ROSE 2008 – IEEE International Workshop
on Robotic and Sensors Environments
Ottawa – Canada, 17-18 October 2008

Feature Selection for a Real-Time Vision-Based Food Inspection System

Mai Moussa Chetima, Pierre Payeur
School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada, K1N 6N5
[mchet090, ppayeur]@site.uottawa.ca

Abstract *– In a world where automation of processes is more and more on demand, machine vision is continuously explored to address several industrial problems such as quality inspection. In the processed-food industry where the external quality attributes of the product are inspected visually before the packaging line, machine vision systems often involve the extraction of a larger number of features than those actually needed to ensure proper quality control. This work experiments with several feature selection techniques in order to reduce the number of attributes analyzed by a real-time vision-based food inspection system. Four filter-based and wrapper-based feature selectors are evaluated on seeded buns and tortillas datasets. Experimental results show that consistency-based and the RELIEF subset evaluation techniques perform the best for all the considered datasets in terms of accuracy. However, variations in the number of attributes selected still vary significantly between these techniques.*

Keywords *– Machine vision, food inspection, feature selection, machine learning.*

I. INTRODUCTION

For several years, the food industry has been trying to automate the product quality verification process in order to decrease production costs and increase the quality and uniformity of the production. Machine vision-based systems are of particular interest when it comes to extracting features of a product for classification purposes. Most of the external quality attributes of a product are usually inspected visually before the packaging line and items that do not satisfy to the set standards are automatically rejected. Such machine vision systems have been used for several inspection applications in the food industry including meat, fruits and vegetables, bakery products, and prepared consumer foods [1][2][3].

Machine learning is at the core of several vision-based inspection systems. But in many applications, it is not known exactly which features are critical for the quality control or how they should be represented. It is therefore difficult and time-consuming to determine on which features to focus the system's attention when configuring the inspection system in order to sustain a high production rate with a minimum of redundancy. One intuitive solution is to include all features that could possibly be relevant and let the learning algorithm decide which features are in fact worthwhile [4]. A more formal way is to identify the relevant features by means of feature selection techniques and make the inspection system concentrate only over a space of a reduced dimension. Such feature selection techniques are often categorized as filters or wrappers. In the filter approach, the feature selector is independent of the learning algorithm and serves as a filter to sieve the irrelevant and/or redundant attributes. The wrapper feature selector rather works around the learning algorithm to actively determine the relevant attributes.

The motivation for this research work is to evaluate the effectiveness of some state-of-the-art feature selection techniques for an application on real-time vision-based food inspection systems that operate on several types of bakery products such as hamburger buns, tortillas, and croissants. The development of an automated process for the determination of the most relevant features that can also adapt to the type of products being inspected represents a major evolution in the technology. It makes the configuration and maintenance of inspection systems more straightforward, even for new products, while improving the uniformity of the production.

Section II discusses four filter-based and wrapper-based feature selectors that have been considered. Section III presents and analyzes the outcomes of an experimental evaluation of those feature selection techniques on datasets generated with an existing real-time vision-based food inspection system.

II. FEATURE SELECTION TECHNIQUES

As mentioned previously, filter and wrapper approaches are examined for the application considered. Three different filter-based techniques are detailed: a correlation-based, a consistency-based, and the RELIEF feature selection methods respectively. One wrapper-based technique is also examined.

A. Correlation-based Feature Selection

Correlation-based feature selection (CFS), introduced by Hall [5][6], evaluates subsets of attributes rather than individual attributes. Hall's rationale of this technique is based on the hypothesis that *"a good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other"* [5]. The first part of this hypothesis is inspired by Gennari *et al.* [7] who stated that features are relevant if their values vary systematically with category membership. This statement has been formalized by Kohavi and John [8] who formulated that a feature V_i is said to be relevant if and only if there exists some v_i and c for which $P(V_i = v_i) > 0$ such that:

$$P\left(C=c \middle| V_i = v_i\right) \neq P(C=c) \quad (1)$$

Theoretical and empirical evidence encourages removing redundant information along with irrelevant features [8][9][10]. A feature is considered redundant if it is highly correlated with one or more other features. CFS uses the following heuristic evaluation to rank feature subsets:

$$Merit_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (2)$$

where $Merit_s$ is the heuristic "merit" of a feature subset S containing k features, $\overline{r_{cf}}$ is the average feature-class correlation, and $\overline{r_{ff}}$ is the average feature-feature correlation. The numerator can be interpreted as an indication of how predictive a group of features is, as for k fixed and greater than 1, the feature-class correlation average $\overline{r_{cf}}$ will be relatively large if the group of features is correlated with the class and small otherwise; and therefore the numerator allows discriminating irrelevant features. On the other hand, the denominator discriminates redundant features because in case of redundant attributes (respectively non redundant), the feature-feature correlation average $\overline{r_{ff}}$ will be large (small), which implies a larger (smaller) denominator, and therefore a smaller (larger) $Merit_s$. The correlation between features is computed using symmetrical uncertainty (SU):

$$SU = 2.0 \times \left[\frac{H(Y) + H(X) - H(X,Y)}{H(Y) + H(X)}\right] \quad (3)$$

where H(Y) is the entropy of a discrete feature Y and H(X, Y) is the entropy of a discrete feature X after observing Y.

B. Consistency-based Feature Selection

In consistency-based feature selection, training instances are projected onto the subset of attributes and then the consistency of the subset is evaluated. It is therefore common practice to use consistency-based subset evaluator in conjunction with a search algorithm that looks for the smallest subset with consistency. Liu and Setiono proposed an inconsistency evaluation criterion in [11]. Two instances are considered inconsistent if they match except for their class labels. The inconsistency criterion is computed as follows: (1) suppose there are k possible class labels $label_1$, $label_2, \ldots,$ $label_k$ in a certain dataset which contains N instances; (2) suppose there are J distinct combinations of attribute values for a subset s of attributes (without considering the class labels of the instances); (3) suppose that D_i is the number of occurrences (or matching instances without considering the class labels of the instances) of the i^{th} combination of attribute values; (4) suppose that among the D_i instances, c_1 instances belong to class label $label_1$, c_2 instances belong to $label_2$, …, and c_k instances belong to class label $label_k$, such that $c_1 + c_2 + \ldots + c_k = D_i$ and let $M_i = \max\{c_1, c_2, \ldots, c_k\}$. Then the inconsistency count is given by:

$$inconsistency \;\; count = \left(D_i - M_i\right) \quad (4)$$

In other words, for matching instances (without considering the class labels of the instances) in a subset s of attributes, the more the class labels match, the less inconsistent (or the more consistent) is the subset s with respect to the class. The inconsistency rate of an attribute subset s is given by the sum of all inconsistency counts divided by the total number of instances:

$$inconsistency_S = \frac{\sum_{i=1}^{J} D_i - M_i}{N} \quad (5)$$

Several search strategies can be used to look for the smallest most consistent subset of attributes.

C. RELIEF Feature Selection

The RELIEF attribute selection technique uses general characteristics of the data to evaluate attributes and operates independently of any learning algorithm. RELIEF was first introduced by Kira and Rendell [12] as a means of estimating the "quality" of attributes with and without dependencies among them. RELIEF is an instance-based attribute ranking scheme that works by randomly sampling an instance from the data and then locating its nearest neighbors from the same and opposite classes. The neighbor from the same class is named *nearest hit* and the one from the opposite class is called *nearest miss*. The values of the attributes of the nearest neighbors are compared to the sampled instance and used to update relevance scores for each attribute. The rationale of the RELIEF algorithm is that useful attributes should differentiate between instances from different classes and have the same value for instances from the same class.

The original version of RELIEF is limited to only two-class problems, which led Kononenko [13] to extend the original RELIEF to deal with noisy, incomplete, and multi-class datasets. The first enhancement that Kononenko addressed was to increase the reliability of probability approximation by searching the k-nearest hits/misses instead of only one near hit/miss. The enhanced version, called RELIEF-F, finds nearest neighbors from each class different than the current sampled instance and averages their contribution for updating estimates W[A], and finally weights the average with prior probability of each class as follows [14]:

set all weights W[A] := 0.0;
for i := 1 to m do
begin
randomly select an instance R;
find k nearest hits H_j;
for each class C $\neq$ class(R) do
find k nearest misses M_j(C)
for A := 1 to NumberOfAttributes do

$$W[A] := W[A] - \sum_{j=1}^{k} diff(A, R, H_j)/(m \times k) +$$

$$\sum_{C \neq class(R)} \left[\frac{P(C)}{1 - P(class(R))} \sum_{j=1}^{k} diff(A, R, M_j(C)) \right] / (m \times k)$$

end

where *NumberOfAttributes* is the total number of attributes in the original dataset, and *diff(Attribute, Instance1, Instance2)* computes the difference between the values of *Attribute* for two instances. For discrete attributes, the difference is either 1 (the values are different) or 0 (the values are the same), while for continuous attributes the difference is the actual difference normalized to the interval [0, 1]. The same function *diff(.)* is used for calculating the distance between instances to find the nearest neighbors.

D. Wrapper Feature Selection

A wrapper feature selection approach uses a machine learning algorithm as a black box, therefore needing only the interface of the induction algorithm. In fact, knowledge of the learning algorithm itself is not necessary [8]. The wrapper feature selector repeatedly searches for a "good" feature subset by using the induction algorithm as part of the evaluation function. In [8], Kohavi and John suggest using *10-fold cross-validation* method for model selection. Cross-validation is explained in the next section.

III. EXPERIMENTAL EVALUATION ON VISION-BASED FOOD INSPECTION SYSTEM

A. Experimental Setup

The vision-based food inspection system used for our experimentation is a technology that automates visual identification and classification of bakery products such as buns, cookies, tortillas, and pizzas. The system, shown in Fig. 1, is equipped with a conveyor belt which moves the bakery products to the rejection and packaging systems. One camera is mounted above the conveyor belt and produces real-time line scans of the top view of products moving with the conveyor belt. A laser light strip is also projected vertically on the conveyor belt and an extra profile camera sensing the laser light in diagonal generates real-time height information on every product. Real-time image processing algorithms combine the height information with the data from the line scan camera in order to estimate a set of parametric features of the products under inspection. Depending on the product, up to 200 features can be extracted. The system analyzes the features of every product and then orders rejection of the product if it is classified as defective, or orders packaging if the product is classified as acceptable. For the purpose of our experiments, all extracted features and the system's decision were saved on every single product to create large datasets. In order to evaluate the accuracy of the proposed feature reduction techniques, it is assumed that the system produces correct classification on all items and therefore the accuracy is evaluated against the original classification of the system [15].

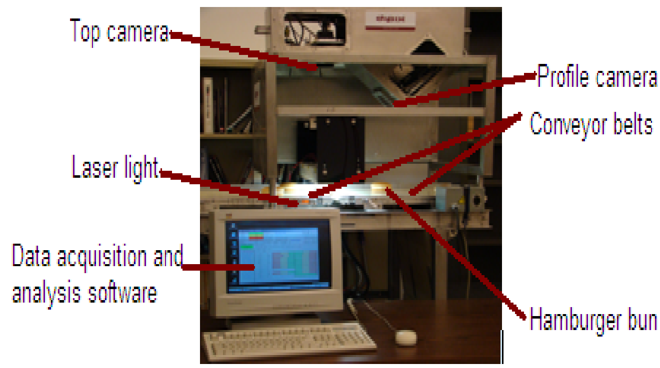

Fig. 1. Experimental System

B. Procedure for the Comparative Evaluation

Experiments were conducted on two of the most common products inspected by the vision-based food inspection system available: buns and tortillas. Seeded buns were selected as they contain more features and more complexity than regular unseeded ones. Buns and tortillas both have "irregular" shapes as none of them has a perfectly defined geometrical shape. For both the buns and the tortillas datasets, 82 continuous features are extracted per product item, plus one boolean feature representing the decision to reject or accept the item. Each dataset contains 3287 product items. Prior to feature selection, the continuous features are discretized using a supervised discretization technique introduced by Fayyad and Irani [16] which combines an entropy-based splitting criterion with a minimum description length stopping criterion [14].

In order to evaluate the accuracy achieved with the various feature selectors, three different machine learning techniques are used that represent different approaches to learning: Naive Bayes (a probabilistic learner), C4.5 (a decision tree learner) and Multi-Layer Perceptron (MLP, a neural networks learner). The Naive Bayes algorithm assumes that features are conditionally independent given the label and computes the posterior probability of each class given the feature values present in the class, and assigns the instance to the class with the highest probability [17][18]. C4.5 is an algorithm which builds a decision tree top-down by recursively finding the best single feature test to conduct at the root node of the tree [19]. The MLP is a hierarchical structure of several perceptrons with weighted interconnections able to capture complex input/output relationships from training data [20].

For wrapper-based feature selectors an interaction with the learning algorithm is required during feature selection.

We used two different approaches to evaluate accuracy: holdout and cross-validation methods [21]. The holdout method, also called test sample estimation, separates the data into two mutually exclusive subsets: the training set and the test set, or holdout set. In *k*-fold cross-validation, the data is randomly split into *k* mutually exclusive subsets (the folds) of approximately equal size. The learner is trained and tested *k* times, each time with (*k*-1) different training sets and one different test set. The overall accuracy is averaged over the *k* folds. We used 10-fold cross-validation as suggested in [21], and repeated it 10 times with different random seeds. We repeated holdout 100 times on our datasets and averaged the overall accuracy.

It is worth mentioning that the continuous features are discretized only for feature selection purposes. After feature selection, the reduced datasets are extracted from the original continuous datasets and then passed to the learning algorithms for accuracy estimation. The same train/test sets and the same folds were used for all learning schemes in order establish a common base for comparison.

C. Experimental Results on Tortillas and Seeded Buns

Table 1 and Table 2 show the number of features selected for tortillas and seeded buns respectively, sorted in ascending number of features selected by the different dimensionality reduction techniques. For the tortillas dataset, the C4.5 wrapper algorithm is top on the list by selecting only 4 features out of 82, followed equally by the CFS feature selector and the MLP wrapper that both select 5 features. The RELIEF algorithm is the last on the list as it keeps up to 33 attributes out of 82. For the buns dataset, the C4.5 wrapper is still on top of the list with 2 features selected, followed by the consistency-based subset evaluation (5 out of 82). The Naive Bayes wrapper and the MLP wrapper both found 6 attributes to be relevant to their respective learning algorithms. RELIEF is again the one rejecting the less attributes.

For both the buns and the tortillas datasets, C4.5 wrapper, MLP wrapper, Naive Bayesian wrapper, CFS and consistency-based subset evaluation tend to consider 10 or less features out of 82 as relevant to qualify the product, whereas RELIEF seems to be cautious by keeping many more features. The fact that only positive differences of probability were kept in the RELIEF implementation implies that weakly relevant features are very likely to be preserved. Kohavi and John also mentioned that the RELIEF algorithm tends to keep most of the relevant features of a dataset even if they are redundant and only a fraction of them is necessary for the concept description [8]. Moreover, wrapper feature selection techniques globally tend to select fewer features than filter-based feature selectors. This could be explained by the fact that wrappers are meant to optimize the feature selection for a particular given algorithm with which they interact during the attribute selection process.

Table 1. Features space reduction performance on the tortillas dataset.

	Number of features selected	Number of features rejected	Feature rejection ratio (%)
C4.5 wrapper	4	78	95.12
CFS	5	77	93.90
MLP wrapper	5	77	93.90
NB wrapper	7	75	91.46
Consistency	10	72	87.80
RELIEF	33	49	59.76

Table 2. Features space reduction performance on the buns dataset.

	Number of features selected	Number of features rejected	Feature rejection ratio (%)
C4.5 wrapper	2	80	97.56
Consistency	5	77	93.90
MLP wrapper	6	76	92.68
NB wrapper	6	76	92.68
CFS	7	75	91.46
RELIEF	59	23	28.04

However, the number of attributes selected as relevant by the different feature selectors should be interpreted with caution. In fact, although a dataset with fewer features would be preferred for production rate enhancement, the accuracy of prediction with the reduced datasets is of capital importance. This makes us favor prediction accuracy over dimension reduction of the dataset for industrial quality inspection applications, as considered here.

Fig. 2 and Fig. 3 show the accuracy estimation with the different feature selection techniques for the tortillas dataset, evaluated using three learning schemes: Naive Bayes, C4.5 and Multi-layer Perceptron (MLP). The accuracy represents the percentage of products that were classified from the reduced set of features in the exact same way as the original classification of the system, as defined in the dataset. On both figures, "Tortilla" represents the original full dataset which did not undergo any feature selection and the vertical bars at the edge of the columns represent the standard deviation. Fig. 2 shows the result of 10 repetitions of 10-fold cross-validation and Fig. 2 presents the result of 100 repetitions of holdout accuracy test.

For all feature selectors, the C4.5 learning scheme globally gives the highest accuracy and the lowest standard deviation at the same time, followed closely by the MLP learner and far beyond by the Naive Bayes learner. Holdout and cross-validation gave comparable results in terms of accuracy and standard deviation. It is important to note that in the experiments presented, both cross-validation and holdout use 90% of the data for training and 10% for test. For 10-fold cross-validation, the full original tortillas dataset had an accuracy of 99.39% with a standard deviation of 0.47% using the C4.5 learning scheme. Consistency-based subset evaluation gave a better accuracy (99.44%) than the full dataset. The RELIEF feature selector is the second best by having the same accuracy estimation as the original dataset, followed by the C4.5 wrapper and the correlation-based feature selector with an accuracy of respectively 0.03% and 0.04% below that of the full dataset. The holdout tests gave approximately the same order by ranking consistency and RELIEF equally accurate to the original full dataset (accuracy of 99.37%), followed by CFS and C4.5 wrapper achieving an equal accuracy of 0.01% inferior to that of the original dataset. The standard deviation for all feature-reduced datasets are all between 0.39% and 0.47% and are considered not high enough to impact the interpretation of our results.

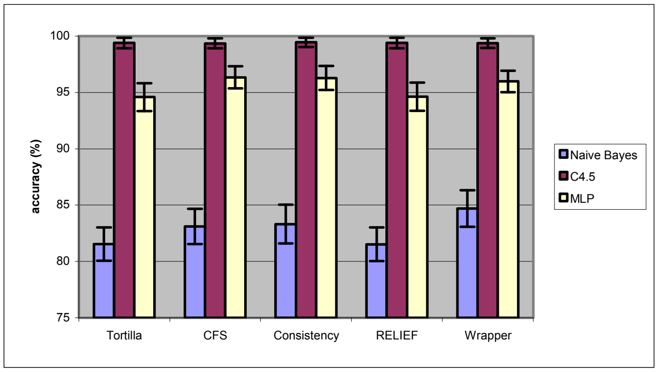

Fig. 2. Ten repetitions of "10-fold cross-validation" accuracy estimation for the tortillas dataset.

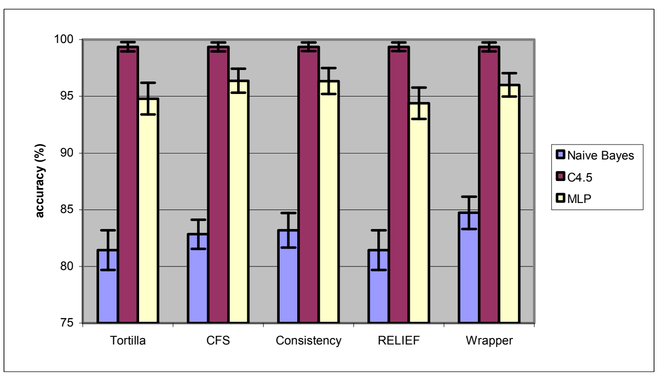

Fig. 3. One hundred repetitions of "Holdout" accuracy estimation for the tortillas dataset.

Fig. 4 shows the results of 10 repetitions of 10-fold cross-validation for the buns dataset. The holdout test results are not presented here as they are similar to the cross-validation results. C4.5 is once again the learning scheme giving globally the best accuracy estimation, followed by the MLP and the Naive Bayes respectively. The original full buns dataset has an accuracy of approximately 99.81% with a standard deviation of 0.25% and none of the reduced dataset is able to achieve a better accuracy. However, the consistency-based subset evaluation and the C4.5 wrapper closely follow the original full buns dataset by both achieving an estimated accuracy only 0.58% inferior to that of the full dataset. RELIEF occupies the third place and CFS the fourth with an accuracy less than 1% lower than that of the full buns dataset.

The fact that the Naive Bayes classifier gives lower accuracy estimations compared to the C4.5 and the Multi-Layer Perceptron can be attributed to the assumption that the algorithm makes about features being conditionally independent. In fact, several of the features in the dataset are correlated, for example features such as the mean diameter of an approximately circular product and its surface area are clearly correlated. MLP was able to capture a certain rule for the classification of the products considered because of the structure inherent to the MLP which allows capturing complex input/output relationships. C4.5 giving very high classification accuracy can be explained by the fact that the vision-based food inspection system inherently uses a decision structure very close to a decision tree.

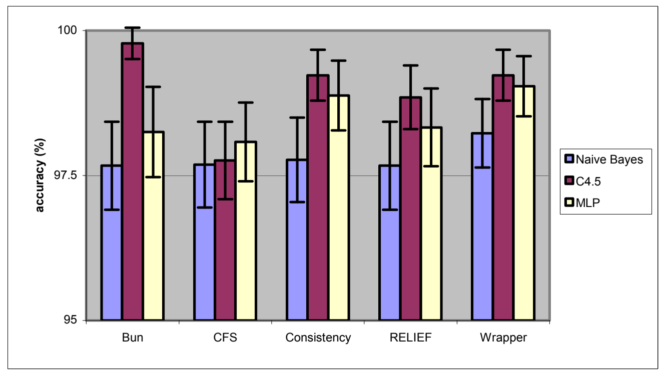

Fig. 4. Ten repetitions of "10-fold cross-validation" accuracy estimation for the buns dataset.

Globally, the advantage of a consistency-based subset evaluation is undisputed accuracy-wise for both the buns and the tortillas datasets. Considering the number of attributes retained by the different feature selectors, consistency-based subset evaluation acquires another advantage over RELIEF by proposing only 10 attributes vs. 33 for the tortillas dataset and 5 vs. 59 for the buns dataset. One might explain the success of consistency-based subset evaluation and RELIEF techniques by their ability to capture attribute interactions. CFS also gives reasonably good results, especially when the relatively small number of features selected by this filter is considered. In the case considered here, several features in the dataset are correlated and CFS appears to be able to identify these correlations. The C4.5 wrapper has a net advantage with respect to the number of features selected and wrappers actually tend to give better results than filters. But this benefit is compromised by the time it takes to train wrappers, which can reach several minutes rather than only a few seconds with filter-based selectors, due to the repetitive interaction with the learning schemes.

According to our experiments, consistency-based subset evaluation seems to be the best suited feature selector for application on a real-time vision-based food inspection system, provided that the products under classification are similar to the ones analyzed here. RELIEF ranks as the second best candidate in terms of accuracy and training time, but tends to keep a lot more features than needed. In fact, not only does consistency-based subset evaluation outperform all the other feature selectors' accuracy for the C4.5 algorithm, but it also generally gives a better accuracy for all the learning schemes we experimented with and for all the datasets. Instead of analyzing and interpreting 82 features for every product under classification and in-real time, the vision-based food inspection system can now focus on less than 10 features and produce classification accuracy similar or even better than with the full 82 features. Reducing the number of features extracted and analyzed in real-time reduces the time taken for processing a single product and therefore allows the food inspection system to support higher production rates. It is however necessary to emphasize that having quality training data covering as many cases as possible is a definite key in helping the algorithms generalize. Therefore particular caution in choosing the data samples shall be applied to prevent poor generalization of the learner.

IV. CONCLUSION

This paper presented and evaluated four feature selection techniques for parametric space reduction in a real-time vision-based food inspection system. Three machine learning algorithms were used to evaluate and compare the accuracy of the classification from extracted features with and without feature selection. Experimental results on seeded buns and tortillas demonstrated that consistency-based subset evaluation outperforms all other feature selectors in terms of accuracy, and is also very competitive in terms of the number of selected attributes. The RELIEF technique also revealed good performance, but has the disadvantage of keeping more features than the other selectors, which might impede production rates. Wrapper approaches tend to give excellent accuracy results, especially with the C4.5 decision tree inducer, but take a longer time to train.

Most of the feature-reduced datasets provided by the attribute selectors gave an estimated accuracy very close to the accuracy achieved with the full datasets, and even higher accuracy when extracted with the consistency-based subset evaluation technique. Apart from the RELIEF algorithm, all feature selectors reduced the parameter space dimension by more than 85%. This demonstrates the relevance of integrating feature selectors into the vision-based food inspection system such that it can focus on fewer features and still provide inspection results of a comparable quality while allowing a higher rate of production.

ACKNOWLEDGEMENTS

The authors acknowledge the partial financial support of Precarn Inc., and the collaboration of Dipix Technologies Inc. and of Son's Bakery to this research.

REFERENCES

[1] M.R. Chandraratne, D. Kulasiri, and S. Samarasinghe, "Classification of Lamb Carcass Using Machine Vision: Comparison of Statistical and Neural Network Analyses", *Journal of Food Engineering*, vol. 82, no. 1, pp. 26-34, 2007.

[2] J. Blasco, N. Aleixos, and E. Molto, "Computer Vision Detection of Peel Defects in Citrus by Means of a Region Oriented Segmentation Algorithm", *Journal of Food Engineering*, vol. 81, no. 3, pp. 535-543, 2007.

[3] T. Brosnan and D.-W. Sun, "Improving Quality Inspection of Food Products by Computer Vision – a Review", *Journal of Food Engineering*, vol. 61, pp. 3-16, 2004.

[4] H. Almuallim and T.G. Dietterich, "Learning with Many Irrelevant Features", *Proc. of the 9th National Conf. on Artificial Intelligence,* pp. 547-552, AAAI Press, 1991.

[5] M.A. Hall, *Correlation-Based Feature Selection for Machine Learning*, PhD thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1998.

[6] M.A. Hall, "Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning", *Proc. of the 17th Intl Conf. on Machine Learning,* pp. 359-366, 2000.

[7] J.H. Gennari, P. Langley, and D. Fisher, "Models of Incremental Concept Formation", *Artificial Intelligence,* vol. 40, no. 4, pp. 11-61, 1999.

[8] R. Kohavi and G.H. John, "Wrappers for Feature Subset Selection", *Artificial Intelligence,* vol. 97, pp. 273-324, December 1997.

[9] P. Langley and S. Sage, "Induction of Selective Bayesian Classifiers", *Proc. of the 10th Conf. on Uncertainty in Artificial Intelligence,* Seattle, W.A, Morgan Kaufmann Publishers, pp. 399-406, 1994.

[10] R. Kohavi and D. Sommerfield, "Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology", *Proc. of the 1st Intl Conf. on Knowledge Discovery and Data Mining,* AAAI Press, pp. 192-197, 1995.

[11] H. Liu and R. Setiono, "A Probabilistic Approach to Feature Selection: a Filter Solution", *Proc. of the 13th Intl Conf. on Machine Learning,* pp. 319-327, 1996.

[12] K. Kira and L. Rendell, "A Practical Approach to Feature Selection", *Proc. of the 9th Intl Conf. on Machine Learning, pp. 249-256,* Morgan Kaufmann Publishers, 1992.

[13] I. Kononenko, "Estimating Attributes: Analysis and Extensions of RELIEF", *Proc. of the 7th European Conference on Machine Learning, pp. 171-182,* Berlin; New York: Springer-Verlag Publisher, 1994.

[14] M.A. Hall and G. Holmes, "Benchmarking Attribute Selection Techniques for Discrete Class Data Mining", *IEEE Transactions on Knowledge and Data Engineering*, vol. 15 no. 6, pp. 1437-1447, 2003.

[15] I.H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[16] U.M. Fayyad and K.B. Irani, "Multi—Interval Discretization of Continuous-Valued Attributes for Classification Learning", *Proc. of the 13th Intl Joint Conf. on Artificial Intelligence*, Morgan Kaufmann, pp. 1022-1027, 1993.

[17] J.R. Anderson and M. Matessa, "Explorations of an Incremental, Bayesian Algorithm for Categorization", *Machine Learning*, vol. 9, pp. 275-308, 1992.

[18] P. Domingos and M. Pazzani, "Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier", *Proc. of the 13th Intl Conf. on Machine Learning*, Bari, Italy, Morgan Kaufmann, pp. 105-112, 1996.

[19] J.R. Quinlan, *C4.5: Programs for Machine Learning*, San Mateo, California: Morgan Kaufmann Publishers, 1993.

[20] S. Haykins, *Neural Networks: a Comprehensive Foundation*, Upper Saddle River, N.J., Prentice Hall, 1999.

[21] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", *Proc. of the Intl Joint Conf. on Artificial Intelligence*, Montreal, QC, Morgan Kaufmann, pp. 1137-1143, 1995.