Machine Recognition of Navel Orange Worm Damage in X-ray Images of Pistachio Nuts

Pamela M. Keagy,[1] Bahram Parvin,[2] and Thomas F. Schatzki[1]

[1]USDA-ARS Western Regional Research Center

[2]Information and Computing Sciences Division, Lawrence Berkeley Laboratory
University of California, Berkeley, California 94720

November 1994

MASTER

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Machine recognition of navel orange worm damage in x-ray images of pistachio nuts

PAMELA M. KEAGY[1], BAHRAM PARVIN[2] AND THOMAS F. SCHATZKI[1]

[1]USDA-ARS Western Regional Research Center, 800 Buchanan St., Albany, CA 94710
[2]Lawrence Berkeley Laboratory, ICSD-ITG MS 50B-2239, 1 Cyclotron Rd., Berkeley, CA 94720

ABSTRACT

Insect infestation increases the probability of aflatoxin contamination in pistachio nuts. A non-destructive test is currently not available to determine the insect content of pistachio nuts. This paper presents the use of film X-ray images of various types of pistachio nuts to assess the possibility of machine recognition of insect infested nuts. Histogram parameters of four derived images are used in discriminant functions to select insect infested nuts from specific processing streams.

Key words - pistachio nuts, discriminant functions, aflatoxin, X-ray radiograph, machine recognition.

1. INTRODUCTION

The US now produces 28% of the world's pistachio nuts and is second only to Iran in total production[1]. International markets are setting strict limits on aflatoxin (a potent natural carcinogen) contamination and US producers are seeking ways to assure that the aflatoxin content of their products is as low as possible. Previous work by Sommer[2] indicated that no aflatoxin was found in healthy nuts with intact hulls before harvest. Nuts with hulls which split before harvest were at risk for aflatoxin contamination while infestation by naval orange worm increased the probability of aflatoxin contamination. Currently, insect infested nuts are removed from the processing line when other defects are present or by manual inspectors when external evidence of infestation is present. Proposed US standards for grades of pistachio nuts restrict insect damage to 1-3% (by weight) of nuts. Grading procedures consist of opening a sample of nuts and visually inspecting for insect damage and other quality factors. A nondestructive method of inspection for insect damage is not currently available.

Internal evidence of insect infestation may be seen by visually inspecting X-ray film images of pistachio nuts. The objective of this work was to detect insect infested nuts with x-ray images and machine recognition.

2. MATERIALS and METHODS

2.1 Samples

Pistachios fresh from the field are hulled, washed and dried within hours of harvest. The first step in the sorting process occurs during washing as nuts which float (floaters) and nuts which sink (sinkers) are kept separate in subsequent drying and sorting operations. The nuts are dried to approximately 6% moisture and then stored for later sorting and roasting. Following storage small nuts, meats and twigs are removed by a scalper. Remaining nuts are passed through an air separator to remove empty and broken shells; a needle picker to separate open and closed shell nuts; and a color sorter (to separate clean, lightly stained (dye stock) and dark stained nuts (shelling stock)). Inspectors manually remove defects (including insect damaged nuts) missed by the previous automatic sorters, and a succession of grates separates meats, small, large and extra large nuts. This process is diagramed in figure 1. Table 1 presents abbreviations and descriptions of the processing streams generated during sorting of pistachio nuts. Samples were obtained from a commercial processor representing all processing streams from the 1992 crop except one. All nuts of the manually removed (HPO, hand pick out) insect - floater stream had been destroyed. Samples from each stream obtained were analyzed for aflatoxin by HPLC.

2.2 Images

Nuts (300 large sinkers, 450 HPO insects, and 150 each from other process streams) from each process stream were individually arrayed on contact paper in one of three orientations (suture plane parallel, perpendicular or at an angle to the film plane) and X-rayed (90 sec at 25 keV [0.25 mm Be window) with a Faxitron series X-ray system 4380N, Hewlett Packard, McMinnville, OR; Industrex B film, Eastman Kodak, Rochester, NY). Films were handled in the dark and exposed without film holders to avoid background patterns caused by the film holders at low X-ray energies. Twelve bit digital images were obtained from the films at a resolution of $(0.125mm)^2$/pixel using a Lumiscan 200 film scanner (Lumisys, Sunnyvale, Ca.). Nuts were then opened to determine presence or absence of insect damage.

Each film contained a 25 step plastic wedge as an exposure standard. The wedge was constructed to provide absorption levels comparable to those found in pistachio nuts. The optical density response of the film was sigmoidal when plotted against the thickness of the step wedge. The variance of the optical density was an inverse function of the optical density. A logit transform was used to linearize the response and equalize the variance. This transform made the density change caused by an insect tunnel more equally visible in the thick and thin parts of the pistachio. The transform also matched the form of the response to the output of a prototype linescan X-ray machine which produces realtime images at $(0.5mm)^2$/pixel. Small variations between film exposures were removed by linearly rescaling each film to a background value of 5 and step 20 of the wedge to 196 on a scale of 0 to 255. Pixel averaging was used to obtain lower resolution images corresponding to $(0.5 mm)^2$/pixel. Nut images at the edge of the X-ray beam were excluded from the image data set to avoid artifacts.

Fig. 2 contains representative images of an insect infested and a good nut from the color sorter reject floater process stream. A good nut is characterized by a bright area of nutmeat surrounded by a less bright cross-section of shell. Darker areas occur indicating space between the shell and meat and occasionally between the two halves of the nutmeat. The normal areas of darkness are generally characterized by sharp edges. In insect damaged nuts, an additional area of darkness is frequently seen which corresponds to tunnels formed by the feeding insect. This erodes the amount of bright area and increases the amount of medium density dark area. Edges created by insect activity are generally less sharp than natural edges. These characteristics suggested that histograms of intensity and derived edge images could be used to select insect infested nuts from good nuts.

2.3 Image Processing

Nuts were segmented from the background and the intensity histogram computed for each nut. These histograms were characterized by their statistical parameters of number of pixels or area, intensity mean, variance, skewness and kurtosis. Further useful histogram statistics were obtained from four derived images consisting of: 1) edge image - maximum slope across each pixel; 2) curvature image - based on first and second fundamental forms (details are given in Appendix 1) ; 3) difference of Gaussians (DOG) image (excitatory sigma=1, inhibitory sigma=1.6) and 4) DOG edge image - maximum slope across each pixel in the DOG image. The HIPS image processing package (SmartImage Software, New York) running on a Sun workstation was used to calculate the image features. Each of these images is illustrated in figure 2 for one good nut and one insect infested nut from the color sorter reject floater process stream.

2.4 Discriminant functions

Very limited numbers of insect damaged nuts were found in processing streams other than the HPO insect sinkers. Known insect infested images from the HPO insect sinkers were combined with images from one other process stream to form the data set for each discriminant function. Discriminant functions were not computed for the air separated sinkers and floaters because very little edible product was found in these process streams and the image characteristics of the broken and empty shells were very different from a normal inshell pistachio nut. This resulted in 14 combinations of data for discriminant functions.

A stepwise discriminant procedure (SAS[3] STEPDISC) was used to select groups of histogram features which were significant at $P<0.05$. The resulting features (usually six or less) were used to calculate linear discriminant functions (SAS DISCRIM procedure) which would identify insects in each processing stream. The discriminant functions were evaluated both by splitting the data into training and test sets and by using the crossvalidation method. Results were similar with either approach. The squared canonical correlation coefficient indicates the ability of a function to identify infested nuts. Higher values indicate better separations as illustrated in the operating characteristic functions of small sinkers and color sorter reject floaters.

3

2.5 Operating characteristic curves

The crossvalidate option of SAS DISCRIM computed the posterior probability of infestation for each nut. Relative operating characteristic curves were computed from the posterior probability data using the program LABROC1 (available by anonymous FTP, from C. E. Metz[4], University of Chicago, ftp random.bsd.uchicago.edu). Depending on the value of the nuts and the frequency of infestation, the tolerance for false positives and false negatives may be very different in different streams.

3. RESULTS AND DISCUSSION

Table 2 shows the data for percent of product, aflatoxin concentration and insect occurrence for those process streams with measurable aflatoxin and one of the major product streams which was free of aflatoxin. For all processing streams, correlation of aflatoxin concentration (ng/gm) and insects per 100 nuts is 0.44 (p=0.0756, n=17 product streams). While discriminant functions were calculated using all but two product streams, detailed results will be presented for only two as examples. Color sorter reject floaters comprise 1.6% of the crop and contain 14 ng/gm aflatoxin and 6 insects per 100 nuts. Small sinkers contain 97 ng/gm aflatoxin and 13 insects per 100 nuts. Notice that the major product stream, large sinkers, had no detectable aflatoxin and no detectable insects. It would not be cost effective to sort this stream for insects.

Table 3 presents the histogram features used to select insect infested nuts among color sorter reject floaters. Six of the available features were significant at the 0.05 level and the squared canonical correlation was 0.48. Table 4 illustrates what this means for selection of insect infested nuts. If a critical value (posterior probability from SAS DISCRIM output) of at least 0.54 is required to label a nut as infested, 83% of the infested nuts (true positives) will be found with 17% of the non-infested nuts also being classified as infested (false positives). This might be an acceptable level of false positives when the selection is only being applied to 1.6% of the crop and this fraction is of lower quality and value. At a critical value of 0.35, 90% of the insects can be eliminate at a cost of 28% of the product. The selection of a specific critical value must be left to the user where cost and regulatory constraints are known.

Table 5 shows the histogram features used to select insect infested nuts among small sinkers with a squared canonical correlation of 0.67. Again six of the twenty features are significant at the 0.05 level but only one of the features was also used in the color sorter reject floater function. The higher canonical correlation indicates better discriminating power as illustrated in Table 6. Here a critical value of 0.858 selects 85% of the insects at a cost of 6% of the good product. Table 7 summarizes the squared canonical correlations of discriminant functions for those process streams which contained measurable aflatoxin. Canonical correlations ranged from 0.40 to 0.67 indicating performance characteristics similar to the two examples discussed above.

Three feature variables were included in 10 or more of the discriminant functions calculated for 14 processing streams. (Functions were not computed for two processing streams which contained very little edible product, air leg light sinkers and floaters.) First among these was an area measure from any of the images. Average nut area for process streams containing aflatoxin in shown in Table 7. Insect infested nuts tend to be smaller than non-infested nuts in the same process stream. Kurtosis of the curvature image was also important. Most of the information in this image is found in the extreme black and white pixels indicating ridges and troughs in the nut. Insects create troughs as they feed on the nuts. Kurtosis is derived from the fourth moment of the distribution so it would be reasonable to expect this statistic to indicate differences in the extreme values. The variance of the DOG-edge image was also an important indicator of insect presence. This may be another indicator of the amount of damage created by the insect in an otherwise smooth nutmeat.

4. SUMMARY

It is possible to identify many, but not all, insect infested nuts with X-ray images and machine recognition. Engineering, economic and regulatory factors need to be considered in choosing the optimum set of image features and decision criteria.

5. ACKNOWLEDGEMENTS

6. REFERENCES

1. California Pistachio Commission, California Pistachio Industry Annual Report Crop Year 1992-93 p46, 1993.

2. Sommer, N. F., Buchanan, J. R. and Fortlage, R. J. "Relation of early splitting and tattering of pistachio nuts to aflatoxin in the orchard". *Phytopathology* 76 692-694,.1986.

3. SAS Institute. SAS/STAT user's Guide. Release 6.03 edition. The Institute: Cary NC, 1988.

4. Metz, C. E. , Shen J-H, Herman BA, "New methods for estimating a binormal ROC curve from continuously-distributed test results." Invited for presentation at the 1990 Joint Statistical Meetings of the American Statistical Society and the Biometric Society; Anaheim, CA, August 1990.

5. Faux, I. D. and Pratt, M. J. "Computational geometry for design and manufacturing." John Wiley & Sons, New York, 1983

6. Parvin, B. and Medioni G. "B-rep from unregistered multiple view range images.", IEEE Conference on Robotics and Automation, May 1992.

7. APPENDIX

Curvature is a measure of the differential surface property of the local intensity distribution, and it is computed from the first and second fundamental forms. These forms uniquely determine certain local invariance quantities of a 3D surface, where, for X-ray images, invariance is expressed in terms of translation, rotation, and scaling. Faux and Pratt[5] expressed the first and second fundamental form from differential geometry in parametric space. In this section we use their notation to develop these forms in the Cartesian space. Let a point on the surface be defined as $P = x\vec{i} + y\vec{j} + z(x,y)\vec{k}$, then the first and second fundamental forms in the Cartesian system are computed to be:

$$G = \begin{bmatrix} 1 + \left(\frac{\partial z}{\partial x}\right)^2 & \frac{\partial z}{\partial x}\frac{\partial z}{\partial y} \\ \frac{\partial z}{\partial x}\frac{\partial z}{\partial y} & 1 + \left(\frac{\partial z}{\partial y}\right)^2 \end{bmatrix}$$

$$D = \frac{1}{\vec{n}} \begin{bmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial z}{\partial x}\frac{\partial z}{\partial y} \\ \frac{\partial z}{\partial x}\frac{\partial z}{\partial y} & \frac{\partial^2 z}{\partial y^2} \end{bmatrix}$$

The normal curvature of a surface is the curvature of the intersecting curve between the surface and the plane containing the surface normal and tangent vector to the curve. The directions in which the normal curvature becomes maximum or minimum are called principal directions corresponding to the principal curvatures. The normal curvature is defines as:

$$K_n = \frac{\dot{X}^T D \dot{X}}{\dot{X}^T G \dot{X}}$$

Through elimination and solving a pair of simultaneous equations, the following quadratic equation is obtained, where the roots of this equation correspond to maximum and minimum principal curvatures.

$$(g_{11}g_{22} - g_{12}g_{21})k_n^2 - (g_{11}d_{22} + d_{11}g_{22} - 2g_{12}d_{12})k_n + (d_{11}d_{22} - d_{12}d_{21}) = 0$$

To construct the curvature histogram, the frequency of occurrence of each curvature value over the entire image is tabulated.

Table 1. Descriptions and abbreviations for pistachio nut processing streams.

Process stream abbreviation and number of images used S=Sinkers F=Floaters		Sample description
Small-S	73	Small nuts, sinkers, removed by scalper
Small-F	40	Small nuts, floaters, removed by scalper
Air-S	0	Broken shells, sinkers, removed by the air separator, very little edible product, discriminant function not computed
Air-F	0	Empty shells, floaters, removed by the air separator, very little edible product, discriminant function was not computed
Needle-S	115	Closed shell nuts. sinkers, no aflatoxin or insects were found
Needle-F	96	Closed shell nuts, floaters, no aflatoxin or insects were found
Color Reject-S	107	Nuts with stained shells, sinkers, rejected by automatic color sorters
Color Reject-F	90	Nuts with stained shells, floaters, rejected by automatic color sorters
HPO Dye-S	115	Lightly stained nuts manually removed by inspectors, sinkers, intended to be marketed as red dyed nuts.
HPO Dye-F	90	Lightly stained nuts manually removed by inspectors, floaters, intended to be marketed as red dyed nuts.
HPO Shell-S	105	Dark stained nuts manually removed by inspectors, sinkers, intended for shelling and sale as meats
HPO Shell-F	92	Dark stained nuts manually removed by inspectors, floaters,, intended for shelling and sale as meats
HPO Insect-S	150	Nuts manually removed by inspectors due to suspected insect damage, sinkers,
HPO Insect-F	0	Nuts manually removed by inspectors due to suspected insect damage, floaters, samples of these nuts were not available
Large-S	214	Good large nuts, sinkers
Large-F	116	Good large nuts, floaters
Extra large-S	109	Good extra large nuts, sinkers
Extra large-F	105	Good extra large nuts, floaters

TABLE 2. Relationship of aflatoxin and insects in pistachio nuts.

PRODUCT STREAM	% OF TOTAL PRODUCT	AFLATOXIN NG/GM	AFLATOXIN % OF CROP TOXIN	INSECTS per 100 NUTS
HPO Insect-S	0.89	89	37	44
Small-S	0.53	97	24	13
Small-F	0.10	190	11	15
Color Reject-F	1.55	14	10	6
HPO Dye-F	0.13	91	9	9
Color Reject-S	10.91	1.4	7	2
Large-F	0.44	15	3	1
Large-S	31.06	0	0	0

Table 3. Histogram features used to select insect infested nuts among color sorter reject floaters.

Statistic	Image
Area	Difference of Gaussians (DOG) Image
Kurtosis	Curvature Image
Variance	DOG Edge Image
Kurtosis	DOG Image
Skewness	DOG Edge Image
Kurtosis	Edge image
Squared Canonical Correlation = 0.48	

Table 4. Discriminant function values for selection of insect damaged nuts from normal color sorter reject floater pistachio nuts. The critical value is the posterior probability of infestation obtained for each nut from the discriminant function. A false positive is a non-infested nut labeled as infested and a true positive is an infested nut labeled as infested.

Critical Value	False Positive	True Positive
0.973	0.001	0.195
0.912	0.009	0.389
0.836	0.033	0.575
0.706	0.084	0.723
0.542	0.171	0.834
0.354	0.278	0.902
0.168	0.435	0.953
0.078	0.616	0.981
0.027	0.809	0.995

Table 5. Histogram features used to select insect infested nuts from normal small sinker pistachio nuts.

Statistic	Image
Area	Edge Image
Skewness	Curvature Image
Variance	DOG Edge Image
Mean	DOG Image
Skewness	DOG Image
Mean	Edge Image
Squared Canonical Correlation = 0.67	

Table 6. Discriminant function values for selection of insect infested nuts from normal small sinker pistachio nuts. The critical value is the posterior probability of infestation obtained for each nut from the discriminant function. A false positive is a non-infested nut labeled as infested and a true positive is an infested nut labeled as infested.

Critical Value	False Positive	True Positive
0.999	0.000	0.098
0.998	0.001	0.300
0.993	0.005	0.495
0.975	0.017	0.686
0.858	0.055	0.853
0.298	0.138	0.946
0.149	0.312	0.988
0.077	0.498	0.997
0.002	0.701	1.000

Table 7. Average nut area and canonical correlations of discriminant functions for process streams which contained measurable aflatoxin.

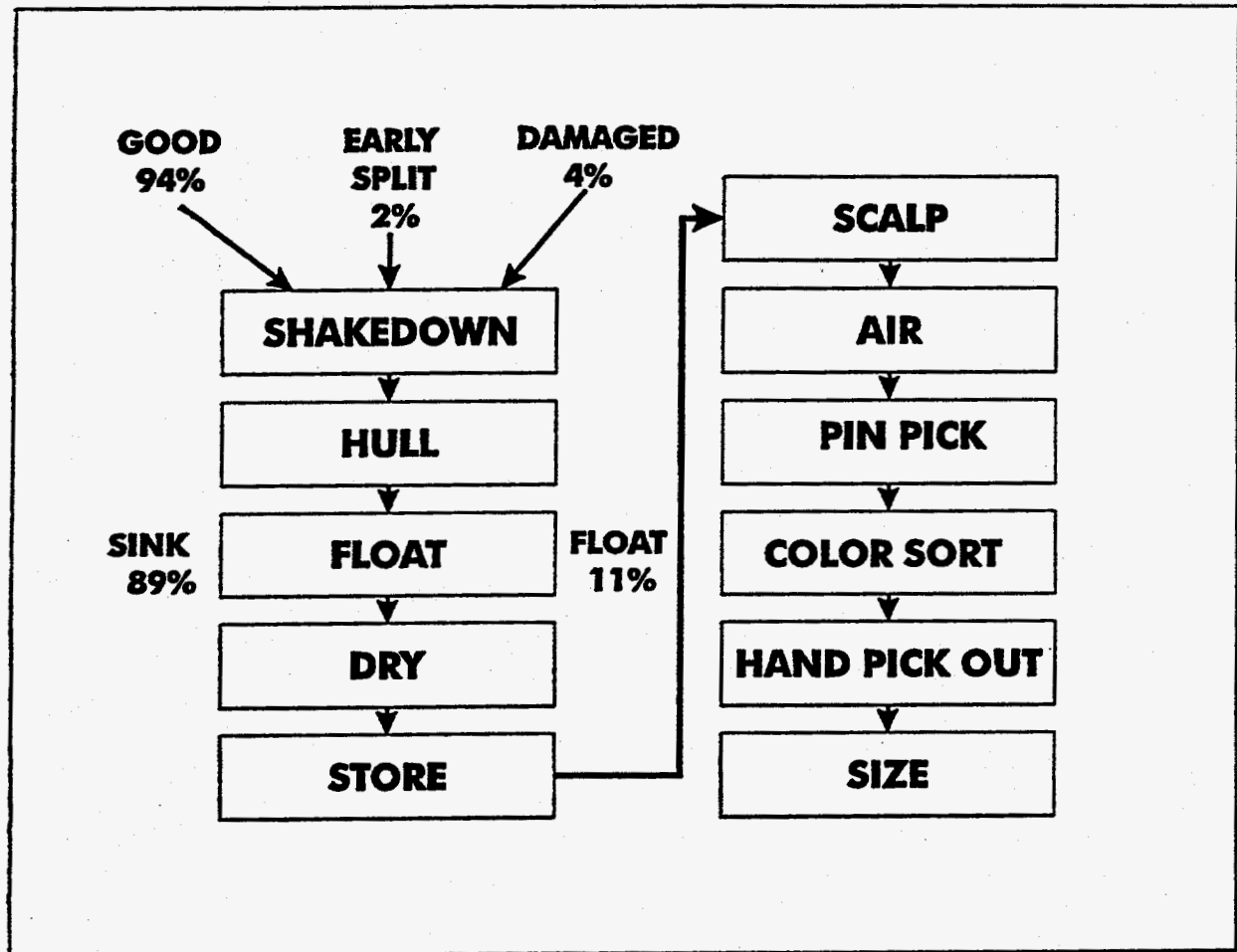| Product Stream	% of Total Product	Canonical Correlation	Area Pixels	Insects per 100 nuts
HPO Insect-S	0.9		650 (72)	44
Small-S	0.5	0.67	450 (70)	13
Small-F	0.1	0.60	602 (109)	15
Color Reject-F	1.6	0.48	737 (101)	6
HPO Dye-F	0.1	0.60	630 (89)	9
Color Reject-S	10.9	0.43	718 (82)	2
Large-F	0.4	0.52	764 (70)	1
Large-S	31.1	0.40	673 (56)	0

Figure 1. Processing steps used during harvesting and sorting of pistachio nuts.

Figure 2. Upper left pair - 0.125mm/pixel X-ray images of insect damaged (left) and good (right) pistachio nuts from the color sorter reject floater process stream; Middle left pair - same nuts with resolution reduced to 0.5mm/pixel by pixel averaging; lower left pair - difference of Gaussians images computed from 0.5mm/pixel images above; upper right pair - curvature images derived using first and second fundamental forms from middle left pair; middle right pair - edge image (maximum first derivative across pixel) from middle left pair; lower right pair - DOG edge image derived from DOG image in lower left pair.
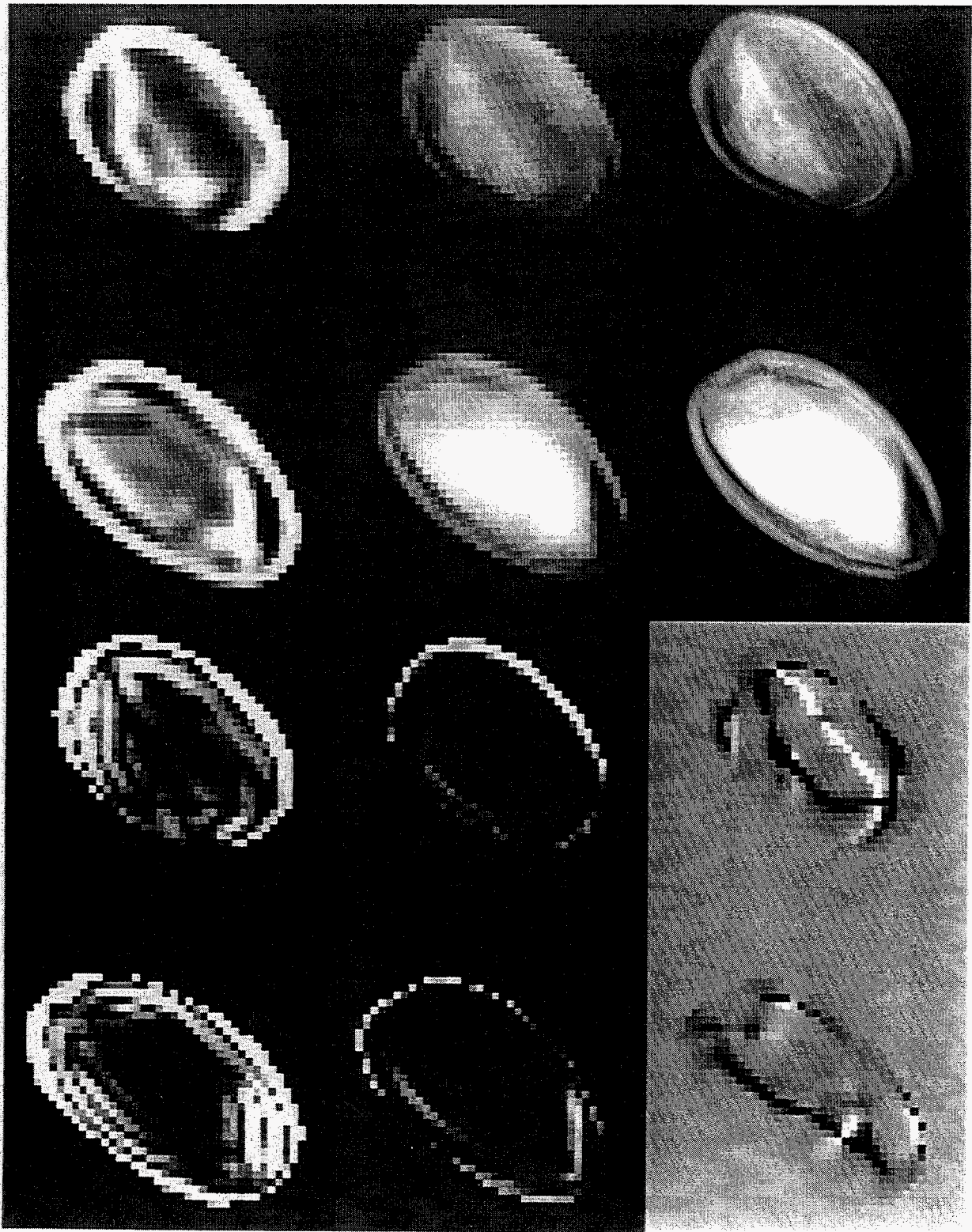
Figure 2