreme case, and then extend it to the more general case. The focal elements of the proposed DST model are

$$\mathfrak{F} = \{T_1, T_2, \ldots, T_P, \Theta\}. \tag{15}$$

We will assume that $P \in \overline{1, N}$; $P = 0$ case implies complete lack of information, and we trivially assign the DST model $m(\Theta) = 1$.

Extreme Case.

a) Masses Allocated to Singletons $m(T_i)$: From the examples in Table I, we notice how the values h_i appear to weigh templates according to how well they match the candidate. With this observation in mind, we propose

$$m(T_i) \propto h_i \implies m(T_i) = \frac{h_i}{K} = \frac{N-P}{K}, \tag{16}$$

where $K > 0$ is the proportionality constant.

b) Mass Allocated to Complete Ignorance $m(\Theta)$: Since the masses allocated to all the focal elements must add to 1, we must have

$$m(\Theta) = 1 - \sum_{i=1}^{P} m(T_i) = 1 - \frac{S_{\mathbf{h}}}{K} = 1 - \frac{P(N-P)}{K}. \tag{17}$$

On the other hand, in the extreme case, consider (12) and (13) where P templates are yielding perfect matches. Clearly, when $P = 1$, template T_1 provides a perfect match while the others offer 0 correlation. We should then allocate no mass for Θ, i.e., $m(\Theta) = 0$. On the other hand, when $P = N$, all the templates exhibit equally 'perfect' matches, and no decision can be made in favor of any template. We should then allocate $m(\Theta) = 1$. Using a linear relationship, for the extreme case, we then use

$$m(\Theta) = \frac{P-1}{N-1}. \tag{18}$$

Compare (17) and (18) to get

$$K = (N-1)P. \tag{19}$$

Thus we arrive at the following DST model:

$$m(A) = \begin{cases} \dfrac{N-P}{P(N-1)}, & \text{for } A = T_i,\ i \in \overline{1, P}; \\[2ex] \dfrac{P-1}{N-1}, & \text{for } A = \Theta; \\[2ex] 0, & \text{otherwise.} \end{cases} \tag{20}$$

General Case.

c) Masses Allocated to Singletons $m(T_i)$: As we did for the extreme case, we again propose

$$m(T_i) \propto h_i \implies m(T_i) = \frac{h_i}{K}, \tag{21}$$

where $K > 0$ is the proportionality constant.

d) Mass Allocated to Complete Ignorance $m(\Theta)$: Again, as before, we must have

$$m(\Theta) = 1 - \sum_{i=1}^{P} m(T_i) = 1 - \frac{S_{\mathbf{h}}}{K}. \tag{22}$$

In the extreme case, we know that $S_{\mathbf{h}} = P(N-P)$. This corresponds to the case when P templates have perfect matches, viz., $v_i = 1$ and $c_i = h_i = N-P$, for $i \in \overline{1, P}$ (see (12) and (13)). In the general case however, when one or more of the P templates have correlation coefficients that are less than 1, as we show in Appendix A, $S_{\mathbf{h}} < P(N-P)$. In other words, for a given value of P, the extreme case (where $v_i = 1$, $i \in \overline{1, P}$) yields the maximum value for $S_{\mathbf{h}}$. So,

$$m(\Theta) = 1 - \frac{S_{\mathbf{h}}}{K} > 1 - \frac{P(N-P)}{K}. \tag{23}$$

The amount of the increase in $m(\Theta)$ from the extreme case should be $[P(N-P) - S_{\mathbf{h}}]/K$. Thus we would get

$$m(\Theta) = \frac{P-1}{N-1} + \frac{P(N-P) - S_{\mathbf{h}}}{K}. \tag{24}$$

Since the masses should add to 1, considering (21) and (24), we get

$$\frac{S_{\mathbf{h}}}{K} + \frac{P-1}{N-1} + \frac{P(N-P)-S_{\mathbf{h}}}{K} = 1. \quad (25)$$

This yields

$$K = P(N-1). \quad (26)$$

Thus, we arrive at the following DST model:

$$m(A) = \begin{cases} \dfrac{h_i}{P(N-1)}, & \text{for } A = T_i,\ i \in \overline{1,P}; \\ 1 - \dfrac{S_{\mathbf{h}}}{P(N-1)}, & \text{for } A = \Theta; \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

Noting that $h_i = N - P$, $t \in \overline{1,P}$, for the extreme case, we note that the DST model in (27) is valid for the extreme case as well. Finally, we note that, for $i \in \overline{1,P}$,

$$Bl(T_i) = \frac{h_i}{P(N-1)};\ \ Pl(T_i) = 1 - \frac{\sum_{j \neq i} h_j}{P(N-1)}. \quad (28)$$

IV. Model Validation

In this section, we provide some examples to demonstrate the validity of the proposed DST model. All examples use $N = 4$, and we will use $\mathbf{m}$ to denote the vector of mass assignments, i.e.,

$$\mathbf{m} = [m(T_1),\ m(T_2),\ m(T_3),\ m(T_4),\ m(\Theta)]^T. \quad (29)$$

Some General Observations.

Example 1: $\mathbf{v} = [y,\ y,\ y,\ y]^T$. This example demonstrates the case when all the correlation coefficients are identical. In this case, we get

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = \mathbf{h} = [0,\ 0,\ 0,\ 0].$$

Thus, $P = 0$, i.e., we have no singleton focal elements. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0,\ 0,\ 1]^T.$$

Note that, $Un(T_i) = [0, 1]$, $i \in \overline{1,4}$, which shows the complete lack of evidence to make any decision. ■

Example 2: $\mathbf{v} = [y,\ y,\ y,\ z]^T$ *with* $z > y$. Here, all the correlation coefficients are identical, except one which has a higher coefficient. In this case, we get

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & z(z-y) \\ 0 & 0 & 0 & z(z-y) \\ 0 & 0 & 0 & z(z-y) \\ y(y-z) & y(y-z) & y(y-z) & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [y(y-z),\ y(y-z),\ y(y-z),\ 3z(z-y)];$$
$$\mathbf{h} = [0,\ 0,\ 0,\ 3z(z-y)].$$

Thus, $P = 1$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0,\ 3z(z-y)/3,\ 1 - 3z(z-y)/3]^T.$$

Note that the mass assignments for T_4 and Θ are dependent on the distance $(z-y)$. As $(z-y)$ increases, the $m(\Theta)$ decreases and $m(T_4)$ increases. The maximum distance achievable is when $z = 1$ and $y = 0$, yielding $m(T_4) = 1$. However, as $(z-y)$ decreases, the $m(\Theta)$ increases and $m(T_4)$ decreases. The minimum achievable distance occurs when $z = y$. As we approach this minimum distance, we converge to the scenario of maximum conflict where $m(\Theta) = 1$. Also note that $Un(T_4) = [3z(z-y)/3, 1]$. ■

Example 3: $\mathbf{v} = [y,\ y,\ z,\ z]^T$ *with* $z > y$. Here, half the correlation coefficients are identical; the other half also has identical coefficients but with a higher value. In this case, we get

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0 & z(z-y) & z(z-y) \\ 0 & 0 & z(z-y) & z(z-y) \\ y(y-z) & y(y-z) & 0 & 0 \\ y(y-z) & y(y-z) & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [2y(y-z),\ 2y(y-z),\ 2z(z-y),\ 2z(z-y)];$$
$$\mathbf{h} = [0,\ 0,\ 2z(z-y),\ 2z(z-y)].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 2z(z-y)/6,\ 2z(z-y)/6,\ 1 - 4z(z-y)/6]^T.$$

Compare this DST model, Example 3, with that of Example 2 above. While the masses are still dependent on the distance $(z-y)$, the values of the masses for T_3 and T_4 have been reduced. This lowering of the value can be attributed to having additional competing templates. Furthermore, notice how $m(\Theta)$ has been increased; it will never be reduced to 0, even if a maximum distance between z and y is reached. The range of values for $m(\Theta)$ is $[0.33, 1]$ because of the amount of focal elements present in the vector. Also note that $Un(T_3) = Un(T_4) = [2z(z-y)/6, 1 - 2z(z-y)/6]$. ■

Numerical Examples.

Example 1: $\mathbf{v} = [0.01,\ 0.03,\ 0.12,\ 0.98]^T$. Here, the template T_4 is the only strong match. Note that we have

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & -0.125 & -0.232 & -0.204 \\ 0.147 & 0 & -0.172 & -0.159 \\ 0.568 & 0.357 & 0 & -0.030 \\ 0.666 & 0.440 & 0.040 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [-0.0110,\ -0.0306,\ -0.0792,\ 2.7244];$$
$$\mathbf{h} = [0,\ 0,\ 0,\ 2.7244].$$

Thus, $P = 1$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0,\ 0.9081,\ 0.0919]^T.$$

Also note that $Un(T_4) = [0.9081, 1]$. ■

Example 2: $\mathbf{v} = [0.98,\ 0.83,\ 0.40,\ 0.30]^T$. Here, the templates T_1 and T_2 act as strong matches. Note that we have

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & -0.125 & -0.232 & -0.204 \\ 0.147 & 0 & -0.172 & -0.159 \\ 0.568 & 0.357 & 0 & -0.030 \\ 0.666 & 0.440 & 0.040 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [1.3818,\ 0.6723,\ -0.3640,\ -0.3930];$$
$$\mathbf{h} = [1.3818,\ 0.6723,\ 0,\ 0].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0.2303,\ 0.1120,\ 0,\ 0,\ 0.6577]^T.$$

Also note that

$$Un(T_1) = [0.2303, 0.8880];\ Un(T_2) = [0.1120, 0.7697].$$ ■

Example 3: $\mathbf{v} = [0.31,\ 0.32,\ 0.43,\ 0.44]^T$. Here, all the templates are weak. Note that

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0.0032 & 0.0516 & 0.0572 \\ -0.0031 & 0.0473 & 0.0528 & \\ -0.0372 & -0.0352 & 0 & 0.0044 \\ -0.0403 & -0.0384 & -0.0043 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [-0.0806,\ -0.0704,\ 0.0946,\ 0.1144];$$
$$\mathbf{h} = [0,\ 0,\ 0.0946,\ 0.1144].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0.0158,\ 0.0191,\ 0.9652]^T.$$

This illustrates that, with our proposed model, even with a weak set, masses are still assigned to the stronger templates, but with an increased value for $m(\Theta)$. Also note that

$$Un(T_1) = Un(T_2) = [0, 0.8652];$$
$$Un(T_3) = [0.0158, 0.9810];\ Un(T_4) = [0.0191, 0.9843].$$ ■

Example 4: $\mathbf{v} = [0.31,\ 0.32,\ 0.73,\ 0.74]^T$. We use this example to compare with Example 3 to demonstrate how uncertainty decreases if T_3 and T_4 become stronger candidates. We have

$$\Delta\mathbf{W} = \begin{bmatrix} 0 & 0.0032 & 0.3066 & 0.3182 \\ -0.0031 & 0 & 0.2993 & 0.3108 \\ -0.1302 & -0.1312 & 0 & 0.0074 \\ -0.1333 & -0.1344 & -0.0073 & 0 \end{bmatrix},$$

and

$$\mathbf{c} = [-0.2666,\ -0.2624,\ 0.5986,\ 0.6364];$$
$$\mathbf{h} = [0,\ 0,\ 0.5986,\ 0.6364].$$

Thus, $P = 2$. The corresponding DST model is

$$\mathbf{m} = [0,\ 0,\ 0.0998,\ 0.1061,\ 0.7942]^T.$$

Notice how T_3 and T_4 are being supported more, while the support for $m(\Theta)$ is being reduced. Also note that

$$Un(T_1) = Un(T_2) = [0, 0.7942];$$
$$Un(T_3) = [0.0998, 0.8940];\ Un(T_4) = [0.1061, 0.9003].$$ ■

V. EXPERIMENT

In this section, we demonstrate the applicability and robustness of the proposed model within the context of numerical character recognition within images.

Data Set. We used the data set in [23] for detecting numerical characters from computer fonts with 4 variations (combinations of italic, bold, and normal) and in 85 font types. Each font type provides a set of 10 images thus creating a total of 850 template images within the data base.

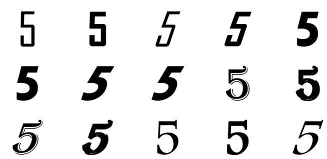

Fig. 2. Template T_5 in 15 different font types.

Formulation of Correlation Coefficients as Evidence. For our experiment, we used only 15 font types. With the templates $\{T_0, \ldots, T_9\}$ denoting the numerical characters $\{0, \ldots, 9\}$, respectively, we therefore used a set of 150 templates. Fig. 2 shows the 15 font types corresponding to the template T_5 (which corresponds to the numerical character "5"). The data base was then partitioned into two groups. The first partition was designated as the training set (i.e., evidence sources) and the other as the testing set. It is important to emphasize that the purpose of this experiment is to simply evaluate the DST model. We have undertaken no preprocessing of the images.

With each font giving a set of 10 templates, when correlated with an unknown image, we obtain the vector $\mathbf{v}$ in (6). We treat this vector corresponding to each font as one source of evidence. With the 15 font types, we therefore obtain 15 evidence sources and their corresponding 15 DST models. These sources are then fused using the DCR in (5). For the testing data set, unknown candidate images are randomly selected from the remaining 700 templates and then we introduce 'salt and pepper' noise to corrupt the image prior to generating the correlation coefficients. Several different noise density, d, levels of 'salt and pepper' noise were employed. Figs 3, 4, and 5 show three levels of d. The noise level applied to a candidate image was quantified via the peak signal-to-noise ratio (PSNR) [24].

Results and Discussion. No noise was added initially to determine the baseline accuracy which was 93.17% for the detection of 600 randomly selected samples using all 15 evidence sources. It typically required a minimum of 3 fusion

Fig. 3. $d = 0.55$

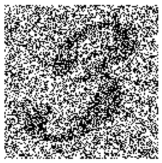

Fig. 4. $d = 0.75$

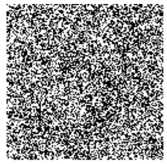

Fig. 5. $d = 0.90$

combinations to converge to the correct match, thus making it obvious that simply taking the highest correlation coefficient would be an inadequate approach. The confusion matrix in Table II demonstrates the difficultly in distinguishing a specific template from one another (in particular, for T_4 and T_1).

TABLE II
CONFUSION MATRIX: $d = 0$, ACCURACY = 93.17%

	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
T_0	55	0	0	0	0	0	0	0	2	0
T_1	0	49	0	0	15	0	0	0	0	0
T_2	0	0	71	0	0	0	0	0	0	0
T_3	0	0	0	54	0	0	0	0	0	0
T_4	0	5	0	0	53	0	0	0	0	0
T_5	0	0	0	0	0	58	0	0	0	0
T_6	0	0	0	0	2	9	40	0	4	0
T_7	0	1	0	0	0	0	0	69	0	0
T_8	0	0	0	0	0	0	0	0	52	0
T_9	3	0	0	0	0	0	0	0	0	58

When high distortion is introduced into the samples ($d = 0.9$, PSNR = 35.02), the overall accuracy is 76.7%. The corresponding confusion matrix appears in Table III, where an increase in false positives and false negatives across all templates is observed.

TABLE III
CONFUSION MATRIX: $d = 0.9$, ACCURACY = 76.7%

	T_0	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
T_0	42	0	0	0	2	0	2	3	3	0
T_1	0	43	1	0	7	0	0	0	0	1
T_2	1	0	55	4	0	0	0	11	0	0
T_3	0	1	0	45	0	8	0	0	11	1
T_4	0	5	0	0	67	2	0	1	0	1
T_5	5	1	0	1	0	52	5	0	1	0
T_6	8	0	0	1	4	12	33	0	6	2
T_7	0	3	2	0	1	0	0	46	0	0
T_8	3	0	0	1	0	1	3	0	42	2
T_9	11	5	0	0	2	0	1	4	0	34

Table II demonstrates that detection errors were mostly attributable to a sample that may not fit well with our 150 evidence sources. This suggests that we could possibly pick different sources of evidence that could better represent the database for improved detection. On the other hand, Table III demonstrates that a failure in detection is attributable to noise distortion.

The importance of our proposed DST model is demonstrated by how it captures the uncertainty as noise is introduced into the sample. As Table IV shows, when noise increases, we are unable to reduce the uncertainty even when multiple sources of evidence are fused. We are still able to detect the winner (by simply picking the template with the highest mass), but now, the DST model provides us invaluable information regarding our confidence (or lack thereof) in the match. This is an important and useful feature of our model.

TABLE IV
NOISE LEVELS VS UNCERTAINTY

$\mathbf{N_D}$	**.05**	**.25**	**.45**	**0.65**	**0.85**	**0.95**
$\mathbf{P_{SNR}}$	47.46	40.52	37.91	36.37	34.70	35.16
$\mathbf{\Theta}$	0.57	0.74	0.86	0.94	0.98	0.99

VI. CONCLUSION

We have proposed a simple DST model for capturing the uncertainty associated with allocating a template for a given candidate signal. This model can be especially useful in situations where there is no clear winning template in terms of the correlation coefficient values. Such a situation creates uncertainty as to the template to be declared the winner. The DST framework allows the model to be developed with little prior knowledge, and the DST uncertainty interval (constructed from the belief and plausibility values) provide valuable information regarding the confidence one can place on this decision. Moreover, DST evidence fusion strategies can be utilized to fuse evidence generated from different sources thus allowing the decision to be refined and improved. This is exactly the strategy that we followed in the experiment carried out in Section V.

We emphasize that the results in Section V are extracted simply by using the correlation coefficient as applied *directly* to the templates and the candidate. In practice, when one employs correlation coefficients for template matching, template and candidate signals are pre-filtered (e.g., light compensation, rotation, etc., for images). The results we give do not employ any pre-filtering, and the decisions can be significantly improved by employing such schemes.

The model proposed in this work is simple in the sense that it captures uncertainty via the assignment of a mass to the complete ambiguity (i.e., Θ). A better model would be to allow other non-singleton propositions to be focal elements. For example, if the correlation coefficients corresponding to the templates T_1 and T_2 are high and the coefficients for the others are low, a better model would generate a focal element from (T_1, T_2). The model in this paper employs Θ as the only non-singleton focal element.

APPENDIX A
MAXIMUM OF $S_{\mathbf{h}} = \sum_{i=1}^{N} h_i$

Consider the correlation coefficient vector $\mathbf{v}$ in (6), where $v_i \in [0, 1]$, $\forall i \in \overline{1, N}$. Generate the following matrix associated with $\mathbf{V}$:

$$\Delta\mathbf{W} = \begin{bmatrix} v_1\Delta v_{11} & v_2\Delta v_{21} & \dots & v_N\Delta v_{N1} \\ v_1\Delta v_{12} & v_2\Delta v_{22} & \dots & v_N\Delta v_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ v_1\Delta v_{1N} & v_2\Delta v_{2N} & \dots & v_N\Delta v_{NN} \end{bmatrix}, \quad (30)$$

where $\Delta v_{ij} = (v_i - v_j)$. Next, generate a column weights vector, $\mathbf{c}$, by summing the elements of each column of $\Delta\mathbf{W}$:

$$\mathbf{c} = \left[N\, v_1^2 - v_1 S_{\mathbf{v}},\ N\, v_2^2 - v_2 S_{\mathbf{v}},\ \ldots \right.$$
$$\left. N\, v_j^2 - v_j S_{\mathbf{v}},\ \ldots,\ N\, v_N^2 - v_N S_{\mathbf{v}}\right]. \quad (31)$$

Without loss of generality, let us assume that only the first P elements of $\mathbf{c}$ are positive which implies that

$$N\, v_i^2 - v_i S_{\mathbf{v}} \begin{cases} > 0, & \text{for } i \in \overline{1, P}; \\ \leq 0, & \text{for } i \in \overline{P+1, N}, \end{cases}$$
$$\iff v_i \begin{cases} > S_{\mathbf{v}}/N, & \text{for } i \in \overline{1, P}; \\ \leq S_{\mathbf{v}}/N, & \text{for } i \in \overline{P+1, N}. \end{cases} \quad (32)$$

The mass measure vector thus generated is

$$\mathbf{h} = \left[N\, v_1^2 - v_1 S_{\mathbf{v}},\ N\, v_2^2 - v_2 S_{\mathbf{v}},\ \ldots \right.$$
$$\left. \ldots,\ N\, v_P^2 - v_P S_{\mathbf{v}},\ 0,\ \ldots,\ 0\right]. \quad (33)$$

Then,

$$S_{\mathbf{h}} = \sum_{j=1}^{P} \left(N\, v_j^2 - v_j S_{\mathbf{v}}\right) = N \sum_{j=1}^{P} v_j^2 - S_{\mathbf{v}} \left(\sum_{j=1}^{P} v_j\right). \quad (34)$$

We now consider the following problem: how much can we increase $S_{\mathbf{h}}$ by changing v_j, $j \in \overline{1, P}$, while making sure that only the first P elements of $\mathbf{c}$ are positive? We note the following:

- For v_i, $i \in \overline{1, P}$, use (32) to get $\partial S_{\mathbf{h}}/\partial v_i = 2Nv_i - S_{\mathbf{v}} - \sum_{j=1}^{P} v_j > 0$.
- For v_i, $i \in \overline{P+1, N}$, $\partial S_{\mathbf{h}}/\partial v_i = -\sum_{j=1}^{P} v_j < 0$.

So, it is clear that the maximum of $S_{\mathbf{h}}$ is achieved when v_i, $i \in \overline{1, P}$ are increased to their maximum value (viz., 1) and when v_i, $i \in \overline{P+1, N}$, are decreased to their minimum value (viz., 0). In doing so, we do not violate the conditions in (32) and therefore we ensure that only the first P elements of $\mathbf{c}$ are positive. So, putting $v_i = 1$, $i \in \overline{1, P}$, and $v_i = 0$, $\imath \in \overline{P+1, N}$, in (34), we get

$$\max S_{\mathbf{h}} = NP - P^2 = P(N - P). \quad (35)$$

Acknowledgment

This work was supported in part by the U.S. National Science Foundation (NSF) via grant #1038257. The foundation of this work was developed at the University of Miami.

References

[1] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Person Prentice Hall, 2008.

[2] D. S. Moore, *The Basic Practice of Statistics*, 3rd ed. W.H Freeman and Company, 2004.

[3] S. S. Kothalkar and D. Manjusha, "Comparative study of image registration methods," *International Journal of Image Processing (IJIP)*, vol. 8, no. 3, pp. 125–147, 2014.

[4] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," in *Aerospace/Defense Sensing, Simulation, and Controls*. International Society for Optics and Photonics, 2001, pp. 95–102.

[5] S. Abbound and D. Sadeh, "The use of cross-correlation function for the alignment of ECG waveforms and rejection of extrasystoles," *Computer and Biomedical Research*, vol. 17, pp. 258–266, 1984.

[6] S. Dobbs, N. Schmitt, and H. Ozemek, "QRS detection by template matching using real-time correlation on a microcomputer," *Journal of Clinical Engineering*, vol. 9, pp. 197–212, 1984.

[7] V. X. Afonso, *Biomedical digital signal processing: C-language Examples and Laboratory Experiments for the IBM PC*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993, pp. 236–264.

[8] V. Krasteva and I. Jekova, "QRS template matching for recognition of ventricular ectopic beats," *Annals of Biomedical Engineering*, vol. 35, pp. 2065–2076, December 2007.

[9] F. J. Chin, Q. Fang, T. Zhang, and I. Cosic, "A fast critical arrhythmic ECG waveform identification method using cross-correlation and multiple template matching," *International Conference of the IEEE EMBS*, vol. 32, pp. 1922–1925, August 2010.

[10] T. Last, C. D. Nugent, and F. J. Owens, "Multi-component based cross correlation beat detection in electrocardiogram analysis," *Biomedical Engineering Online*, vol. 3, no. 26, pp. 1–14, July 2004.

[11] U. Mahbuba, H. Imtiaza, T. Roya, S. Rahmana, and A. R. Ahadb, "A template matching approach of one-shot-learning gesture recognition," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1780–1788, November 2013.

[12] G. Ganeshapillai and J. Guttag, "Real time reconstruction of quasiperiodic multi parameter physiological signals," *EURASIP Journal on Advances in Signal Processing*, vol. 173, no. 1, pp. 1–15, 2012.

[13] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiology Measurement*, vol. 29, no. 1, pp. 15–32, 2008.

[14] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.

[15] S. Blackman and R. Popoli, *Design and Analyis of Modern Tracking Systems*. Norwood, MA: Artech House, 1999.

[16] D. L. Hall and J. Llinas, Eds., *Handbook of Multisensor Data Fusion*. Boca Raton, FL: CRC Press, 2001.

[17] C. Thomas and N. Balakrishnan, "Modified evidence theory for performance enhancement of intrusion detection systems," in *Proc. International Conference on Information Fusion (ICIF)*, Cologne, Germany, Jun./Jul. 2008, pp. 1–8.

[18] P. Smets, "Practical uses of belief functions," in *Proc. Conference on Uncertainty in Artificial Intelligence (UAI)*, K. B. Laskey and H. Prade, Eds. San Francisco, CA: Morgan Kaufmann, 1999, pp. 612–621.

[19] A. Benavoli, L. Chisci, A. Farina, and B. Ristic, "Modelling uncertain implication rules in evidence theory," in *Proc. International Conference on Information Fusion (FUSION)*, Cologne, Germany, Jun./Jul. 2008, pp. 1–7.

[20] R. R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., *Advances in the Dempster-Shafer Theory of Evidence*. New York, NY: John Wiley and Sons, 1994.

[21] R. R. Yager and L. Liu, Eds., *Classic Works of the Dempster-Shafer Theory of Belief Functions*, ser. Studies in Fuzziness and Soft Computing. Springer-Verlag Berlin Heidelberg, 2008.

[22] K. Sentz and S. Ferson, "Combination of evidence in Dempster-Shafer theory," Sandia National Laboratories, Albuquerque, NM, Tech. Rep. SAND2002-0835, Apr. 2002.

[23] T. E. de Campos. (2012, Oct.) The Chars74K dataset: Character recognition in natural images. University of Surrey. Guildford, Surrey, UK. [Online]. Available: http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/

[24] E. Leavline and D. A. G. Singh, "Salt and pepper noise detection and removal in gray scale images: An experimental analysis," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, no. 5, pp. 343–352, 2013.