Multispectral image analysis for object recognition and classification

C.R. Viau[a], P. Payeur[a], A.-M. Cretu[b]
[a]School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Ave., Ottawa ON Canada; [b]Département d'informatique et d'ingénierie, Université du Québec en Outaouais, 101, Saint-Jean-Bosco, Gatineau (Québec) Canada

ABSTRACT

Computer and machine vision applications are used in numerous fields to analyze static and dynamic imagery in order to assist or automate decision-making processes. Advancements in sensor technologies now make it possible to capture and visualize imagery at various wavelengths (or bands) of the electromagnetic spectrum. Multispectral imaging has countless applications in various fields including (but not limited to) security, defense, space, medical, manufacturing and archeology. The development of advanced algorithms to process and extract salient information from the imagery is a critical component of the overall system performance.

The fundamental objective of this research project was to investigate the benefits of combining imagery from the visual and thermal bands of the electromagnetic spectrum to improve the recognition rates and accuracy of commonly found objects in an office setting. A multispectral dataset (visual and thermal) was captured and features from the visual and thermal images were extracted and used to train support vector machine (SVM) classifiers. The SVM's class prediction ability was evaluated separately on the visual, thermal and multispectral testing datasets.

Keywords: visual thermal image analysis, feature extraction, support vector machine (SVM), classification

1. INTRODUCTION

A continuing challenge for computer and machine vision applications remains the recognition of objects of interest in real and complex scenes. Object recognition can be accomplished with a certain level of accuracy by using image or template matching whereby several images of the objects are stored in memory and compared to the presented scene. A correlation process is used to identify the object in the scene that appears most like the stored templates. The correlation process is typically performed in the spatial (pixel) domain but can also be performed in the frequency domain. One of the issues with template matching approaches is ensuring that the template or descriptor remains relatively accurate over time. As time unfolds, the object's physical and visual properties can change and may no longer resemble the templates or the descriptors. Conversely, the templates or descriptors may have been obtained under specific conditions that do not match the current scene.

Alternatively, object recognition applications can be based on machine learning and artificial intelligence (AI) algorithms such as neural networks, decision trees, genetic algorithms, and support vector machines (SVM) to name a few. These AI algorithms are trained (typically in an offline process) to recognize features that distinguish the true object from its surroundings. These algorithms usually require a large dataset of imagery exposing the desired object under various viewing angles and conditions. If the object is previously known and a suitable training dataset is available, this type of algorithm can generate high recognition probabilities. However, as with the template matching algorithms, machine learning algorithms are likely to yield low success rate if the object is not previously known or if its appearance is different than in the training dataset.

Object recognition and classification research found in the open literature generally use image datasets from a specific band of the Electromagnetic (EM) spectrum such as X-ray, ultraviolet (UV), visual (visible) or thermal (infrared, IR). Multispectral image analysis is typically used in military and surveillance applications.

The following research investigates how features from visual and thermal imagery can be used jointly to improve the recognition rates of commonly found objects in an office setting. Naturally, the choice of objects was limited to those that radiate thermal energy. A multispectral dataset (visual and thermal) was captured and specific features were extracted to train several SVM classifiers. The SVM's class prediction abilities were evaluated separately on the visual, thermal and the multispectral dataset. Commonly used performance metrics were applied to assess the sensitivity,

Automatic Target Recognition XXVI, edited by Firooz A. Sadjadi, Abhijit Mahalanobis,
Proc. of SPIE Vol. 9844, 98440N · © 2016 SPIE · CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2223360

specificity and accuracy of each classifier. It was decided that the training and testing datasets would be captured using different environments in order to replicate real world scenarios where a machine vision application such as a mobile robot maybe trained to detect an object in an unknown environment. Potential applications for this research could include first responders and security forces.

The remainder of this paper is organized as follows. Section 2 presents a brief review of related works in the field of image features detection, classification and performance metrics. Section 3 discusses the data collection process, the datasets used for the experiments and the image preprocessing. Section 4 presents the methodology which discusses the choice of features and classifiers used as part of the experiments. Section 5 presents the experimental process specifically developed to support this research. Section 6 discusses the evaluation procedures and experimental classification results obtained. Finally, Section 7 summarizes the experiments and suggests future work.

2. LITERATURE REVIEW

2.1 Image Features

In computer vision applications, image features are properties of a scene or of a specific object within that scene that can be extracted to describe the entity. A feature can be something as simple as an object's size or its intensity and can be as detailed as its texture. Features are typically regrouped into three categories; shape, color and texture. A series of features used to describe a scene or an object is referred to as a feature vector or descriptor. For example, a commonly used shape descriptor is the Hu Moments[1] which is composed of a seven dimensional vector. An example of a texture descriptor is the Legendre Moments[2] which is extracted from a local binary pattern and is invariant to translation, scaling and uniform contrast changes. Feature vectors or descriptors such as Hu and Legendre moments can be used by a feature matching algorithm in conjunction with a correlation process or by a machine learning system to identify objects in a scene that closely resemble an object of interest.

Many other object descriptors have been proposed such as SIFT[3], Speeded Up Robust Features[4] (SURF), Features from Accelerated Segment Test[5] (FAST), Binary Robust Independent Elementary Features[6] (BRIEF), Oriented FAST and Rotated BRIEF[7] (ORB). Others algorithms based on Mean-Shift[8], Continuously Adaptive Mean-Shift[9] (CAMSHIFT), covariance, Principal Component Analysis (PCA), and various edge/corner detection methods such as Canny[10], Harris[11], Sobel and SUSAN[12] have been used to detect and track objects of interest. In cases where feature extraction is used for tracking, these algorithms often work in conjunction with a variation of a Kalman or Particle filter to predict the size, position, velocity and acceleration of the targets.

Cayouette, Labonté and Morin[13] investigated the possibility of incorporating a Probabilistic Neural Network (PNN) in an Automatic Target Recognition (ATR) system for an imaging IR seeker emulator. They conducted several tests by shuffling the training and validation datasets and achieved between 95% and 99.43% success rate in correctly identifying aircraft and flares. The features they initially considered included intensity-based features (maximum intensity, average intensity, intensity variance, and intensity distribution) and shape-based features (area, coordinates of the centroid, perimeter, roundness, angle of the principal axis of minimum inertia, small principal moment of inertia, large principal moment of inertia, aspect ratio, maximum radial distance, minimum radial distance, average radial distance and radial distance). From this list of intensity and shape features, they observed the discriminability between the two classes for each feature and selected 13 features. The 13 features were normalized to make them invariant to rotation and translation (see Section 4.1 for feature description).

2.2 Classifiers

A classifier is an implementation of a classification scheme whereby an algorithm is used to learn the characteristics of a class or a pattern from a training dataset and subsequently attempts to recognize the pattern in a separate testing dataset. There are several types of classification schemes or machine learning algorithms such as decision trees, neural networks, SVM, probabilistic methods, Nearest Neighbor, Hidden Markov Model, Bayesian and their variants. A commonality between these machine learning algorithms is that they need to be trained prior to being capable of predicting class associations. The learning is typically done in a supervised or unsupervised way. Simply put, in supervised learning techniques, the class association or the class label is provided to the machine learning algorithm. In unsupervised learning approaches, the class association is not provided and the classification algorithm must look for similarities in the dataset for class prediction. Each method has its advantages over the other and ultimately the type of application and the available data greatly influence the method used for the training of the classification algorithm.

Among these multiple solutions, SVM were chosen in the context of this work because SVM have been used extensively in the literature for various types of statistical pattern recognition applications and demonstrated a high potential for this specific application. SVMs are fast to train, can operate on linear and non-linear data in n-dimensional space, and are usually a top performer in statistical classification challenges. Duda[14] described some of the early work on SVM published in the early to mid-90s which were based on previous work done on margin classifiers (linear machines with margins). SVMs are categorized as linear discriminant classifiers and the general idea behind them is to map data patterns, which cannot be separated by a linear decision boundary, into a much higher dimension. The transformation to the higher dimensional space is achieved through a non-linear mathematical transformation (also known as a kernel) where the input patterns can then be separated by a linear decision boundary or hyperplane. The optimal hyperplane, in the new higher dimensional space, is determined by maximizing the margin (i.e. the separation distance) to the nearest training points. These training points used to compute the margin are known as the support vectors. A larger margin between the support vectors and the optimal hyperplane typically results in better generalization ability by the classifier[14]. Duda also notes that the support vectors are the hardest to classify but most useful in the design of the classifier.

There are numerous forms of kernels used to transform the feature space into a higher dimension and some of the typical ones include[14] Linear Radial Basis Function, Polynomial, Radial Basis Function (RBF) and Sigmoid. SVMs are, in their basic form, binary or two-category classifiers but can be extended to handle multi-category classification problems. This is achieved by combining several binary classifiers[15] (i.e. Class A and not-A, Class B and not-B, Class C and not-C) where the output of the binary classifier carrying the largest weight is selected as the predicted class. The drawback to this approach is that there may be ambiguous regions which cannot be assigned to one class.

3. DATA COLLECTION AND DATASETS

A review of the open literature did not find a suitable dataset for the purposes of this research. The datasets found typically consisted of imagery from the visual or thermal bands of the EM spectrum but rarely from both. In the rare datasets[16] that did have both visual and thermal imagery of the same scene, the view point was very far away and made it very difficult to extract details from the potential objects. In another case, the imagery contained only one class of objects which was again not suitable for the purpose of this research. As a result, no suitable dataset was available and a custom set of matching visual and thermal imagery was captured.

In defining the objectives of this research, it was decided that the imagery collected would be in an indoor setting to have better control on the environment. This requirement added additional challenges since only a limited class of objects actually radiate thermal energy.

3.1 Camera Specifications and Image Analysis Software

The image dataset was acquired using a Fluke Ti10 Thermal Imager. The hand held camera encloses a visual and thermal detector which allows its user to acquire nearly co-located images of an object in both visual and thermal bands. The Ti10 allows the user to select one or more palette to display the apparent temperature of the scene in the camera's field of view. For the purpose of this project, the thermal image were mapped to a greyscale palette to facilitate comparison to greyscale visual images.

The collected dataset consisted of common office items with a thermal signature. Examples of these items include a recently charged mobile phone, a coffee cup with a hot beverage, a laptop charger, a desk lamp and a portable heater. A sample of the five classes of objects is illustrated in Figure 1.

Fluke's SmartView 3.7.23 was used to analyse the imagery captured by the Ti10 thermal imager. The image captured by the Ti10 camera was exported to the Fluke .ISO file format and using the SmartView software converted to visual and thermal (greyscale) bitmap images.

Figure 1. Examples of the five classes of objects used for this research

3.2 Image Preprocessing

The dataset consisted of a total of 173 image pairs captured using the Ti10 thermal imager and was divided into 44 training image pairs and 129 testing image pairs. All of the training images contained a single object class per image. In the case of the testing dataset, several images contained multiple object classes in the same image. This resulted in 44 instances of objects used in the training dataset and 165 instances of objects for the testing dataset. The breakdown of each class in the training and testing datasets is described in Table 1.

Table 1. Dataset description

Class	Name	# of occurrence in the dataset	
		Training	Testing
1	Mobile Phone	7 (15.9 %)	23 (13.9%)
2	Coffee Mug	9 (20.5%)	44 (26.7%)
3	Laptop Charger	10 (22.7%)	28 (17.0%)
4	Desk Lamp	9 (20.5%)	40 (24.2%)
5	Portable Heater	9 (20.5%)	30 (18.2%)
	Total	44	165

The training images were captured using the same background for all objects at various viewing angles and distances to the camera. Examples of the training dataset are illustrated in Figure 2 to Figure 6. Each sample contains a visual band image (top) and a thermal band image (bottom).

The training dataset was not used for testing of the classifiers. Similarly, the testing dataset was not used to train the classifiers. To generate the testing dataset, the same five objects were positioned in different places within two different office spaces. The class object imagery was captured at different distances, under different lighting and in many cases with several objects in the same scene. In order to preserve the thermal signature of the coffee mug relatively constant in all of the testing dataset, the mug was refilled several times with boiling water. Similarly, the cell phone was placed back on a wireless charger for several minutes and the portable heater was restarted for several minutes as well.

In order to challenge the segmentation algorithms, the testing images were captured with the class objects positioned in typical office settings such as on bookshelves, on a work desk next to other objects of the same size and color as well as on a textured carpet. Several testing dataset examples are illustrated in Figure 7 to Figure 11, including the visual band image (top) and the thermal band image (bottom).

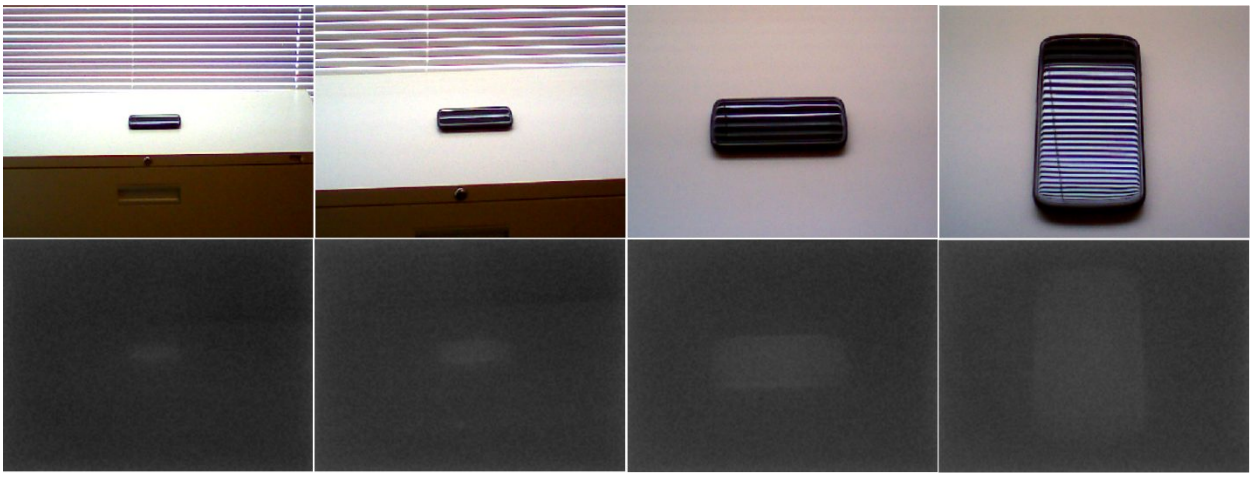

Figure 2. Sample training dataset for Class 1 (Mobile Phone)

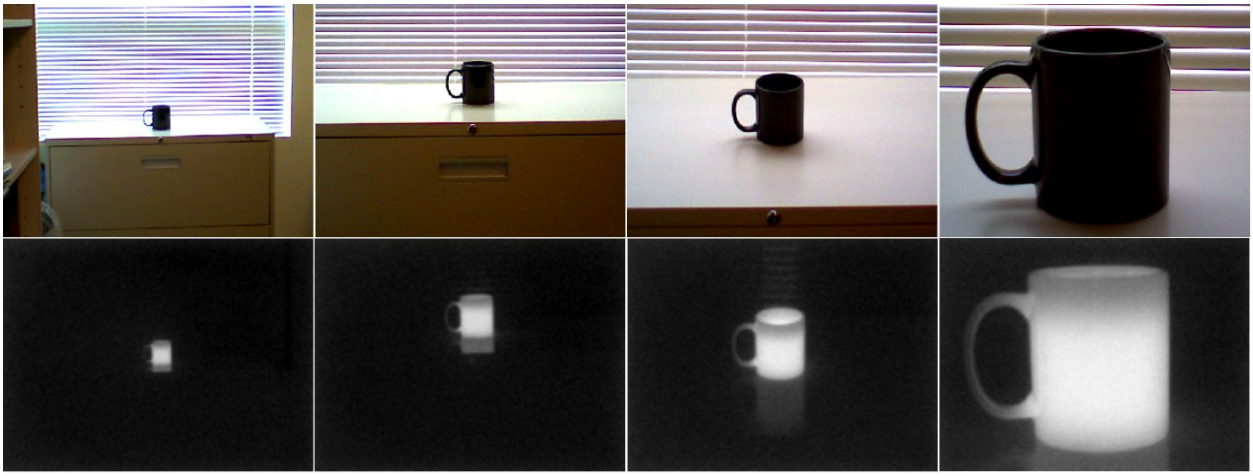

Figure 3. Sample training dataset for Class 2 (Coffee Mug)

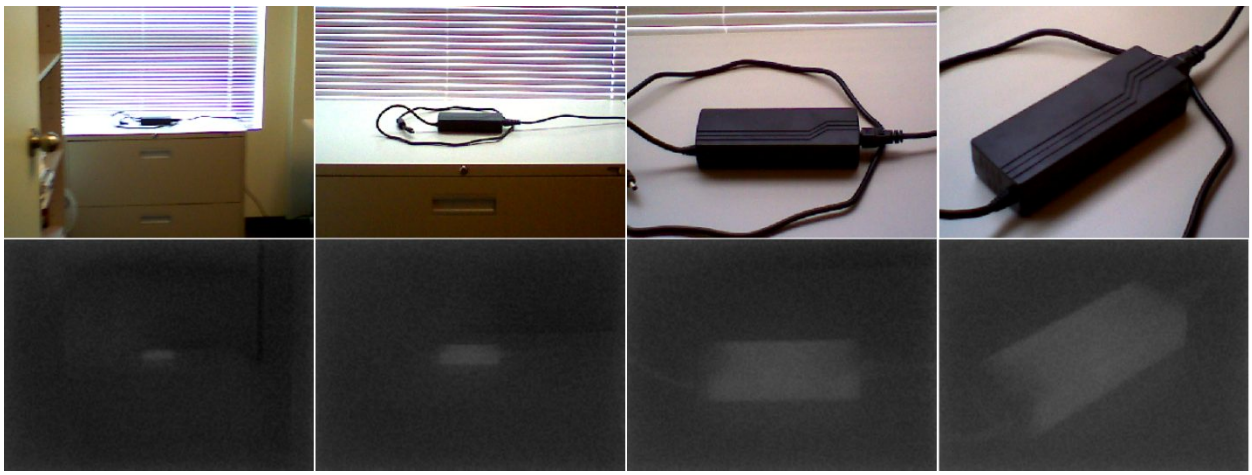

Figure 4. Sample training dataset for Class 3 (Laptop Charger)

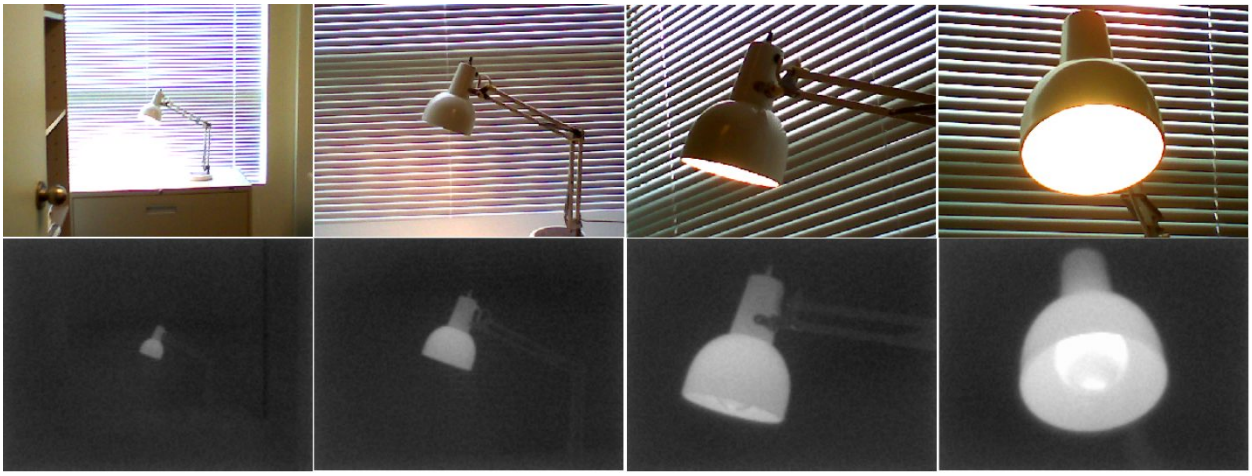

Figure 5. Sample training dataset for Class 4 (Desk Lamp)

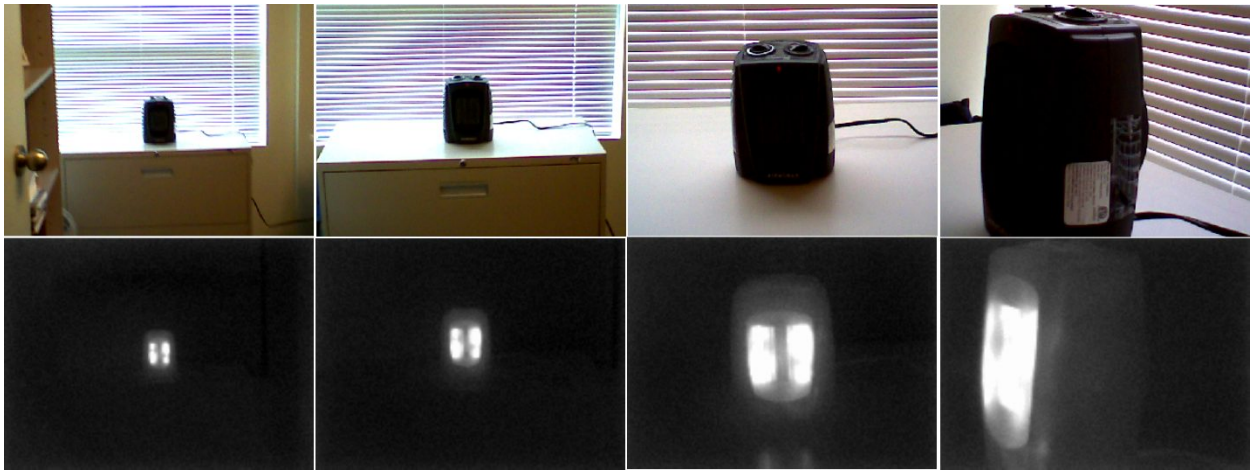

Figure 6. Sample training dataset for Class 5 (Portable Heater)

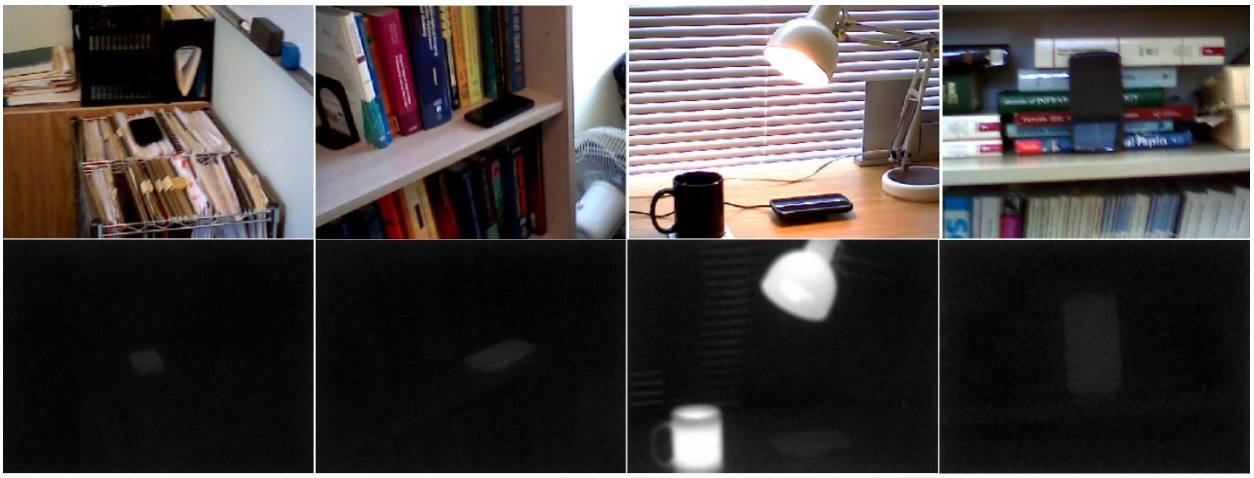

Figure 7. Sample testing dataset for Class 1 (Mobile Phone)

Figure 8. Sample testing dataset for Class 2 (Coffee Mug)

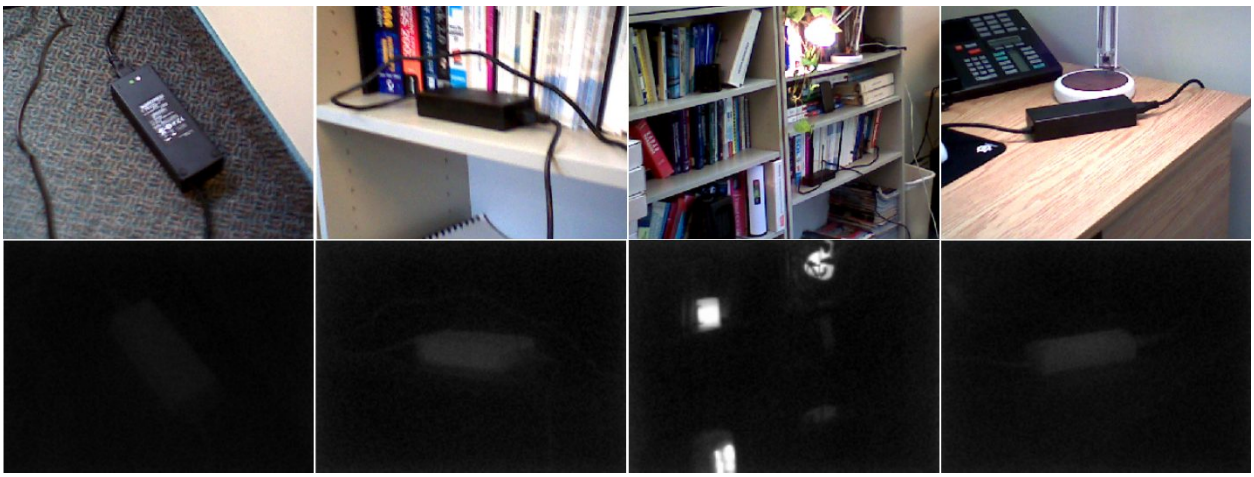

Figure 9. Sample testing dataset for Class 3 (Laptop Charger)

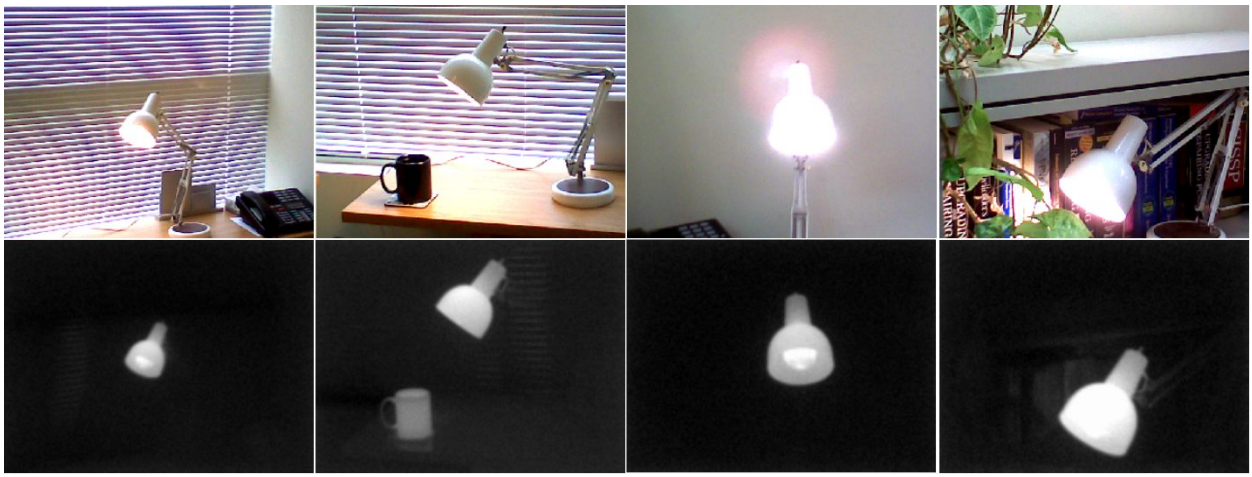

Figure 10. Sample testing dataset for Class 4 (Desk Lamp)

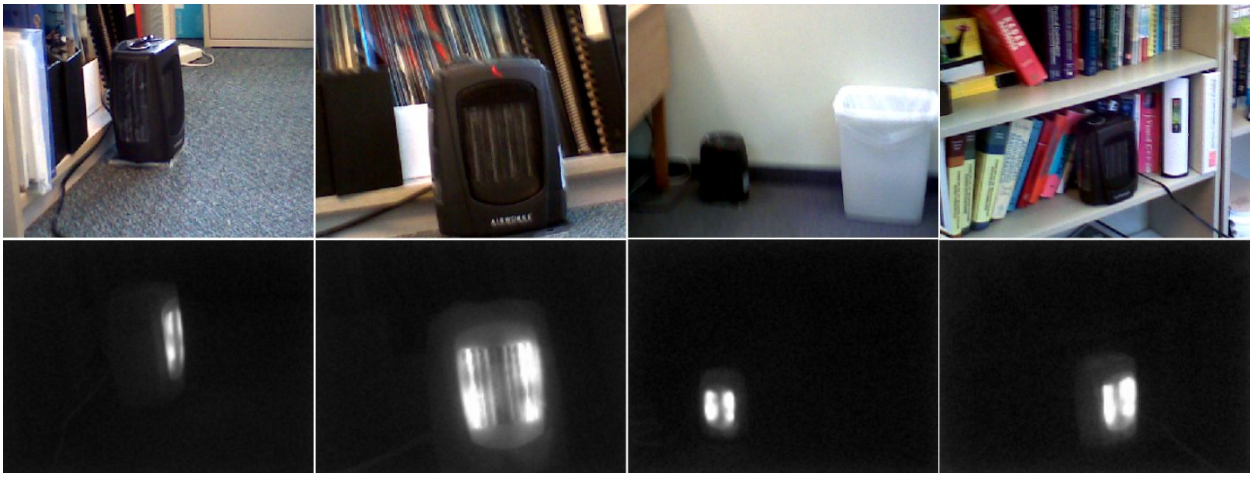

Figure 11. Sample testing dataset for Class 5 (Portable Heater)

4. METHODOLOGY

This section provides a general overview of the methodology used to develop the necessary software tools to achieve the objectives of the research. The principal workflow is illustrated in Figure 12 and consists of first segmenting the image to extract the objects of interest, extracting the desired features from the segmented images, finally training the SVM classifiers and evaluating their class predictability against the testing dataset. Each of these main components of the research are discussed in the following subsections with the exception of the Segmentation component which is presented in a separate publication[17].

Figure 12. Principal workflow

4.1 Image Features Selection

As briefly described in Section 2.1, there are numerous types of features that can be extracted from visual and thermal imagery and are typically categorized by shape, color and texture. For the purpose of this research, it was necessary to identify a set of features that could be extracted from both visual and thermal imagery. Although beyond the scope of this research, the selected features should be easily and efficiently computed such that they could eventually be implemented in hardware for a real-time application. The final consideration was the dimensionality of the feature vector as it would be used to train classifiers.

It was decided to implement the features presented by Cayouette *et al.*[13] primarily because the features had already been evaluated on thermal imagery and demonstrated great discriminability in the context of the application presented. In the present work, the same features were evaluated on visual imagery as well as thermal imagery to demonstrate their discriminability capabilities. The features are presented in Table 2.

The list of features selected represent intensity and shape characteristics of the objects of interest. They do not however represent any texture characteristics and this was intentional because of the low imagery quality produced by the Ti10 camera. In the visual dataset, many of the images were blurred and contained very few textured details. In the thermal imagery, the low dynamic range of the sensor combined with highly radiating objects (such as the bulb of the desk lamp and the elements of the portable heater) resulted in localized pixel saturation in many images. For these reasons, only intensity (both in the visual and thermal images) and shape features were retained.

Table 2. Features Description

Feature Type	Feature	Formula	Feature ID	
			Visual Band	Thermal Band
Intensity	Normalized Maximum Intensity	$\frac{Z_{\max}}{\overline{Z}}$	1	14
	Normalized Average Intensity	$\frac{\overline{Z}}{A}$	2	15
	Normalized Variance of the Intensity Distribution	$\frac{\mu_Z^2}{A^2}$	3	16
	Normalized Third Moment of the Intensity Distribution	$\frac{\mu_Z^3}{A^3}$	4	17
Shape	Normalized Square Root of the Minimum Moment of Inertia	$\frac{\sqrt{I_{\min}}}{A}$	5	18
	Normalized Square Root of the Maximum Moment of Inertia	$\frac{\sqrt{I_{\max}}}{A}$	6	19
	Normalized Maximum Radial Distance	$\frac{D_{\max}}{\sqrt{A}}$	7	20
	Normalized Minimum Radial Distance	$\frac{D_{\min}}{\sqrt{A}}$	8	21
	Normalized Average Radial Distance	$\frac{\overline{D}}{\sqrt{A}}$	9	22 –
	Normalized Second Moment of the Distance Distribution	$\frac{\mu_D^2}{A}$	10	23
	Angle of the Principal Axis of Minimum Inertia	θ	11	24
	Aspect Ratio	$\sqrt{\frac{I_{\max}}{I_{\min}}}$	12	25
	Roundness	$\frac{P^2}{4\pi A}$	13	26

4.2 Classifier Selection

The final component is the classifier selection and description. As presented in Section 2.2, there are several types of classifiers that could be used for this application and as the "no free lunch" theorem suggests, no one classifier is above all in all cases. The choice of the Support Vector Machine (SVM) was based on the availability and well documented open-source library LIBSVM[18]. SVM have been used extensively in the literature for various types of statistical pattern recognition applications and based on the authors' previous experience with the library, demonstrated a high potential for this specific application. SVMs are fast to train, can operate on linear and non-linear data in n-dimensional space, and are usually a top performer in statistical classification challenges. Specifically for these experiments, several multi-class SVM were used with the polynomial kernel.

The open literature[19] suggested that the classification results can be optimized by varying the cost and gamma parameters. As part of the classifier application, a grid search was performed for each feature combination evaluated to optimize the classification results. The cost parameter was evaluated between 2e-4 and 2e5 while the gamma parameter was evaluated between 2e-5 and 2e4. The grid search initially consisted of 100 points and the SVM was trained and tested on each of the 100 grid points. From these 100 results, the highest classification rate was identified along with the associated cost and gamma parameters. An automated process refined the grid search until the best classification rate was achieved. Generally, the optimum cost and gamma parameters for a specific feature combination were identified within 300 to 600 iterations. Once the optimum classification rate was identified, the specific feature combination along with the classification rate, the cost and gamma parameters were stored in an output text file for post classification analysis. The classification analysis results are presented in Section 6.

5. EXPERIMENTS DESIGN

For every training and testing image captured, the 13 features defined in Table 2 were extracted from the visual image and the same 13 features were extracted from the matching thermal image for a total of 26 features per object. Three separate experiments were conducted to determine if the proposed features were suitable for distinguishing between the five classes of objects selected.

The first experiment was to determine the best classification rates using only the visual features, the second experiment was to determine the best classification rates using only the thermal features. The third experiment was to find a feature vector or descriptor combination from both the visual and thermal image pairs which would hopefully produce better results than the visual or thermal alone.

For the first two experiments, the dimensionality of the feature vector could be up to 13 features. However, it was decided to limit the number of features to five in order to reduce the possible number of combinations. In the case of the third experiment, the dimensionality could be up to 26 features. It was decided to limit the feature vector to a maximum of five visual features and five thermal features for a maximum of ten features, once again to limit the possible number of feature combinations. Lottery mathematics[20] was used to determine the number of possible feature vector combinations for experiments 1 and 2.

Assuming that the feature vector had $k = 5$ features with n unique possible values between 1 and 13, then the number of possible combinations (c) is defined as follows:

$$c(n,k) = \frac{n!}{k!(n-k)!} \tag{1}$$

However, since the feature vector can have up to $k = 5$ unique features with values between 1 and 13, Equation 1 has to be evaluated for all possible values of k as follows:

$$c(n,k) = \sum_{k=1}^{5} \frac{n!}{k!(n-k)!} \tag{2}$$

Evaluating Equation 2 for $n = 13$ from $k = 1$ to 5 results in 2,379 possible combinations for the visual dataset and 2,379 possible feature combinations for the thermal dataset.

For the third experiment, the goal was to identify a feature vector that included both visual features and thermal features. In this experiment, the feature vector had to include a minimum of two features (one from the visual image and one from the thermal image) but was limited to 10 features (5 from the visual and 5 from the thermal). If there are 2,379 possible combinations in the visual dataset and the same number in the thermal dataset, the total possible combinations was determined by multiplying the possible combinations in the visual dataset (i.e. 2,379) by the possible combinations in the thermal dataset (i.e. 2,379) for a total possible combinations of 5,659,641.

For the first two experiments, all 2,379 visual band and 2,379 thermal band possible feature combinations were evaluated. In the case of the third experiment, all of the possible 5,659,641 feature combinations were evaluated. The feature combinations were evaluated by systematically combining the 2,379 visual features with the 2,379 thermal features, training, optimizing and testing the SVM classifier. The 5,659,641 combinations were divided into 40 subsets which were evaluated in parallel over several days. The classification results and analysis are provided in Section 6.

6. RESULTS AND ANALYSIS

6.1 Individual Features

One of the objectives of this research was to identify the optimum feature or features to maximize the classification results. In this section, individual features were evaluated and the highest True Positive Rate (TPR) using a single feature in the visual band was 55.8% achieved using Feature 7 (*Normalized Maximum Radial Distances*). In the thermal dataset, the highest TPR was 49.7% also using Feature 7. It is clear from these results that using only 1 feature is not sufficient and that several visual features or several thermal features or a combination of visual and thermal features are required to achieve classification improvement. The following section presents the classification results achieved when combining several visual features to train the SVM classifier.

6.2 Visual Band Features

As discussed in Section 5, it was decided to limit the number of feature combinations in the visual band and thermal band to 5 features. Table 3 illustrates the 5 best feature combinations with the highest TPR using visual band features only. Recall that the classification experiments were conducted on a testing dataset which contained images from all 5 classes and was not used to train the SVM classifiers. The first column of Table 3 represents the overall ranking; the second column represents the TPR while the next five columns represent the feature combination used to achieve the TPR result. As illustrated in Table 3, the highest TPR achieved using up to 5 features in the visual band was 65.5%. This is a 9.7% improvement in comparison to the best of the individual features results (55.8%) reported in the previous section. The highest TPR using a combination of visual band features was achieved with one intensity feature [2] and 4 shape features [5 7 10 12]. Finally, it can be noted that no combination with less than 4 features appeared in the top classification results.

Table 3. Visual band feature classification results

Ranking	TPR (%)	Feature Combination
1	65.5	[2 5 7 10 12]
2	64.2	[2 5 6 7 10]
3	63.6	[5 6 7 9 10]
4	63.0	[7 8 10 12]
5	62.4	[5 6 7 10]

6.3 Thermal Band Features

As discussed in Section 5, it was also decided to limit the number of feature combinations in the thermal band to 5 features. Table 4 illustrates the 5 best feature combinations with the highest TPR using thermal band features only. As illustrated in Table 4, the highest TPR achieved using up to 5 features in the thermal band was 69.1%.

Table 4. Thermal band feature classification results

Ranking	TPR (%)	Feature Combinations
1	69.1	[14 19 20 26]
2	68.5	[14 16 20 26]
3	68.5	[14 19 20 22 26]
4	68.5	[14 15 20 22 26]
5	68.5	[17 20 23 25 26]

This is a 19.4% improvement in comparison to the best individual thermal band feature results (49.7%) reported in Section 6.1. The TPR achieved using a combination of thermal band features only is better than the results achieved using a combination of visual band features only (65.5%) as reported in Section 6.2. The highest TPR using a combination of thermal band features was achieved with 1 intensity feature [14] and 3 shape features [19 20 26]. It is worth noting that Feature 20 and Feature 7 are the same (*Normalized Maximum Radial Distance*) and are prominent features in both the visual and thermal bands. Although the features are computed using the same mathematical formula in both bands, the extracted values are not necessarily the same since they may have different shapes in each bands.

6.4 Combined Visual and Thermal Band Features

In this section, the results of the third experiment conducted to determine if one or more combination(s) of visual and thermal band features could further improve the TPR achieved in the previous sections are reported and discussed. In this experiment, the combination vectors contained between 2 (minimum of 1 from the visual band and 1 from the thermal band) and 10 features (maximum of 5 from the visual and 5 from the thermal bands). As discussed in Section 5, a total of 5,569,641 feature combinations (composed of visual and thermal features) were evaluated to find the optimum combination to maximize the classification rates.

Table 5 illustrates the 5 best evaluated combinations (using visual and thermal feature descriptors) with the highest TPR achieved of 71.5%. This is a 21.8% improvement (71.5% vs. 49.7%) over the individual thermal band TPR, a 15.7% improvement (71.5% vs. 55.8%) over the individual visual band TPR, a 6% improvement (71.5% vs. 65.5%) over the visual band combination, and a 2.4% improvement (71.5% vs. 69.1%) over the thermal band combination results.

Table 5. Combined visual and thermal band feature classification results

Ranking	TPR (%)	Feature Combinations
1	71.5	[2 4 6 14 19 20 22 26]
2	70.9	[3 7 11 15 20 22 23 26]
3	70.9	[3 7 14 15 20 22 26]
4	70.3	[6 7 10 11 13 17 20 21 25 26]
5	70.3	[2 3 4 6 14 20 22 25 26]

In terms of feature descriptors, the feature combinations that achieved the highest classification rates were:

Visual band only: [2 5 7 10 12] (1 intensity feature, 4 shape features)

Thermal band only: [14 19 20 26] (1 intensity feature, 3 shape features)

Combined visual-thermal: [2 4 6 14 19 20 22 26] (3 visual band, 5 thermal band features)

It was determined that the *Normalized Maximum Radial Distance* feature (Feature 7 in the visual band and Feature 20 in the thermal band) was the most prominent feature, and the feature that provided the most discriminability between the classes. The *Normalized Maximum Radial Distance* feature was part of the best feature combination in the visual-only, in the thermal-only and combined visual-thermal classifiers. Also worth noting that the *Roundness* feature (Feature 26) extracted from the thermal band appeared in 49 of the top 50 combined visual-thermal combinations.

7. CONCLUSIONS AND FUTURE WORK

The fundamental objective of this research was to investigate the potential offered by combining features from visual and thermal imagery to improve the recognition rates and accuracy of commonly found objects in an office setting. The research demonstrated that the best results were achieved by the combined visual-thermal classifier which improved the overall True Positive Rate (TPR) of the system to 71.5%. This is an improvement of 6% and 2.4% respectively over the visual-only and thermal-only classifiers.

The features selected for this work were not typical, or commonly used features such as SIFT, FAST, SURF, Hu or Hough Moments discussed in Section 2.1, but rather a unique set of features that were previously used in another classification study[13] using thermal imagery only. In Cayouette *et al.*'s study[13], classification rates over 95% were achieved using only 2 classes of objects (aircraft and flare) that clearly differentiated from each other. A direct comparison between the results achieved in this study and those achieved by Cayouette *et al*. is difficult to make because in this study 5 different classes were used instead of 2, visual and thermal imagery were used instead of thermal only and much fewer training/testing samples were available (209 vs. 1264).

It is possible that other feature descriptors (different features) could achieve better classification rates than the descriptors selected for this work. Furthermore, it was decided to limit the dimension of the descriptor length to 5 features (up to 13 available) for the visual and thermal-only classifiers and to 10 features (up to 26 available) for the combined visual-thermal classifier. It is possible that a feature descriptor with a dimension greater than 10 could provide better classification rates but would also increase computational time.

In the early stages of the experimental design, it was decided that two separate datasets would be acquired; one to train the classifiers and one to test the classifiers. The training dataset was generated in a different environment setting than the testing dataset. As described in Section 3, the training imagery was captured using a common background while the testing dataset was captured in much more complex backgrounds intended to create different levels of difficulty for the segmentation algorithm. It is possible that if the training dataset had been captured in the same setting as the testing dataset, that the classifier might have achieved better classification rates (more similarities between the training and testing samples). In addition, the size of the training dataset was relatively small in comparison to the testing dataset. The results presented in this work were achieved using between 7 and 10 training samples per class. It is possible that better classification rates could be achieved if more training samples were used.

In many classification studies, a single large dataset is divided into a training and testing subset. In these studies, the datasets are randomly divided and re-tested with new subsets each time to obtain an average classification rate using cross-validation techniques such as n-fold, hold-out and leave-one-out. Various techniques of meta-learning such as Arcing, Bagging[21] and Boosting[22] can be used to further enhance the classification results by choosing an optimum subset to train the classifiers. It is possible that by combining cross-validation and boosting techniques, better classification rates could have been achieved.

Finally, the image quality of the Ti10 camera was poor at best. In many image pairs, the visual image was blurry and had low level of details. In the thermal samples, the small dynamic range of the sensor resulted in limited thermal details of the objects. A better quality imager could very likely improve the classification results achieved in this study and could potentially allow for additional features based on object texture in both the visual and thermal bands to be used as part of the descriptor.

REFERENCES

[1] Hu, M.-K., "Visual pattern recognition by moment invariants," IEEE Trans. Inf. Theory 8(2), 179–187 (1962).

[2] Teague, M. R., "Image analysis via the general theory of moments," J. Opt. Soc. Am. 70(8), 920 (1980).

[3] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vis. 60(2), 91–110 (2004).

[4] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., "Speeded-Up Robust Features (SURF)," Computer Vision Image Understanding 110(3), 346–359 (2008).

[5] Rosten, E., Drummond, T., "Machine Learning for High-Speed Corner Detection," [Computer Vision – ECCV 2006], A. Leonardis, H. Bischof, and A. Pinz, Eds., Springer Berlin Heidelberg, 430–443 (2006).

[6] Calonder, M., Lepetit, V., Strecha, C., Fua, P., "BRIEF: Binary Robust Independent Elementary Features," [Computer Vision – ECCV 2010], K. Daniilidis, P. Maragos, and N. Paragios, Eds., Springer Berlin Heidelberg, 778–792 (2010).
[7] Rublee, E., Rabaud, V., Konolige, K., Bradski, G., "ORB: An efficient alternative to SIFT or SURF," 2011 IEEE Int. Conf. Comput. Vis. ICCV, 2564–2571, IEEE (2011).
[8] Fukunaga, K., Hostetler, L., "The estimation of the gradient of a density function, with applications in pattern recognition," IEEE Trans. Inf. Theory 21(1), 32–40 (1975).
[9] Bradski, G. R., Computer Vision Face Tracking For Use in a Perceptual User Interface, Intel Technology Journal Q2(Q2):214-219 (1998).
[10] Canny, J., "A Computational Approach to Edge Detection," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8(6), 679–698 (1986).
[11] Harris, C., Stephens, M., "A Combined Corner and Edge Detector," Proc. 4th Alvey Vis. Conf., 147–151 (1988).
[12] Smith, S. M., Brady, J. M., "SUSAN—A New Approach to Low Level Image Processing," Int. J. Comput. Vis. 23(1), 45–78 (1997).
[13] Cayouette, P., Labonte, G., Morin, A., "Probabilistic neural networks for infrared imaging target discrimination," Proc SPIE 5094, F. A. Sadjadi, Ed., 254–265 (2003).
[14] Duda, R. O., Pattern classification, 2nd ed, Wiley, New York (2001).
[15] Cretu, A.-M., Payeur, P., "Biologically-inspired visual attention features for a vehicle classification task," Int. J. Smart Sens. Intell. Syst. 4(3), 402–423 (2011).
[16] James W. Davis, V. S., "Background-Subtraction Using Contour-Based Fusion of Thermal and Visible Imagery," Computer Vision Image Understanding 106(2-3), 162–182 (2007).
[17] Viau, C., Payeur, P., Cretu, A.-M., "An improved watershed segmentation algorithm with thermal markers for multispectral image analysis," Proc SPIE 9844, Automatic Target Recognition XXVI, Baltimore, USA (2016).
[18] Chang, C.-C., Lin, C.-J., "LIBSVM: A Library for Support Vector Machines," ACM Trans Intell Syst Technol 2(3), 27:1–27:27 (2011).
[19] "Support vector machine", Wikipedia Free Encycl. <https://en.wikipedia.org/wiki/Support_vector_machine> (21 August 2015).
[20] "Lottery mathematics", Wikipedia Free Encycl. (2015).
[21] Breiman, L., "Bagging Predictors," Mach. Learn. 24(2), 123–140 (1996).
[22] Freund, Y., Schapire, R. E., "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," J. Comput. Syst. Sci. 55(1), 119–139 (1997).