Analysis of the shape of noise corrupted signals or spectra for comparison and feature extraction

Hartwig Hetzheim
Institute of Space Sensor Technology and Planetary Exploration, DLR, Berlin
hartwig.hetzheim@dlr.de Fax +493067055 512

Abstract

In the shape of a curve hidden information is superposed by disturbances. This information in the measured curve can be isolated by decomposition based on mapping of properties on selected measures. For the estimation of the similarity of curves, parts and composed properties of the shapes are selected. Because all kinds of properties cannot be determined, estimation theory based on probability cannot be applied. The importance of a property has to be considered and an extension of the Lebesgue measure is used. Such measures, extended Lebesgue functions, and a cut value for unimportant properties describe the properties. With a Lebesgue integral properties are fused and new properties composed. By an iterative process properties are separated successively. The method is applied for retrieval of spectra and for comparison boundaries of clouds in multi-spectral stereo images.

Keywords : Fusion of properties, curve analysis, retrieval of spectra, stereo image processing, time series

1. Introduction

Normally, the signals or spectra are corrupted by noise and the shape is not precise defined. In some signal processing tasks the shape of the curve contains important information. The interest here is to find parts of curves, where the accordance with parts of another curve, measured or synthesised, is maximal. The interest is not to describe or to approximate a curve by functions as in [1]. The retrieval of profiles from molecular spectra, based on the collision broadening, is an example where the shape analysis successfully solves the inverse problem. The estimation of the boundary of clouds by estimation of parts of fuzzy contours with the highest similarity is an example of shape analysis for stereo image processing. The shape analysis can also be used to find parts of similarity between related curves in time series problems and the detection of hidden signals. In all this tasks exists the problem that the information about the spectra or the signals is not complete. Normally, as for the probability theory, such knowledge of all existing parts is a condition for the application of estimation methods. Because all kinds of possible parts of the shape are not known, the normalisation is not possible and different properties have different importance. The estimation of the shape f of an observable value η observed by the value ξ in the form

$$\inf_{f} \mathbf{M}(h - f(x))^2$$

is often not optimal if the importance of the properties are not included in the calculation. For the estimation, methods have to be developed, which use only the monotonicity and loos the additivity. For these estimation methods, also the stochastic properties are used. The basis for this is an extension of the Lebesgue measure theory.

2. Extension of Lebesgue measure and function

Stochastic properties with finite values can be measured by the Lebesgue measure [2], especially if they are not steady. The properties are measured on the open interval $b_v - a_n$. This is related to the elementary geometrical content $|i|$ of

$$|i| = \prod_{n=1}^{n} (b_v - a_n) = \sum_{V=1}^{n} (-1)^n c_1 \cdots c_n$$

Here, c_n are the interval points and the Lebesgue measure is the summation of the contents of these intervals. The characteristic function of the calibration measure $g(x)$ describes the membership for a property x within i and is 1 if i is a member of the set and 0 if not. The characteristic function for the set $\bigcap_{k=1}^{n} i_k$ is given by $\prod_{k=1}^{n} i_k(x) = \inf_{1 \le k \le n} \{g(x_k)\}$.

This measure gives the possibility to define a function of arbitrary parts of intervals as desired. However, the measure is additive. This contradicts the importance for the decision and the non-complete knowledge. Otherwise, the intervals for the Lebesgue measure are all different by pairs. Therefore, the Lebesgue measure has to be extended. A coupling between the components represents the importance of a property. Such a coupling by a factor λ is given by the fuzzy elementary measure $h(x)$ introduced by Sugeno[3] as

$$h_l(x_1 \cup x_2) = h_l(x_1) + h_l(x_2) + l\, h_l(x_1) h_l(x_2)$$

Here, λ is given by

$$l = (1 + l\, h(x_1))(1 + l\, h(x_2)) - 1$$

Such a relationship is used for the extension of the Lebesgue measure for combining intervals, which represent different types of properties. On the Lebesgue point measure are defined other point measures, the Lebesgue functions. These measures imply different properties, which overlap in the intervals, and so intervals are coupled by these properties given on these elementary intervals. For such extended intervals $j(x)$ can be written

$$j_l(x_1 \cup x_2) = j(x_1) \cup (j(x_1) \cap j(x_2)) \cup\ l(j(x_1) \cup j(x_2))$$

The sub-measures s for the closed set $A \subseteq M$ are determined by $s(M) = \sup_{A \subseteq M} s(A)$. On these interval measures, unsteady functions are measured. The unsteadiness of the functions is generated by superposition of process with different values within closed intervals, e.g. the changing of grey values in images by superposition of points of surfaces of different heights. The functions are defined on the parts of intervals of large but limited values on the intervals. This values can be stochastic if they are finite, what is fulfilled for usual physical values. The Lebesgue functions can also extended. Such functions are coupled together by their importance. For a compact definition, the stochastic functions are normalised and mapped on the closed interval [0,1]. By such a function, defined on the extended measure, properties of different kinds are represented in the same manner. So, non-additive properties are described where also stochastic properties are included. For each a priori expected stochastic property the martingale method for estimation is used. The stochastic properties are coupled together. This is described by a coupled system of stochastic differential equations for the stochastic components including the observation equation. The equations are described by spatial relationships between parts of curves and neighbouring parts of the curve and also for a selected region of different curves. With the help of different martingales, the estimation values of assumed components within different regions are calculated. The subtraction of so estimated values from the original values of the curve give the stochastic contribution. The stochastic properties are coupled with the large-scale (non-stochastic) properties and then fused for a separation of a curve in their elementary parts. This is the fundament for a fusion of measures and related functions for a generation of composed generalised properties.

3. Fusion of stochastic properties by an extended Lebesgue integral

The fusion of properties described by Lebesgue measures and Lebesgue functions e is possible with the Lebesgue integral. The set $\bigcup_{k=1}^{n} e_k$ combine all Lebesgue functions defined on the Lebesgue measure i and the characteristic function for $x \subseteq A$ is obtained as $1 - \prod_{k=1}^{n} (1 - e_k(x)) = \sup_{1 \le k \le n} \{e_k(x)\}$. The fluctuations of the functions over the Lebesgue measures are decomposed by intervals of variations, so the unsteadiness and jumping functions are well described. The Lebesgue functions are also defined in [0,1]. The combination of selected measures, i.e. infimum with all possible functions, i. e. supremum, gives the combination supremum infimum for the combination. This is based on the Lebesgue measures, but behind the monotonicity the additivity for the components have to be fulfilled. This contradicts the importance of the properties and is solved by an extension.

The extension of the Lebesgue integral has a cut, limiting the influence of properties. If the importance of a property is less than an assumed level α, then the influence of this property is neglected. For the extended Lebesgue integral over the Borel set B is obtained

$$^{B}\!\!\int_A = \sup\{\inf\{e(x,a), g(x),\ l\}\}$$

The importance is included by the condition for the function $e(x,a) = \begin{cases} 0 & \forall\ x \ \ where \ \ e(x) \le a \\ e(x) & \forall\ x \ \ where \ \ e(x) > a \end{cases}$

If the Lebesgue measure and the related Lebesgue function in the Lebesgue integral is replaced by a combination of the fuzzy measure and the fuzzy function then the fuzzy integral [3] with the notation $\int_{fuz}$ is obtained:

$$\int_A {}_{fuz}\ e_a(x) \oplus dg = \sup_{a \in [0,1]} \{\min[a, g(A \cap e_a(x))]\}$$

The simpler combination by extension of the Lebesgue integral is more understandable for the description of parts of curves than the fuzzy integral. Because

Lebesgue functions are also measured by a measure, the result of the integral is again a Lebesgue function.

$$g_{2_{i,j}}(A, x,\ l) = \int_A^B e_a(x) \oplus dg_1(x,\ l)$$

Roughly spoken, the fusion of properties describes the maximal grade of agreement between the objective evidence and the expectation. Based on this, complex composed stochastic properties can be constructed.

4. Representation of properties of curves by extended Lebesgue measures and functions

The description of stochastic and large-scale properties in form of a function on a Lebesgue measure is important for a successful separation of properties from measurements. The aim is to decompose all superposed functions within the measured curve. For this, the stochastic properties in the curves are extracted by non-linear filtering, rank algorithm, wavelet analysis, or analysis of the stochastic moments. The decomposition is done hierarchically in every step by subtraction of the detected component from the before subtracted measured curve. The algorithm is finished until the curve is flat enough. For the Lebesgue measure are used properties such as: variance is within a boundary, number of ranks related to different distances, gradient of neighbouring values is within a selected boundary, differences of the neighbouring pixel-values are within a selected boundary. For the Lebesgue function properties are used such as: differences of values in different distances, wavelet transformation function with a selected parameter, weak changes of the values in related areas. The large-scale properties such as local extrema, gradients, turning points, jumps, higher moments, skewness, kurtosis, mean absolute deviation are also included.

For the investigation of the curves, the Lebesgue functions are defined on selected generalised parts of the curve, described by measures. This means that the parts of the curve are not only built by geometrical parts of the curve, but also by complex measures. With such a composed description, hidden stochastic and large-scale properties are detected and decomposed in different components. For the description of complex interactions, mixed properties defined on complex measures are composed. This cannot be obtained by usual methods.

The algorithm for the decomposition works iterative. If a component is detected, then this contribution is subtracted from the curve. The obtained curve is analysed for this property once more or for another property. The effect of such subtraction for stochastic properties is shown in Fig. 1-3.

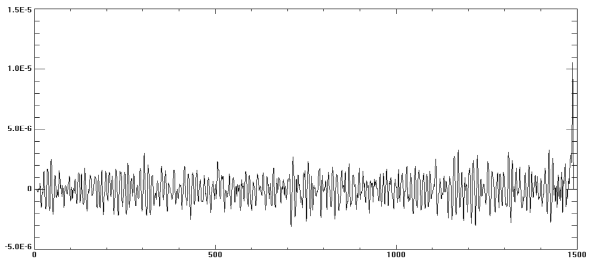

Fig. 1 Curve with short interaction in the stochastic

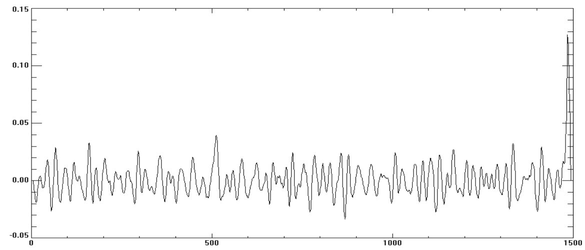

Fig. 2 Curve above very short interaction subtracted

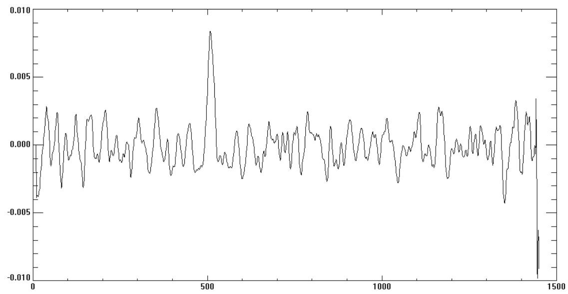

Fig. 3 Curve above medium interaction subtracted

For the fusion exist rules such as for the constant c

$$\int_A (e(x)+c)\ dg(x,\ l) \leq \int_A e\,dg(x,\ l) + \int_A c * dg$$

or

$$\int_A c * dg(x,\ l) = \inf(c, g(x,\ l))$$

which are not fulfilled for normal Lebesgue integrals.

5. Decomposition of composed molecular spectra in elementary spectra by best fitting of properties

The retrieval of height profiles from molecular spectra of minor constituents corrupted by noise and other disturbing effects is a typical example for analysing the shape. The pressure broadening of the spectra is the physical fundament to retrieve profiles and this is based on the shape of the spectra. The BrO spectrum, measured near 624 .7 GHz, has 15 lines in a small range. The superposition of these coupled lines with different shapes of contributions in different heights gives enough information for retrieval, but disturbances and neighbouring strong lines superpose this information. The spectrum is decomposed by subtraction of a forward-calculated spectrum with selected parameters as the number of molecules, and the altitude where the molecules are. These parameters are hidden in the shape. The kind of broadening from the maximum over the middle to the wings of the line contains the information of the superposition of contributions of different heights.

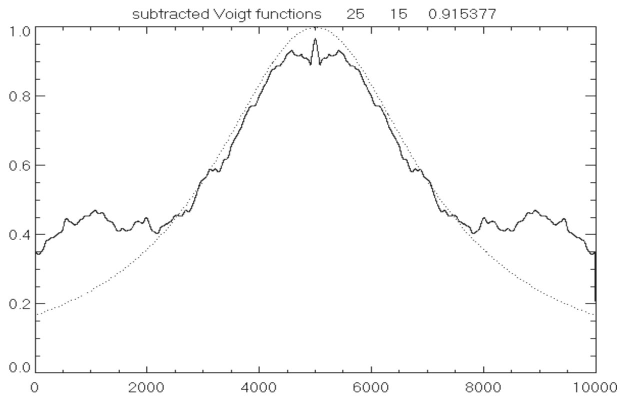

Fig. 4 Adaptation of a forward spectrum on a real one

Based on such an iterative algorithm the forward-calculated spectra with the best coincidence, related to Lebesgue measures, are subtracted form the measured spectra. These coincidences are searched for selected parts of the shape; for small curves only the top part, for broad spectra the width wings. The shape is adapted by different parts of it and by a combination of different properties. All are calculated and the maximal value of the Lebesgue integral gives the decision for the best adaptation. This is more precise than a fitting related to mean distances over the entire shape of the curve. The interaction of diverse properties and the different importance of selected properties substantiate this. So is retrieved the BrO profile, done in EU project HIMSPEC for understanding the processes for depletion of ozone in the arctic area [4].

5. Isolation of information within clouds based on contour analysis

For the comparison of clouds the characteristic information of the clouds have to be collected in points of clouds. Because the contours of the clouds are diffuse and edges do not exist, the clouds are difficult to describe. For multi-spectral images the properties between the different images are combined together. The simple combination of properties obtained by single images describe not precise the interaction between different wavelength channels. Such interconnecting combinations are the basis for the Lebesgue measures and functions. The Lebesgue integrals compose mixed contributions of two or more images. For example are combined different stochastic effects in images of different wavelength channels for the characterisation of the texture of the clouds combined with temperature influences. The Lebesgue integral is so built by a function on a region, which is defined by a measure for the selected region in another image. By such iterative procedure the information is more and more extracted from the images. The stepwise separated information is then represented by bitmap images. By algebraic and logic combinations of these bitmaps together with thresholds, the clouds are defined by texture measures [5].

For the detection of the heights of clouds, the original diffuse (stochastic) boundary is reduced to a well-defined contour. By comparison of contours in images of different spectral channels information about the height are obtained. The basis for this is the temperature sensitivity related to images on different wavelength. If stereo images exist, the contours between both images are compared by curve analysis on basis of the Lebesgue integral. Lebesgue measures and functions represent the properties of the shape of the curves. The minimum of the integral value between both contour parts of the stereo image is used for matching. Because a cloud is connected and of same height (tropical cloud towers may neglect), the cloud heights can be estimated by some corresponding points within the clouds. Such points can be found on the synthetic contour by a procedure similar as for the shape of spectra. These investigations are sponsored by the EU project CLOUDMAP (ENV4 CT97-0399) for detection of cirrus clouds and contrails.

References

[1] Tartner, M.E. and M.D.Lock: Model-Free Curve Estimation, Chapman & Hall, N. York, London 1993.
[2] v. Mangoldt, M., Knopp,K.: Einf. in die Höhere Mathematik, 4. Band, Hirzel Verlag., Leipzig, 1979.
[3] Sugeno, M.: Theory of fuzzy integrals and its application,Doctoral Thesis, Tokyo Inst. of Techn., 1974.
[4] Hetzheim, H. et all: Adaptive controlled retrieving of BrO profiles of ASUR measurements in Winter'99. Proc. of the Fifth European Workshop on Stratospheric Ozone, St. Jean de Luz, Sept. 99.
[5] Hetzheim, H. : Separation of different textures in images using fuzzy measures and fuzzy functions and their fusion by fuzzy integrals, Proc., ESIT'99, Crete, pp. 34-36 and CD, June, 1999.