

100-107863  
91514  
P-13

## **Data Analysis using Scale-space filtering and Bayesian Probabilistic Reasoning**

**Deepak Kulkarni**  
Sterling Federal Systems  
AI Research Branch, Mail Stop 244-17  
NASA Ames Research Center  
Moffett Field, CA 94035

**Kiriakos Kutulakos**  
Computer Sciences Department  
University of Wisconsin  
Madison, WI 53706

**Peter Robinson**  
RECOM Technologies  
AI Research Branch, Mail Stop 244-17  
NASA Ames Research Center  
Moffett Field, CA 94035

(NASA-TM-107863) DATA ANALYSIS USING  
SCALE-SPACE FILTERING AND BAYESIAN  
PROBABILISTIC REASONING (NASA) 13 p

N92-25349

Unclas  
63/61 0091514

**NASA** Ames Research Center  
Artificial Intelligence Research Branch

Technical Report FIA-91-05

March, 1991

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE Dates attached		3. REPORT TYPE AND DATES COVERED	
4. TITLE AND SUBTITLE  Titles/Authors - Attached				5. FUNDING NUMBERS	
6. AUTHOR(S)					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Code FIA - Artificial Intelligence Research Branch Information Sciences Division				8. PERFORMING ORGANIZATION REPORT NUMBER  Attached	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)  Nasa/Ames Research Center  Moffett Field, CA. 94035-1000				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION / AVAILABILITY STATEMENT  Available for Public Distribution  <i>Pete Friedland</i> 5/14/92 BRANCH CHIEF				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  Abstracts ATTACHED					
14. SUBJECT TERMS				15. NUMBER OF PAGES	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT		



# **Data Analysis using Scale-space filtering and Bayesian Probabilistic Reasoning**

**DEEPAK KULKARNI**  
STERLING FEDERAL SYSTEMS  
AI Research Branch, Mail Stop 244-17  
NASA Ames Research Center  
Moffett Field, CA 94035

**KIRIAKOS KUTULAKOS**  
COMPUTER SCIENCES DEPARTMENT  
University of Wisconsin  
Madison, WI 53706

**PETER ROBINSON**  
RECOM TECHNOLOGIES  
AI Research Branch, Mail Stop 244-17  
NASA Ames Research Center  
Moffett Field, CA 94035

## **ABSTRACT**

This paper describes a program for analysis of output curves from Differential Thermal Analyzer. The program first extracts probabilistic qualitative features from a DTA curve of a soil sample, and then uses Bayesian probabilistic reasoning to infer the mineral in the soil. The qualifier module employs a simple and efficient extension of scale-space filtering suitable for handling DTA data. We have observed that points can vanish from contours in the scale-space image when filtering operations are not highly accurate. To handle the problem of vanishing points, perceptual organization heuristics are used to group the points into lines. Next, these lines are grouped into contours by using additional heuristics. Probabilities are associated with these contours using domain-specific correlations. A Bayes tree classifier processes probabilistic features to infer the presence of different minerals in the soil. Experiments show that the algorithm that uses domain-specific correlations to infer qualitative features outperforms a domain-independent algorithm that does not.

**AREA : Perception and Signal Understanding**

**CONTACT: Deepak Kulkarni**  
**ELECTRONIC MAIL: KULKARNI@PTOLEMY.ARC.NASA.GOV,**  
**PHONE NUMBER: (415) 604-4869**



# 1 Introduction

Many data interpretation programs extract qualitative features from data and carry out further analysis using them. Scale-space filtering [1,2] is a technique that can produce qualitative descriptions of continuous curves. This paper describes an extension of this technique that has been applied to the analysis of Differential Thermal Analyzer (DTA) curves. DTA is an instrument that geologists use to analyze the contents of a soil sample. An investigator uses DTA to heat the soil sample and an inert reference material at the same time, and measures the difference in the temperature of the two at discrete temperature intervals. DTA output is a graph of this temperature difference vs. the temperature of the reference material. Figure 1 shows an example of DTA output. As a soil is heated in a DTA, minerals in the soil sample undergo phase transitions and reactions at various stages. An exothermic reaction produces heat and an endothermic reaction consumes heat. An exothermic reaction results in a hill in the output curve. In contrast, an endothermic reaction results in a valley in it. The temperature at which a reaction occurs is diagnostic of minerals present in the soil. For example, an endothermic reaction at 570°C would indicate the presence of quartz. A number of reactions may occur simultaneously and may correspond to small and large features superimposed on each other. The resultant curve is complex, and a domain expert can not identify features in the curve with certainty. For example, in Figure 1, experts interpret that there is some chance of an endotherm at 907°C at A. We have developed an expert system that can first extract probabilistic qualitative features from the DTA output of a sample and use them in recognizing its contents. The system uses a simple and efficient extension of scale-space filtering suitable for handling DTA data and for producing a probabilistic scale-space description. The system also has a Bayes tree classifier that can use probabilistic qualitative features in the curve to recognize the contents of the sample.

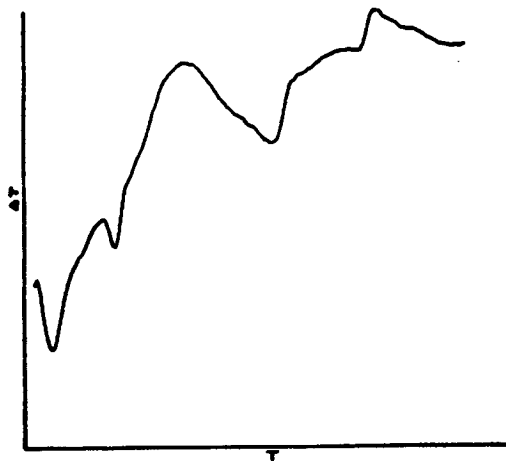


Figure 1: DTA output for a sample

## 2 Generating Qualitative Descriptions of DTA Curves

Scale-space filtering has proved useful in the analysis and representation of one-dimensional signals [1,2,3,4]. This is especially the case in the analysis of curves from DTA experiments, because of

the special cause-effect relationships between the thermal events (i.e. reactions) taking place in the experiments and the resulting curve variations: (a) If a thermal event is fully contained in a temperature interval and it is the only event taking place in that interval, then the event gives rise to a single peak or valley in the DTA curve. (b) The effects of thermal events fully contained in disjoint temperature intervals are independent, and (c) if two or more thermal events overlap, their effects on the DTA curve are additive.<sup>1</sup> These properties allow us to characterize the thermal events taking place in a DTA experiment by identifying and localizing peaks, valleys or overall trends in the DTA curve. In the case of curve valleys and peaks the associated extremum lies between a pair of inflection points. This pair of points therefore identifies a temperature interval over which a curve variation of the above types is, in general, known to take place (Clark [6] discusses how one can determine whether or not a particular point of inflection corresponds to some signal valley or peak). The main contribution of the scale-space filtering approach is in identifying these temperature intervals by pairing the detected points of inflection.

## 2.1 Scale-Space Filtering

The theory of scale-space filtering [1,2,5], addresses the problem of qualitatively describing an input signal  $f(x)$ . The approach is based on the assumption that the variations present in  $f(x)$  are caused by the interaction of several physical processes [1,2]; the signal's final description is a characterization of these processes. Signal variations are detected by finding the points of inflection in  $f(x)$ . Since the points of inflection are detected by applying derivative operators, the number and position of these points depends not only on  $f(x)$  but on the spatial extent (i.e. the scale) of the operators. No single scale is assumed correct; rather, each scale gives rise to a number of points of inflection. The goal of the approach is to group, across all scales, the points of inflection caused by the same physical process [2]. Therefore the curve can be described not by the individual points of inflection detected but by the groups they form.

As described in [3,2], we can analyze the input signal  $f(x)$  at various scales by first convolving it with Gaussian kernels of varying standard deviation  $\sigma$ . For fixed  $\sigma$ , the points of inflection in the smoothed signal  $f(x, \sigma)$  are the points where  $\frac{\partial^2 f(x, \sigma)}{\partial x^2}$  crosses zero. As the original signal gets increasingly smoothed, the only zero-crossing points remaining are those that correspond to the most prominent variations in  $f(x)$ . When plotted on the  $(x, \sigma)$ -plane the points of inflection of  $f(x, \sigma)$  form contours similar to the ones shown in Figure 2a, creating the so-called scale-space image for  $f(x)$ . Yuille and Poggio [5] proved that these contours are 'well-behaved' only in the case of the Gaussian kernel; no new contours are created as  $\sigma$  increases and the contours form continuous, smooth curves with a single extremum. In this case, all the zero-crossing points in the scale-space image can be traced by starting at the finest level of detail (i.e. the unsmoothed signal) and terminating at the first scale in which  $f(x, \sigma)$  has no points of inflection. Motivated by Witkin's [2] assumption that each contour in the scale-space image corresponds to a single physical event, two zero-crossing points in the (unsmoothed) signal are paired only if they belong to the same zero-crossing contour. The paired zero-crossing points define the interval within which the event takes place. This observation applied to the DTA curve analysis problem gives a way to determine the temperature intervals over which thermal events take place in DTA experiments.

---

<sup>1</sup>We are developing a separate module that handles exceptional situations that are not consistent with these properties.

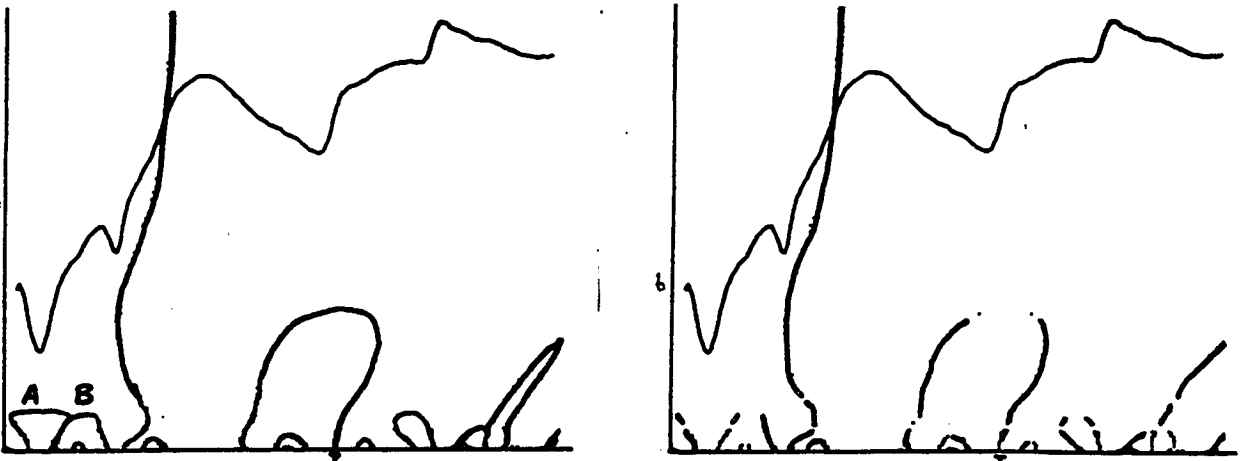


Figure 2: (a) Typical scale-space image of a DTA curve. Note that contours  $A$  and  $B$  are intersecting. (b) The computed scale-space image contains disconnected contours (e.g.  $X$ ), due to approximation errors in the derivative and zero-crossing calculations. The images are superimposed on the signal  $S$ .

## 2.2 Implementation

A major concern in the implementation of the above approach was the appearance of the zero-crossing contours in the computed scale-space image: In the scale-space images of DTA curves, contour intersections are not uncommon (Figure 2a). In addition, the use of finite-size approximations to the Gaussian filter for smoothing the curve implies that the ideal case of smooth and continuous zero-crossing contours cannot be assumed. Therefore, simply tracking the zero-crossing points from lower scales to higher ones is inadequate to group them into contours. Even further, as the standard deviation  $\sigma$  of the smoothing filters grows the possibility of numerical errors in the derivative calculations becomes increasingly important: If  $g(T, \sigma)$  is a Gaussian operator of standard deviation  $\sigma$  (defined as in [1]) and  $dta(T)$ ,  $dta(T, \sigma)$  are the original and smoothed curves respectively, then we have

$$\begin{aligned}
 g(T, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{T^2}{2\sigma^2}\right\} \\
 \frac{\partial^2}{\partial T^2} dta(T, \sigma) &= \frac{\partial^2}{\partial T^2} [dta(T) * g(T, \sigma)] \\
 &= dta(T) * \left[ \frac{\partial^2}{\partial T^2} g(T, \sigma) \right].
 \end{aligned}$$

The magnitude of  $\frac{\partial^2}{\partial T^2} g(T, \sigma)$  decreases rapidly as  $\sigma$  increases, to values very close to zero. Since numerical errors can affect the detection of zero-crossing points in the scale-space image, it is important to be able to process scale-space images that are not taken to be completely accurate. Figure 2b shows a typical scale-space image in which some contours appear to be disconnected.

The theory of scale-space filtering assumes that the variance of the Gaussian smoothing filters varies continuously in the interval  $(0, \infty)$ , creating a continuous scale-space image. In any implementation this image must be sampled to give a discrete approximation. The density of the samples



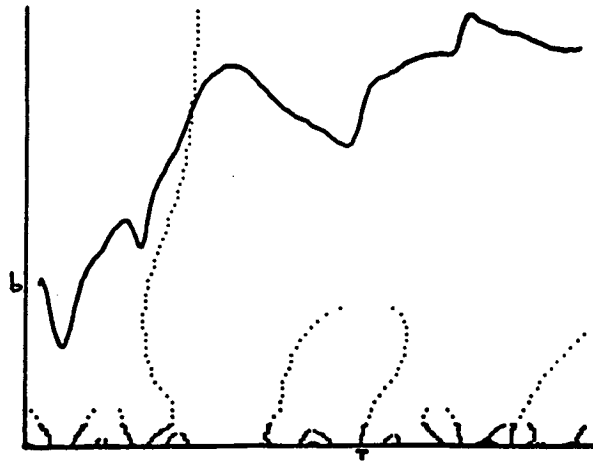


Figure 3: The sampled scale-space image of Figure 2. Note that the sampling rate is larger at lower scales.

greatly affects performance, since the CPU requirements of the convolution operations become large as the scale of the filter increases. It is therefore an important issue to be able to handle scale-space images that are only sparsely sampled. A typical sampled image is shown in Figure 3. Note that because of the sparse sampling of the image, the scale-space contours appear split into two lines starting at  $\sigma = 0$  and extending to higher scales in a near-vertical fashion. Also note that several sampled contours have points missing at some scales.

With the above issues in mind, qualitative descriptions of DTA curves are generated in four steps: (i) A sparsely-sampled scale-space image is computed for the curve (Figure 4a), (ii) zero-crossing points in the image are grouped into lines starting at the lowest scale and extending to higher scales (Figures 4b,c), (iii) the lines are paired to form scale-space contours (Figure 4d), and (iv) the scale-space contours detected are used as a description of the underlying thermal events generating the DTA curve.

The discrete scale-space image is computed by first convolving the signal with finite-size approximations to  $\frac{\partial^2}{\partial T^2}g(T, \sigma)$ . The scales of the filters are taken from a fixed set of values, sampling the image at larger  $\sigma$ -intervals as  $\sigma$  grows. Our experimental results show that the use of a fixed number of scales is adequate for our analysis task. Approximations to  $\frac{\partial^2}{\partial T^2}g(T, \sigma)$  are created by truncating the infinite-size filters at a point where the area under the truncated part of the filters becomes less than an error  $\epsilon$ . Therefore, when  $\epsilon$  is fixed the standard deviation determines the size of the truncated filter. Instead of defining zero-crossing thresholds based on this error [7,8] we chose to use the simple approach of slightly distorting the filter so as to force its total mass to be zero. The zero-crossing points can then be found by detecting all negative-to-positive or positive-to-negative transitions in the filtered curve (Figure 3).

In order to efficiently and reliably group the zero-crossing points into lines in situations similar to the one in Figure 5 we use the perceptual organization techniques developed by Lowe and Binford [9,10,11]. These techniques try to estimate the likelihood of particular relations between points or lines happening by accident. Since the points of a zero-crossing contour create two almost straight lines that eventually converge at some scale, we are interested in finding groups of collinear points.

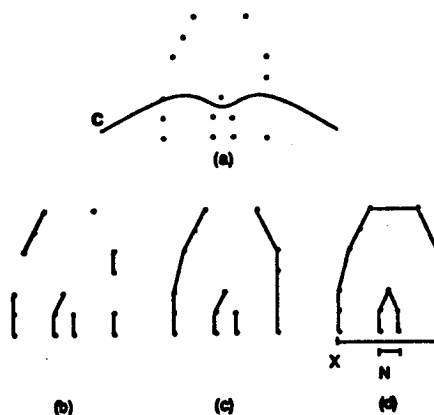


Figure 4: The sequence of steps taken in the analysis of DTA curves. (a) A discrete scale-space image is computed for a curve  $C$ . (b) The points in scale-space are grouped into segments. (c) The segments are grouped into lines. (d) The lines are paired to form the scale-space contours. The curve is described as the result of two reactions, one exotherm reaction taking place over the temperature interval  $X$  and one endotherm reaction taking place over the interval  $N$ .

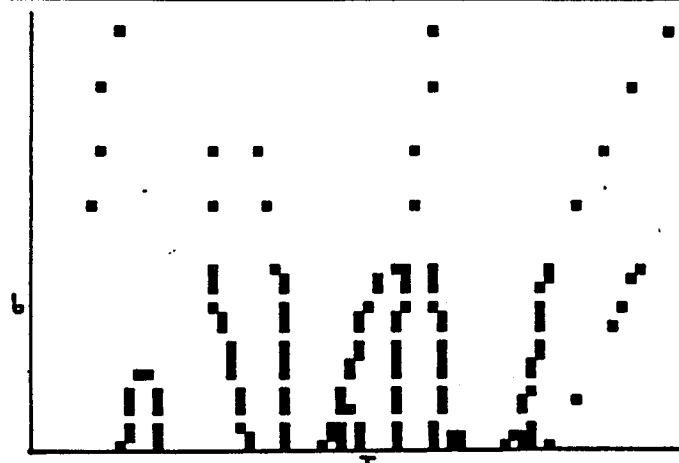


Figure 5: The density of zero-crossing contours at lower scales results in a large number of points that must be reliably grouped into lines. This figure is a magnified part of a typical scale-space for DTA curves.

In this case, a number of points would be accidentally grouped together if some of the points do not belong to the same contour (Figures 6a-c). Perceptual organization then allows us to group the zero-crossing points into lines while minimizing likelihood of these groupings being accidental.

The points are grouped by scanning the scale-space image row-by-row, starting from the lowest scale first. Each point at the lowest scale (i.e. the bottom row of the scale-space image) will belong to a different line. When a point is found at the next row it is placed in the group most likely to contain the newly-found point. This group is found by computing the likelihood that the newly-found point is put in the group by accident. When the top row of the image is reached all points

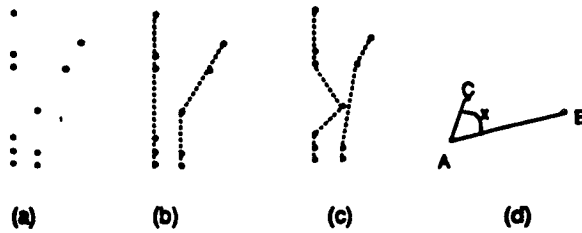


Figure 6: Using perceptual organization to group points. (a) Points belonging to two zero-crossing lines. (b) A likely grouping of the points. (c) A less plausible grouping of the points. (d) Estimating the likelihood of  $C$  being accidentally aligned with  $A$  and  $B$ : This likelihood can be expressed using the angle between  $AC$  and  $AB$ .

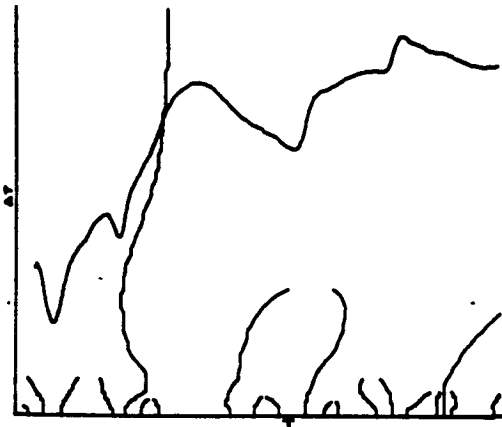


Figure 7: Detecting the lines in the scale-space image of Figure 2.

in the image will be in some group. These groups represent the lines forming contours in the scale-space image. In order to quantify the probability that a particular point  $p$  is placed in a group  $G$  of points by accident we assume that the points in the group lie on a background of uniformly distributed points. Also, we are given a measure of point collinearity and the density of the randomly distributed background points. The probability of  $p$  being placed in  $G$  by accident is equal to the expected number of background points that have stronger collinearity relations with the points in  $G$  than  $p$  (with respect to the defined collinearity measure). Figure 6d gives an example of such a measure in the case where  $G$  contains only two points [9]. We have made several improvements to the above basic algorithm in order to reduce the search conducted by the grouping process, including an initial stage where points in the image are first traced to form small segments (Figure 4b); the resulting segments are then grouped into lines using perceptual organization techniques similar to the ones described above [10] (Figures 4c and 7).

Lines in the scale-space image are of one of two (opposite) types: *Maxtype*, which contain zero-crossings corresponding to maxima of  $\frac{\partial}{\partial T} dta(T, \sigma)$  and *mintype* which is contain zero-crossings corresponding to minima of  $\frac{\partial}{\partial T} dta(T, \sigma)$ . If two lines terminate at scales  $\sigma_1$  and  $\sigma_2$ , we define their scale-ratio to be the ratio of the larger scale divided by the smaller scale. Thus, if lines  $L_1$  and  $L_2$

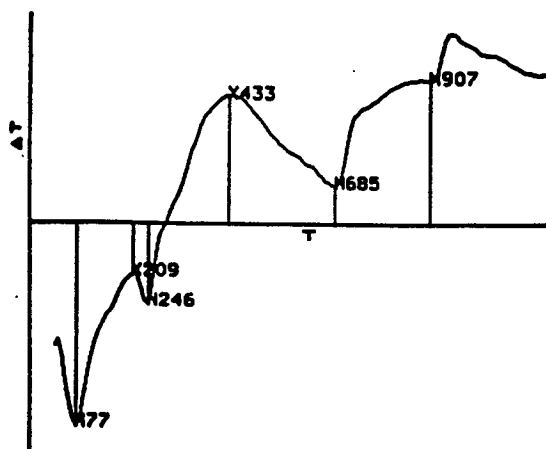


Figure 8: The output of the qualifier. Associated probabilities are not shown in the figure.

have maximum scales of 5 and 10 respectively, then their scale-ratio is 2. When a line of *maxtype* is paired with another line of *mintype* at a higher temperature (i.e. the *mintype* line is on the right of the *maxtype* line), the two correspond to a hill. When a line of *maxtype* is paired with another line of *mintype* at a lower temperature, the two correspond to a valley. The system employs the following procedure to identify partner lines for a line  $L$  having maximum scale  $\sigma_L$  and whose lower endpoint hits the  $\sigma = 0$  axis at temperature  $T$ . It considers as potential candidates all lines with type opposite to that of  $L$  and with distance from  $L$  less than a threshold defined as a function of  $\sigma_L$ . It identifies a line that has the minimum scale-ratio with  $L$  among the candidates with lower endpoints at temperatures greater than  $T$ , and defines it as the hill partner. Similarly, it identifies a line that has the minimum scale-ratio with  $L$  among the candidates with lower endpoints at temperatures lower than  $T$ , and defines it as the valley partner. Line  $L$  is paired with its hill partner to form a hill. Line  $L$  is paired with its valley partner to form a valley. The system then associates a probability with every hill and valley it finds. To do this, it uses domain specific statistics that are correlations between the behavior of the function at the extrema in the derivatives, the second derivative at the peak and the probability of its being a real contour. For example: if the value of function is decreasing at the minima in the derivative, and it is increasing at the maxima in the derivative, and the second derivative at the peak of an endotherm is high, then the probability of the peak being an endotherm is .9. The output of the qualifier is shown in figure 8.

### 3 Classifier

In this section, we describe how the classifier module uses qualified description to recognize the contents of a sample. The module has a Bayes tree rooted in each mineral node. Figure 9 shows an example of a tree for the mineral 'Smectite'. The terminal nodes in the tree are predicates of the form: data has an endotherm / exotherm with peak temperature between  $a$  and  $b$ . They are tested against the features defined by the qualifier. The intermediate nodes represent phase transitions and reactions. Nodes are connected to each other by arrows. An arrow from node  $A$  to node  $B$  represents the relation that  $A$  causes  $B$ . Thus in figure 9, smectite causes water loss,

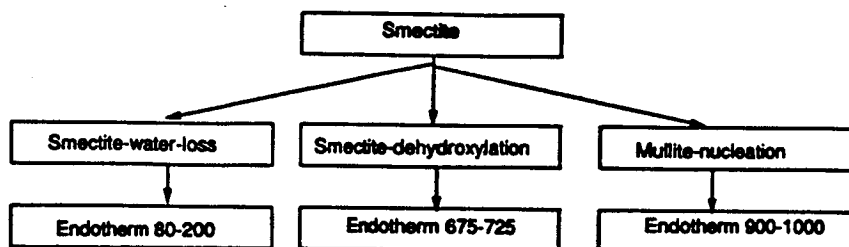


Figure 9: A Bayes tree for the mineral Smectite.

---

dehydroxylation and mullite nucleation. Furthermore water loss causes an endotherm between 80 and 200, dehydroxylation causes an endotherm between 675 and 725 and mullite nucleation causes an endotherm between 900 and 1000. The root, the intermediate nodes and the terminal nodes can take the values true and false. A link from node A to node B in the tree stores four probability values:  $P(a/b)$   $P(\neg a / b)$   $P(a / \neg b)$ , and  $P(\neg a / \neg b)$ . As we described earlier, qualifier associates probabilities with each feature in the curve. The classifier uses this information to assign probabilities to the terminal nodes in the tree. Given probabilities in the terminal nodes, a standard Bayes tree propagation algorithm [12] is used calculate the probabilities of all non-terminal nodes including the root. For the curve in figure, the system finds only two minerals - glycine and smectite - to have probability greater than .5.

## 4 Experimental results

Our line-pairing scheme is based on the conjecture that an algorithm that uses domain specific correlations should perform better than one that does not. We will now describe an experiment that we carried out to test this. For the sake of clarity, we will refer to the line-pairing algorithm described at the end of Section 2.2 as A, and another algorithm that pairs the lines with minimum Manhattan distance between their upper endpoints in the scale-space image as M. These two algorithms were tested on a random sample of 20 curves from the data bank in the DTA laboratory in NASA Ames Research Center. Algorithm M detected 49% of the peaks reported by an expert. In contrast, algorithm A detected 89% of the peaks identified by an expert as having probability greater than 0.5. In this evaluation we compared the performance of the algorithms with that of an expert. In a number of domains, the features identified by an expert differ from those identified by a non-expert. In the DTA domain, our algorithm can identify 70% of the peaks identified by a non-expert, but 89% of th peaks identified by an expert. While the curve will be given an unique interpretation by

a non-expert, it can have different interpretations in different domains. Qualifier described in this paper can produce different qualitative descriptions of a curve from correlations specific for different domains. It also meets the real-time requirements of the DTA task, as it takes 360.2 seconds of CPU time on SPARC workstation for extracting the features.

The integrated system was tested on 16 known soil samples, and the system is able to identify the contents of the samples. 75% of the terminal nodes of the minerals present in these samples were assigned a probability greater than 0.5 by the classifier. Furthermore out of 121 terminal nodes, 64 of the terminal nodes belonged to minerals not present in the soil. The reason for this is that a number of different minerals share common symptoms.

## 5 Conclusion

This paper described an extension of scale-space filtering method that is useful for extracting probabilistic qualitative features from data sampled at discrete time intervals.

The system groups lines in the scale-space image into contours by examining the maximum scale of lines in the scale space graph. It uses domain specific correlations to associate probabilities with these contours. Experiments show that this scheme outperforms a domain-independent scheme. We also showed that a Bayes tree algorithm can make use of the probabilistic description produced by the qualifier in interpretation of the curve. Our future plan to carry out a systematic study of how the CPU time and the classification performance changes when the sampling of the scale-space image gets more and more sparse. This would determine at which point the algorithms break down and how much performance can be gained by filtering the curve the smallest number of times.

## 6 Acknowledgements

We would like to thank Rocco Mancinelli and Lisa White, who provided the DTA test data. Our research has benefitted from numerous discussions with Dave Thompson, Rich Levinson, and Pat Langley.

## References

- [1] Witkin, A., Scale-Space Filtering, *Proc. 8th Int. Joint Conf. Artificial Intelligence*, 1983, 1019-1021.
- [2] Witkin, A., Scale-Space Methods, in *Encyclopedia of Artificial Intelligence*, S. Shapiro, ed., Wiley, New York, 1987, 973-980.
- [3] Mokhtarian, F., and Macworth, A., Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes, *IEEE Trans. Pattern Analysis Machine Intelligence* 8, 1986, 34-43.
- [4] Hildreth, E., Edge Detection, in *Encyclopedia of Artificial Intelligence*, S. Shapiro, ed., Wiley, New York, 1987, 257-267.
- [5] Yuille, A., and Poggio, A., Scaling Theorems For Zero Crossings, *IEEE Trans. Pattern Analysis Machine Intelligence* 8, 1986, 18-25.

- [6] Clark, J., Authenticating Edges Produced by Zero-Crossing Algorithms, *IEEE Trans. Pattern Analysis Machine Intelligence* 11, 1989, 43-57.
- [7] Lu, Y., and Jain, R., Behavior of Edges in Scale Space, *IEEE Trans. Pattern Analysis Machine Intelligence* 11, 1989, 337-356.
- [8] Zuerndorfer, B., and Wakefield, G., Extensions of Scale-Space Filtering to Machine-Sensing Systems, *IEEE Trans. Pattern Analysis Machine Intelligence* 12, 1990, 868-882.
- [9] Lowe, D., and Binford, T., Perceptual Organization As A Basis For Visual Recognition, *Proc. AAAI-83*, 1983, 255-260.
- [10] Lowe, D., Three-Dimensional Object Recognition from Single One-Dimensional Images, *Artificial Intelligence* 31, 1987, 355-395.
- [11] Lowe, D., The Viewpoint Consistency Constraint, *International Journal of Computer Vision* 1, 1987, 57-72.
- [12] Pearl, J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, 1988, 144-200.