01 Jan 1995

Performance of AI Methods in Detecting Melanoma

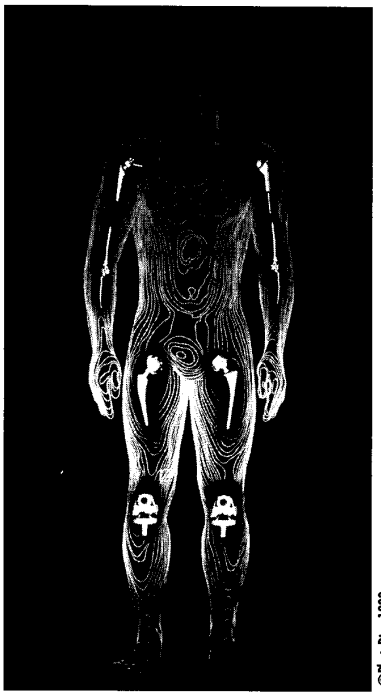
Arve Kjoelen

M. J. Thompson

Scott E. Umbaugh

Randy Hays Moss
Missouri University of Science and Technology, rhm@mst.edu

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/ele_comeng_facwork/1818



Performance of AI Methods In Detecting Melanoma

Arve Kjoelen[1], Marc J. Thompson[1], Scott E. Umbaugh[1], Randy H. Moss[2], and William V. Stoecker[3]

[1]Department of Electrical Engineering
Southern Illinois University at Edwardsville
[2]Department of Electrical Engineering
University of Missouri, Rolla
[3]Department of Computer Science
University of Missouri-Rolla and Department of Internal Medicine
University of Missouri-Columbia

The incidence of malignant melanoma — the deadliest form of skin cancer — is now more than 15 times higher than it was in the 1930s [1, 2, 3]. Medical costs are soaring, and skin biopsies have become the most frequently reimbursed Medicare procedure [4]. When diagnosed in the early stages, melanoma is relatively easy to treat, and patients show survival rates near one hundred percent [1, 5]. Automated diagnosis, if deemed feasible, may increase the chances of early detection and lower the cost of unnecessary biopsies. Even if the success rate is not sufficiently high for automated diagnosis, this tool could prove a useful adjunct in the screening of skin tumors on a mass scale.

Computer vision methods have been previously applied to the problem of skin tumor diagnosis [6, 7, 8, 9, 10, 11, 12]. In general, computer vision methods are used to find tumor borders, segment out the tumor from the rest of the image, and extract features from the tumor. Then, automatic induction or other methods are used to generate a classification rule based on the extracted features. Next, the classification rule is tested on a different set of images to determine its accuracy. This computer vision front-end provides the necessary inputs to an expert system that is used for diagnosis. In this article, we focus on artificial intelligence (AI) classification methods to differentiate melanoma from non-melanoma.

Materials and Methods

Images

The images used in this research were digitized by researchers at the University of Missouri-Rolla from 35-mm color slides obtained from a private dermatology practice and from the New York University Department of Dermatology. The digital images had spatial resolution of 512x512 pixels, a brightness resolution of 256 levels per color plane (8 bits), and consisted of three color planes (red, green and blue) for a total of 24 bits per pixel. The set of images used here consisted of 92 melanoma images and 169 benign images, for a total of 251 images.

Software

The software for this research was developed in the standard (ANSI) C programming language on a SUN workstation operating under the Sun OS operating system. Classification methods for the skin tumors were generated by the 1st-Class automatic induction software [13] and by the AIM [14] numeric modeling tool, both operating on an IBM compatible personal computer.

Automatic Induction

Induction is the process of producing a general classification algorithm from a set of specific examples [12]. The mechanism used by 1st-Class is based on an algorithm known as ID3 [14]. The ID3 algorithm is the induction engine of the 1st-Class software, and operates by generating decision trees based on input examples [14]. A representative, albeit short, decision tree coded in the C programming language is shown in Fig. 1. The AIM induction tool, on the other hand, bases its analysis on mathematical models known as polynomial networks. A polynomial network combines the neural network concept with statistical regression techniques [14], and consists of a network of functional nodes that compute an output function based on a number of inputs. For this experiment, the output of AIM was a number between 0 and 1, with 0 considered non-melanoma and 1 considered melanoma. Figure 2 gives an example of an AIM polynomial equation. Both 1st-Class and AIM can generate source code in the C programming language. This code was incorporated into the software developed to classify the skin tumors.

Features

The set of 16 features used in this experiment were extracted from the digitized color images using computer vision techniques. In the cases of elevation and area, the feature was manually estimated by a dermatologist as being present or not present. The following features were used.

Irregularity — a measure of the irregu-

 0739-5175/95/$4.00©1995

```
#define BENIGN 0
#define MELANOMA 1
int analyze(double Irregularity, double Asymmetry, int Red_variance, int Green_variance,
int Blue_variance, double Red_rel_chroma, double Green_rel_chroma, double
Blue_rel_chroma,
  int Sphericlength, int SphericangleA, int SphericangleB, int Lightness,
  int Chromaticity, int Hue, int Elevation, int Area)  { if (Area 0.50) {
   if (Irregularity < 1.27) {
      if (Red_variance 33.50)
         return(BENIGN);
      else if (Red_variance > = 33.50)
         return(MELANOMA);
   }
   else if (Irregularity >= 1.27)
      return(MELANOMA);
 }
 else if (Area = 0.50) {
   if (SphericangleA < 71.00)
      return(BENIGN);
   else if (SphericangleA >= 71.00) {
      if (Red_variance < 9.50) {
         if (Hue > 33.00)
            return(BENIGN);
         else if (Hue = 33.00)
            return(MELANOMA);
      }
      else if (Red_variance >= 9.50) {
         if (Irregularity < 1.14) {
            if (Red_rel_chroma < 0.12)
               return(MELANOMA);
            else if (Red_rel_chroma >= 0.12) {
               if (Lightness < 55.50) {
                  if (Irregularity < 1.13)
                     return(MELANOMA);
                  else if (Irregularity >= 1.13)
                     return(BENIGN);
               }
               else if (Lightness >= 55.50) {
                  if (Irregularity < 1.09)
                     return(BENIGN);
                  else if (Irregularity >= 1.09)
                     return(MELANOMA);
               }
            }
         }
         else if (Irregularity >= 1.14)
                  return(MELANOMA);
      }
   }
 }
}
```

1. A decision tree produced by 1st-Class coded in the C programming language.

larity of the tumor border [9]. This feature is estimated by the ratio of the square of the perimeter to the area of the tumor;

Asymmetry — a measure of the asymmetry of the tumor, estimated by the nonoverlapping areas after an imaginary "folding" operation along the best axis of symmetry[8];

Variance of the red, green and blue color components in the tumors -an indication of the tumor texture;

Relative chromaticity of the three color components — The intensity of the three color components in the tumor relative to the color components of the surrounding skin [10];

Spherical coordinates — A representation of color developed specifically for detection of variegated coloring in skin tumors [10]: with length (intensity) and two angles representing the relative content of the three colors;

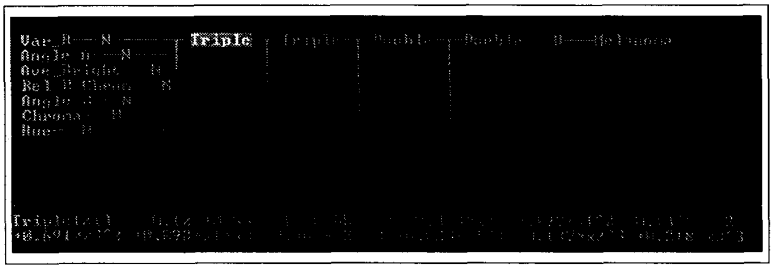
IHS coordinates — A representation of color consisting of lightness (brightness), chromaticity (color/wavelength), and saturation (amount of white in, or impurity of color) [10, 12];

Elevation — A feature indicating whether or not the tumor was elevated by 2 mm or moreover the surrounding skin, estimated by a dermatologist;

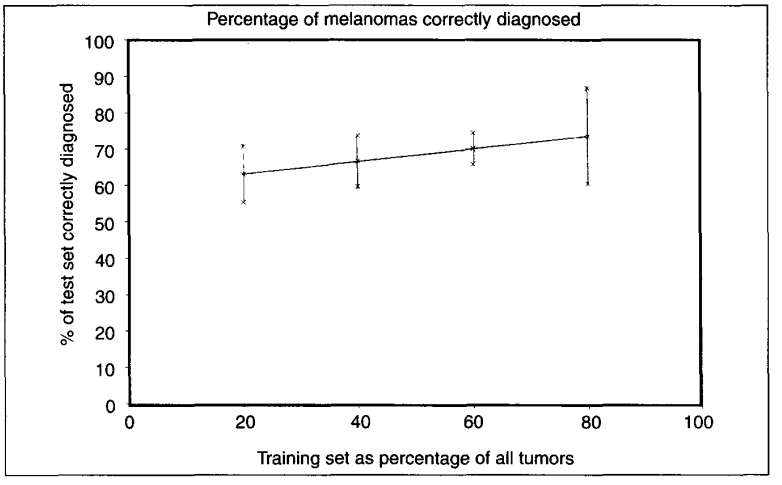
Area — A feature indicating whether or not the greatest tumor diameter exceeded 6 mm.

Training/Test Set Paradigm

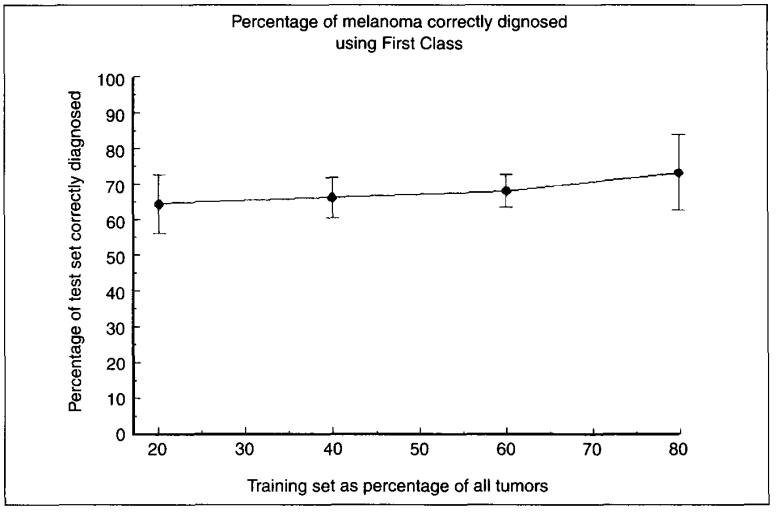
In statistical studies, data are often separated into two sets. One set is used for training or developing the algorithms, and the other is used for testing the algorithms developed. This procedure allows for unbiased results from the test set [10]. The problem of selecting the training and test sets is complex. In order to develop the best algorithm possible, the size of the training set should be maximized; but in order to have confidence in the results, the test set should be made as large as possible. In this study, the sizes of the training and test sets were varied. In addition, ten experiments were run for each combination of training and test set sizes. The training sets were assigned a number according to their size. For instance, of the ten training sets that consisted of sixty percent of the entire tumor set, the fifth set would be numbered 6005. This number was entered as a seed to the C rand function, which returned a random sequence of integers. These integers were scaled to the range 0-250, and used as indices into a list of all



2. An example of the polynomial network generated by AIM. The input variables are on the left side. The N represents a normalizing factor for the input. The triple and double are nodes in the network. The formula of the highlighted triple is shown at the bottom of the figure. X1, X2, and X3 are the first three input variables.



3. Percentage of melanomas correctly diagnosed as a function of training set size using 1st-Class. Results of part one of the experiment, atypical moles are included. The vertical lines represent one standard deviation.



4. Percentage of melanomas correctly diagnosed as a function of training set size using 1st-Class. Results of part one of the second experiment, atypical moles are included. The vertical lines represent one standard deviation.

251 tumors. The tumors corresponding to the indices were then selected for inclusion in the training set. The use of multiple experiments for each combination of training and test set sizes yielded statistical information regarding result validity.

Two two-part experiments were performed with each AI tool. Part one of each experiment consisted of applying 1st-Class and AIM towards the entire set of 251 images, while for part two, a subset of 91 images was excluded from the experiment. The excluded images were atypical moles — a benign type of skin lesion. The atypical moles were excluded because it was suspected that an AI tool would have trouble distinguishing them from malignant melanomas. This group of benign but atypical melanocytic nevi (moles) is undergoing an evolution in classification by dermatologists and dermatopathologists. Some uncertainty still exists regarding definition and clinical behavior of this group.

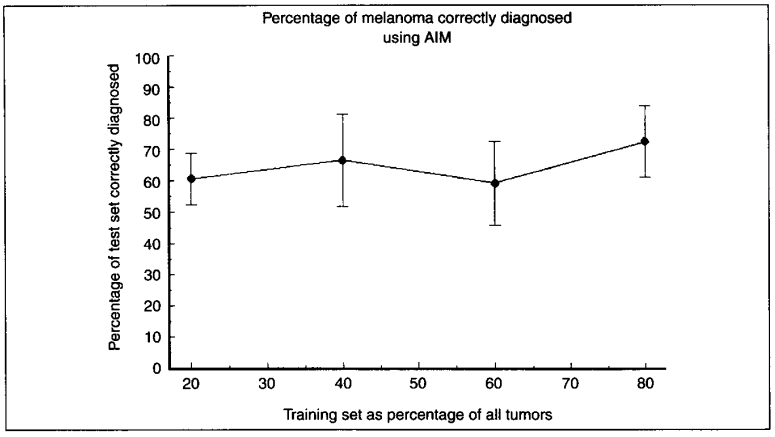
For the first experiment with AIM, we chose to use a cutoff of 0.5, so that output values greater than or equal to 0.5 were interpreted as melanoma, while a value below 0.5 indicated a benign tumor. In the second experiment with AIM, this threshold was lowered to 0.25. All other parameters, including the training and test sets, remained unchanged.

The second experiment with 1st-Class was performed to verify the statistical significance of the first experiment. A slightly different method was used to generate the training and test sets to ensure that they would be different from those of the first experiement.

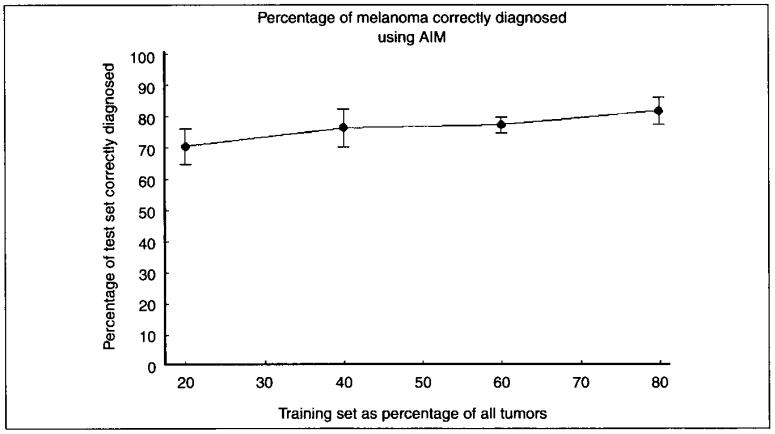
It should be noted that the ability to diagnose correctly melanoma is by far the most important property that an automated system must have. The consequence of failure to diagnose correctly a malignant tumor may lead to the eventual death of the patient. On the other hand, misclassifying a benign tumor as malignant will cause temporary and comparatively insignificant emotional distress to the patient. We refer to this fact by stating that the cost of misclassifying melanoma is much higher than the cost of misclassifying a benign tumor.

Results

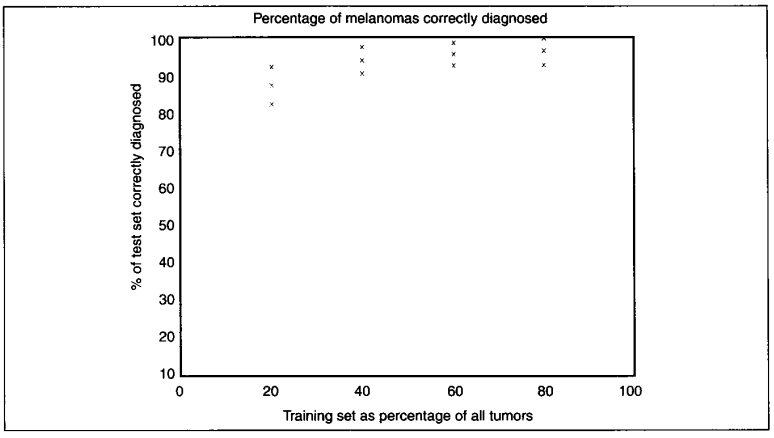
The success rates using 1st-Class to diagnose melanoma in the test images in the first and second experiments are depicted in Figs. 3 and 4, respectively. The figures show a plot of the average success rate from ten randomly selected training



5. Percentage of melanomas correctly diagnosed as a function of training set size using AIM. Results of part one of the experiment, atypical moles are included with a threshold of 0.5. The vertical lines represent one standard deviation.



6. Percentage of melanomas correctly diagnosed as a function of training set size using AIM. Results of part one of the experiment, atypical moles are included with a threshold of 0.25. The vertical lines represent one standard deviation.



7. Percentage of melanomas correctly diagnosed as a function of training set size using 1st-Class. Results of part two of the experiment, atypical moles are excluded. The vertical lines represent one standard deviation.

and test sets for different set sizes. One standard deviation above and below the mean is indicated by the vertical lines in each figure. In both experiments, moderate success was achieved, with 70 percent of the malignant tumors correctly classified when the size of the training set was 60 percent. The standard deviation was also low — less than 5 percent. Equivalent success rates using AIM are shown in Figs. 5 and 6. In Fig. 5, the diagnostic threshold on the AIM output was 0.5, while the threshold was 0.25 in Fig. 6. As expected, the number of melanomas correctly identified increased when the threshold was lowered. AIM performed better than 1st-Class when using a threshold of 0.25, but had lower accuracy when a 0.5 threshold was employed.

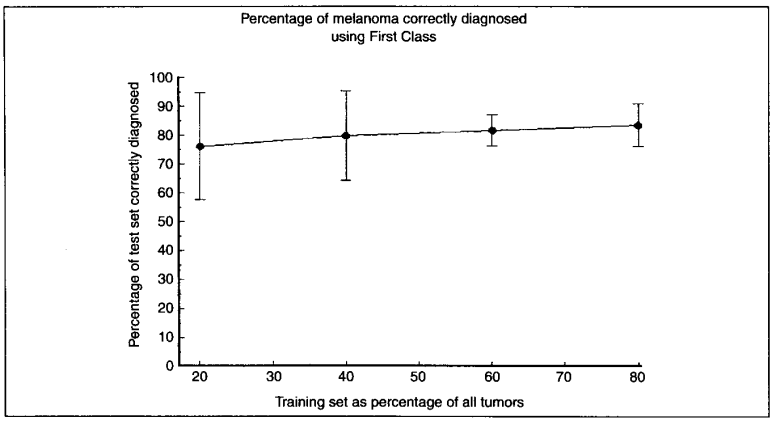
When the atypical moles were excluded, the diagnostic accuracy of both AI tools increased. Figures 7 and 8 depict the results obtained with 1st-Class. Figure 7 indicates accuracy rates as high as 95 percent were obtained when 60 percent of the images were used for training. The fact that the two 1st-Class experiments illustrated in Figs. 7 and 8 resulted in slightly different accuracy rates (about 10 percent, on average) serves to illustrate the somewhat unpredictable nature of AI tools.

Figures 9 and 10 show the increased success rates using AIM for the case when the atypical moles were excluded. For Fig. 9, the output threshold was 0.5, while for Fig. 10 the threshold was 0.25.

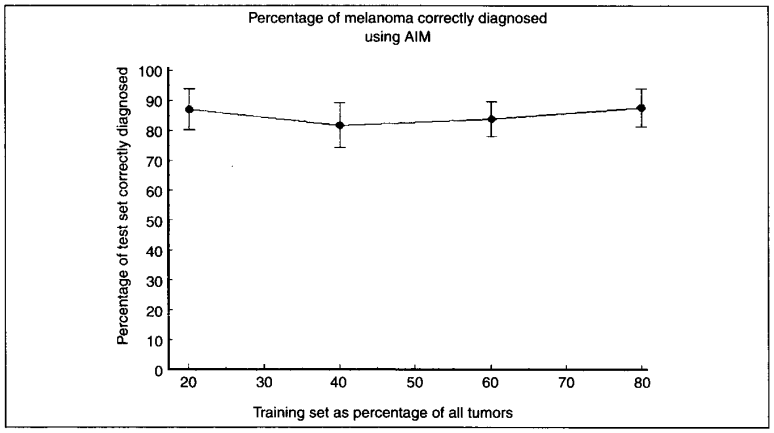
Discussion

As Fig. 3 indicates, the standard deviation for a training set size of 80 percent using 1st-Class tops out at about 10%. The relatively consistent results and small standard deviations obtained in the experiments, lead us to conclude that our results are representative of those that would be obtained in an actual application. Overall, 1st-Class produced results with smaller standard deviations than did AIM. We think this indicates greater reproducibility and possibly greater reliability on the part of 1st-Class.

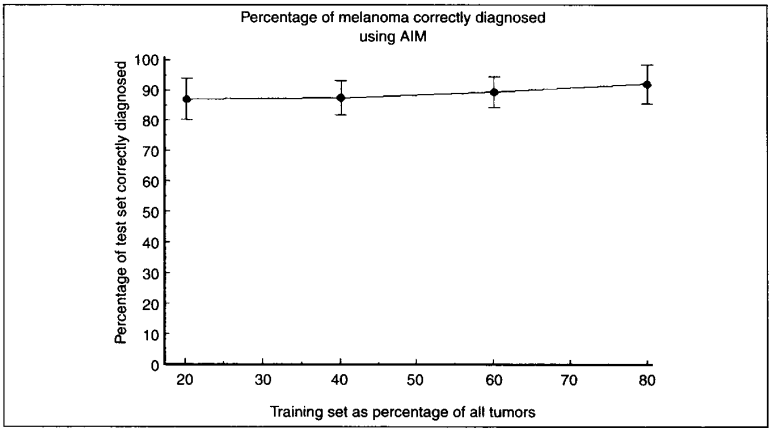
The increased accuracy for part two of the experiment indicates that the presence of atypical moles confused the automatic induction mechanism, as suspected. This finding is important, because it means that the diagnosis ratio can be significantly improved if an effective method can be found to rule out atypical moles.



8. Percentage of melanomas correctly diagnosed as a function of training set size using 1st-Class. Results of part two of the second experiment, atypical moles are excluded. The vertical lines represent one standard deviation.



9. Percentage of melanomas correctly diagnosed as a function of training set size using AIM. Results of part two of the experiment, atypical moles are excluded with a threshold of 0.5. The vertical lines represent one standard deviation.



10. Percentage of melanoma correctly diagnosed as a function of training set size using AIM. Results of part two of the experiment, atypical moles are excluded with a threshold of 0.25. The vertical lines represent one standard deviation.

Features

The 1st Class automatic induction software used only a subset of the 16 available features to generate each classification rule. The most frequently used features were asymmetry, irregularity, hue, and area. These features appeared in about 90% of the decision trees. Less often used were the red variance, the length (from spherical coordinates), angle a (from spherical coordinates) [10,13], and the green variance. Blue variance, relative chromaticity, and lightness were used in about 35% of the classification rules. The other features were seldom or never used. Three of the four most frequently used features were high-level features. This may indicate that some of the low-level features should be discarded or modified in future experiments. The features most often used by 1st-Class are considered by experts to be reliable indicators of melanoma [1, 2, 4].

Because AIM works by forming a polynomial network, the analysis of AIM is not as straightforward as with a decision tree. We have indications, however, that AIM places less weight than 1st-Class on features generally considered to be reliable indicators of melanoma, and that this may account for the large standard deviations resulting from some of the AIM networks.

Conclusions

This research has shown that features extracted from color skin tumor images by computer vision methods can be reliable discriminators of malignant tumors from benign ones. Reliability was demonstrated by the monotonically increasing success ratios with increasing training set size and by the small standard deviations from the mean success rates. An average success rate of 70 percent in diagnosing melanoma was attained for a training set size of 60 percent.

The presence or absence of atypical moles in the training and test sets was shown to have a dramatic impact on the effectiveness of the generated classification rules. This was the case with both AIM and 1st-Class, and indicates a high potential for success if a method can be found for discriminating between atypical moles and melanoma.

Acknowledgments

This research was supported in part by NIH grant #R43 CA60294-0, SIUE Funded University Research Award #F-EN411 and an SIUE Research Equipment and Tools Award.

References

1. **Friedman R, Rigel D, Kopf A:** Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. CA 35:130-151; 1985.

2. Kopf AW, Rigel DS, Friedman RJ: The rising incidence and mortality rate of malignant melanoma. *J. Dermatol. Surg. Oncol.* 8: 760-761; 1982.

3. **Silverberg E:** Cancer statistics, 1985. CA 35:19-35; 1985.

4. **Stoecker WV, Moss RH:** Digital imaging in dermatology. Editorial, *Computerized Medical Imaging and Graphics*. 16: 145-150; 1992.

5. **Koh NH, Lew RA:** Prout, M.N. Screening for melanoma/skin cancer: theoretic and practical considerations. *J. Am. Acad. Dermatol.* 20: 159-172; 1989.

6. **Umbaugh SE, Moss RH, Stoecker WV:** An automatic color segmentation algorithm with application to identification of skin tumor borders. *Computerized Medical Imaging and Graphics*. 16: 163-177; 1992.

7. **Dhawan AP, Sicsu A:** Segmentation of images of skin lesions using color and texture information of surface pigmentation. *Computerized Medical Imaging and Graphics*. 16: 163-177; 1992.

8. **Stoecker WV, Li WW, Moss R.H:** Automatic detection of asymmetry in skin tumors. *Computerized Medical Imaging and Graphics*. 16: 191-197; 1992.

9. **Golston JE, Stoecker WV, Moss RH:** Automatic detection of irregular borders in melanoma and other skin tumors. *Computerized Medical Imaging and Graphics. 16:* 199-203; 1992.

10. **Umbaugh SE:** Computer Vision in Medicine: Color metrics and image segmentation methods for skin cancer diagnosis. Ph.D. Dissertation, Electrical Engineering Department, University of Missouri-Rolla. UMI Dissertation Services. Ann Arbor, MI; 1990.

11. Umbaugh SE, Moss RH: Stoecker, W.V: Automatic color segmentation of images with application to detection of variegated coloring in skin tumors. IEEE Eng. Med. Biol. 10: 57-62; 1991.

12. **Umbaugh SE, Moss RH, Stoecker WV:** Applying artificial intelligence to the identification of variegated coloring in skin tumors. *IEEE Eng. Med. Biol.* 8: 43-52, 1989.

13. **Thomas W, Hapgood W:** 1st-Class Fusion. 1st-Class Expert Systems, Inc. Wayland, Mass, 1987.

14. AIM User's Manual. AbTech Corp., Charlottesville, VA, 1993.

15. **Ballard DH, Brown CH:** *Computer Vision.* Prentice-Hall. New York, N.Y. 1982, 33.

Scott E. Umbaugh is currently an Assistant Professor in Electrical Engineering at Southern Illinois University at Edwardsville. He received the B.S.E. degree with honors from SIUE in 1982, the M.S.E.E. from the University of Missouri-Rolla in 1987, and the PhD in electrical engineering from UMR in 1990. While at UMR he was a chancellor's Fellow. He worked as a computer design engineer for ITT North Electric in Columbus, Ohio and McDonnell Douglas corporation and Affinitec corporation in St. Louis, Missouri, from 1981 through 1986.

He currently serves on the Editorial Board for the *IEEE Engineering in Medicine and Biology Magazine* and the *Pattern Recognition* journal. He is a member of the IEEE Education Society, the IEEE Engineering in Medicine and Biology Society, the Pattern Recognition Society, the American Society for Engineering Education, and Sigma Xi. His professional interests include artificial intelligence, computer vision, medical applications of image processing, and electrical engineering design education. Dr. Umbaugh can be reached at the Department of Electrical Engineering, Southern Illinois University at Edwardsville, Edwardsville, IL, 62026-1801, or seu@minuet.siue.edu.

Arve Kjoelen received his B.S.E.E. in 1993 and his M.S.E.E. in 1994 from Southern Illinois University at Edwardsville, centering on the areas of image processing, computer vision, signal processing and digital design. He received an Undergraduate Research Academy fellowship and a Competitive Graduate Award from the University, and served as a Research Assistant in 1993. His thesis focused on wavelet-based compression of medical images. His professional interests include medical, environmental and systems applications of image processing and computer vision. He is also interested in hardware implementations of image processing algorithms. Arve Kjoelen has been an IEEE member since 1991. He currently works for Ross & Baruzzini in St. Louis, Missouri.

Marc Thompson is currently an engineer at Union Electric Company in St. Louis, Missouri. He received his B.S. degrees in Electrical Engineering and Physics from Southern Illinois University at Edwardsville in March 1993. He is currently pursuing his M.S. degree in Electrical Engineering at Southern Illinois University at Edwardsville. His current research interests include image processing and computer vision.

Randy H. Moss is Professor of Electrical Engineering at the University of Missouri-Rolla (UMR). His research interests are in the areas of image processing, pattern recognition, and computer vision. He is especially interested in medical and industrial applications of machine vision techniques.

Dr. Moss received the B.S.E.E. and M.S.E.E. degrees in Electrical Engineering from the University of Arkansas, where he was a National Merit Scholar, and the Ph.D. from the University of Illinois, where he was a National Science Foundation Graduate Fellow. He serves as an Associate Editor of *Pattern Recognition* and of *Computerized Medical Imaging and Graphics*. His awards include the 1988 Society of Automotive Engineers Ralph R. Teetor Educational Award, the Society of Manufacturing Engineers 1987 Outstanding Young Manufacturing Engineer Award, honorable mention for the 1986 Eta Kappa Nu C. Holmes MacDonald Outstanding Teaching Award, the Tenth Annual Pattern Recognition Society Award (awarded in 1985 for the best paper published in *Pattern Recognition* in 1983), eight UMR Faculty Excellence Awards, and a UMR Outstanding Teacher Award.

Dr. Moss is a senior member of IEEE and a member of Sigma Xi, the Pattern Recognition Society, Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi.

William V. (Van) Stoecker received the B.S. degree in Mathematics in 1968 from the California Institute of Technology, the M.S. in Systems Science in 1971 from the University of California, Los Angeles, and the M.D. in 1977 from the University of Missouri, Columbia. He is Adjunct Assistant Professor of computer Science at the University of Missouri-Rolla and clinical Assistant Professor of Internal Medicine-Dermatology at the University of Missouri-Columbia. His interests include computer-aided diagnosis and application of computer vision in dermatology.