

Computerised image analysis of biocrystallograms originating from agricultural products

J.-O. Andersen ^{a,*}, C.B. Henriksen ^a, J. Laursen ^b, A.A. Nielsen ^c

^a *Department of Agricultural Sciences, The Royal Veterinary and Agricultural University, Agrovej 10, DK-2630 Taastrup, Denmark*

^b *Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Thorvaldsgensvej 40, DK-1871 Frederiksberg C, Denmark*

^c *Department of Mathematical Modelling, The Technical University of Denmark, DK-2800 Lyngby, Denmark*

Accepted 30 October 1998

Abstract

Procedures are presented for computerised image analysis of biocrystallogram images, originating from biocrystallization investigations of agricultural products. The biocrystallization method is based on the crystallographic phenomenon that when adding biological substances, such as plant extracts, to aqueous solutions of dihydrate CuCl_2 , biocrystallograms with reproducible dendritic crystal structures are formed during crystallisation. The morphological features found in the structures are traditionally applied for visual ranking or classification, e.g. in comparative studies of the effects of farming systems on crop quality. The circular structures contain predominantly a single centre from where ramifications expand in a zonal structure. In previous studies primarily texture analysis was applied, and the images analysed and classified by means of a circular region-of-interest (ROI), i.e. the region specified for analysis. In the present study the objective was to examine how the discriminative information relevant for classification purposes is distributed over the zonal structure, and how the information is affected by the varying location of the crystallisation centre. The texture analysis procedures were applied to a so-called degradation series of 33 images, including seven groups representing discrete 'treatment levels'. The biocrystallograms were produced over seven consecutive days, on the basis of a single carrot extract degrading while stored at 6°C. This degradation is known to induce systematic changes in morpholog-

* Corresponding author. Tel.: +45-35-283520; fax: +45-35-282175; e-mail: joa@kvl.dk.

ical features over a number of successive days. The biocrystallograms were scanned at 600 dpi, with 256 grey levels. Eight first-order statistical parameters were calculated for four resolution scales, and 15 second-order parameters for five scales, giving a total of 107 observations for each image. Classification of an individual image was performed by means of stepwise discriminant analysis. Four main types, and several subtypes and sizes of ROI were examined. The 33 images as well as a subset of 21 images were examined. When imposing a restriction on the centre location in the subset, thereby reducing the within-group variance, the scores were markedly improved. Classifications of the total set and the subset showed scores up to 84.8 and 100%, respectively. A number of parameters showed a monotonic relationship with degradation day number. Multiple linear regressions based on up to eight parameters indicated strong relationships, with R^2 up to 0.98. It is concluded that the procedures were able to discriminate the seven groups of images, and are applicable for biocrystallization investigations of agricultural products. Perspectives for the application of image analysis are briefly mentioned. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Image analysis; Texture analysis; Biocrystallization; Copper chloride crystallisation; Quality assessment

1. Introduction

The biocrystallization method, also termed ‘sensitive crystallisation’ and ‘copper chloride crystallisation’, was originally introduced by Pfeiffer (1931). The method is based on the crystallographic phenomenon that when adding specific ionic substances, and generally all organic substances, to an aqueous solution of dihydrate CuCl_2 , biocrystallograms with reproducible dendritic structures are formed during crystallisation (Kleber and Steinike-Hartung, 1959). It is applied primarily in agricultural research concerning crop quality, complementary to chemical analyses of vitamins, proteins etc. A favoured field of application is comparative studies of the effects of different farming systems and fertilisation practices on the morphological features found in the crystal structures (Engqvist, 1989; Balzer and Balzer, 1993).

Crystallograms produced on the basis of pure CuCl_2 exhibit a merely peripheral distribution of crystals on the circular glass underlay, with a diameter of 90 mm, see Fig. 1. As opposed to this biocrystallograms produced on the basis of biological substances, such as plant extracts, exhibit crystal structures covering the whole of the glass underlay. The structures exhibit a variety of macro- and microscopical morphological features reflecting the specific admixed substances. The structures are zonal, co-ordinated with substance-specific ramifications around typically 1–2 centres, see Fig. 2. The ramification structure can be divided into three major zones: the central zone around the crystallisation centre, the median zone containing the major ramification structure, and the peripheral zone. The location of the crystallisation centre generally does not coincide with that of the geometrical centre. See Fig. 3 for a diagrammatic depiction of the zonal structure of a carrot extract

biocrystallogram. The colour phenomena observed in the crystal structures originate from the blue–green CuCl_2 , and in the case of a carrot extract biocrystallogram, from the yellow–orange extract applied. Among numerous single organic compounds examined proteins and N-containing compounds exhibit unique abilities to co-ordinate the crystal structure (Neuhaus, 1957).

In connection with quality assessment of agricultural products the biocrystallograms are traditionally ranked or classified visually by means of discrete scales, based on various morphological features. The upper median zone is reported by some authors to contain the most consistent information relevant for classification purposes (Engqvist, 1970). Generally it is argued that specific morphological features, and the types and degrees of crystal co-ordination reflect qualitative and nutritional properties of the examined product. However major differences are found among the authors concerning emphasis on specific morphological features,

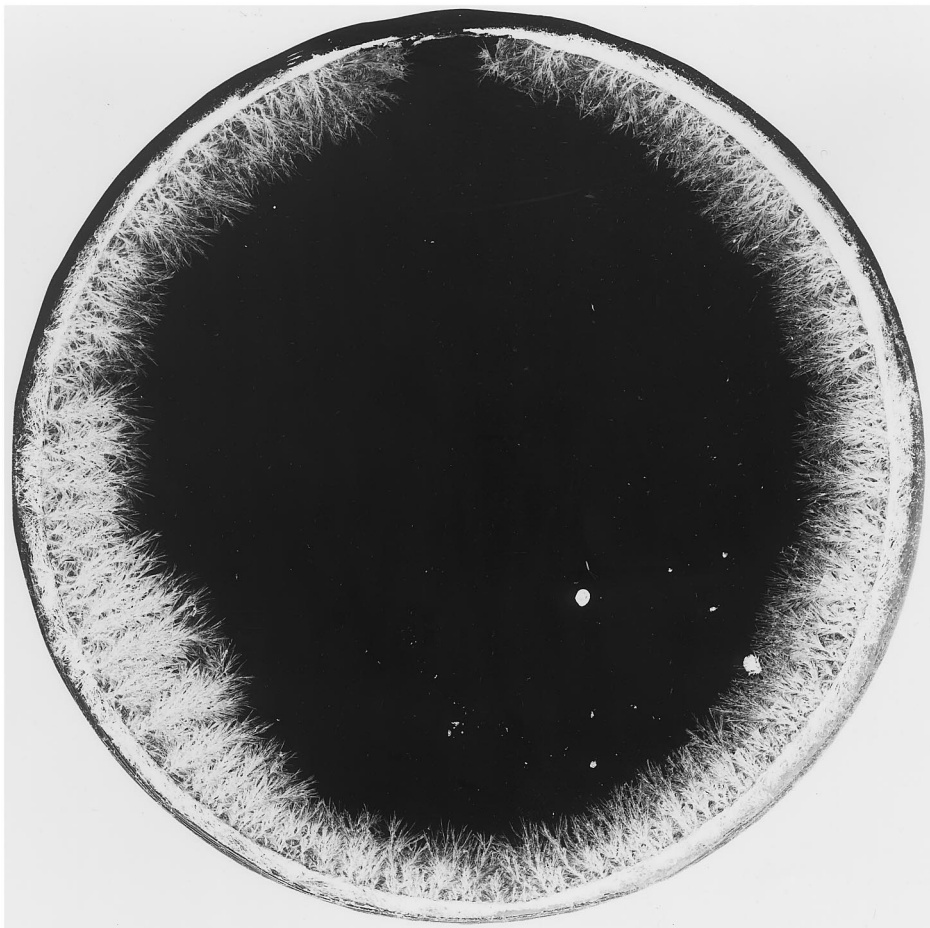


Fig. 1. Crystallogram produced on the basis of aqueous dihydrate CuCl_2 .

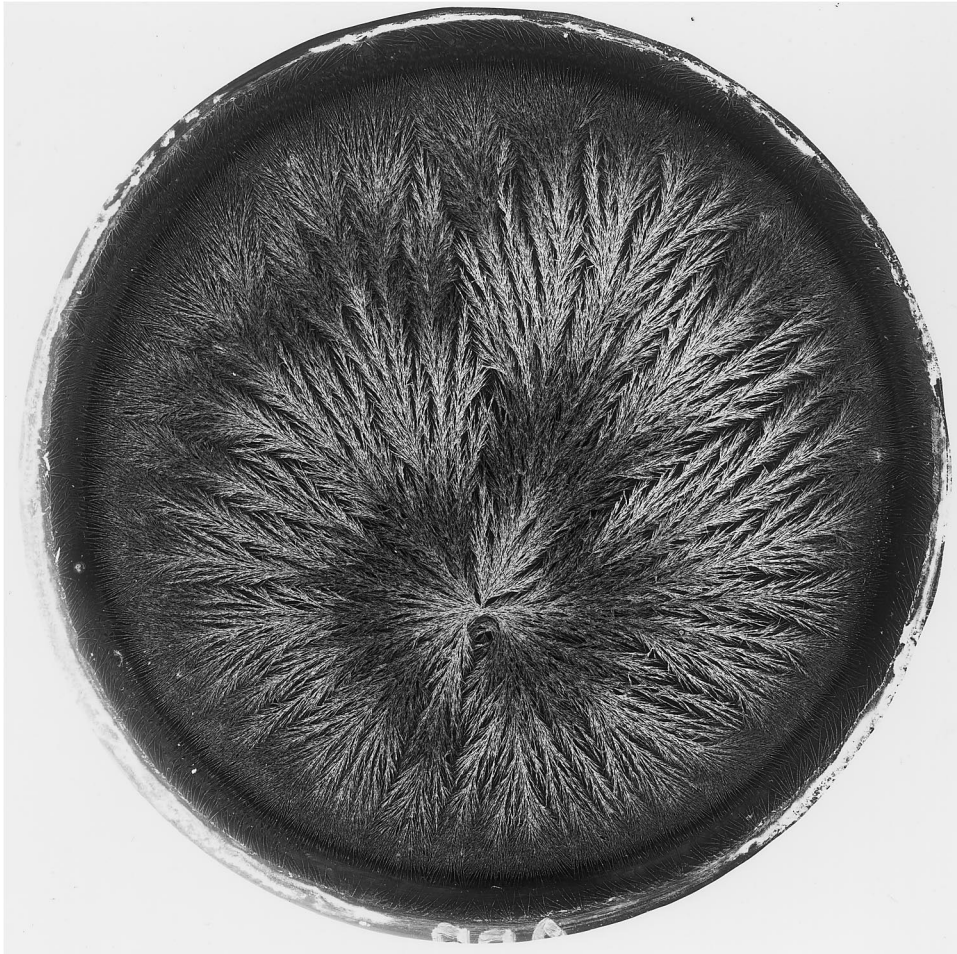


Fig. 2. Biocrystallogram produced on the basis of an aqueous carrot extract.

see e.g. v. Hahn (1962) and Engqvist (1970). It can be argued that a major limitation for a wider application of the biocrystallization method is the lack of standardised and objective methods of quantifying and classifying the morphological features.

Computerized image analysis techniques may meet the demand for such methods. Ideally an image analysis procedure should reflect all the characteristics of a biocrystallogram as a three-dimensional, coloured ramification structure, co-ordinated with zones relative to a centre. However due to the present limitations set by computational capacity and speed more simple approaches are preferable. In a limited number of previous studies encouraging results are reported (Lê Gia, 1996; Lê Gia et al., 1996). The studies are based on primarily texture analysis procedures, 32 grey levels, and a single type of circular region-of-interest (ROI), i.e. the region

of the biocrystallogram image specified for analysis. This ROI contains roughly the central and median zone.

Texture statistics are generally divided into first-order, second-order and higher-order statistics. First-order statistics refer to the marginal grey level distribution, second-order statistics to the joint grey level distribution of pairs of pixels, and higher-order statistics to the joint grey level distribution of three or more pixels. The basic computation of 14 features was introduced by Haralick et al. (1973) and later complemented by others (Connors et al., 1984; Parkkinen et al., 1990).

When loosely defining a texture as being spatially homogeneous, as opposite to a ramification structure expanding from a centre, it is seen that a biocrystallogram

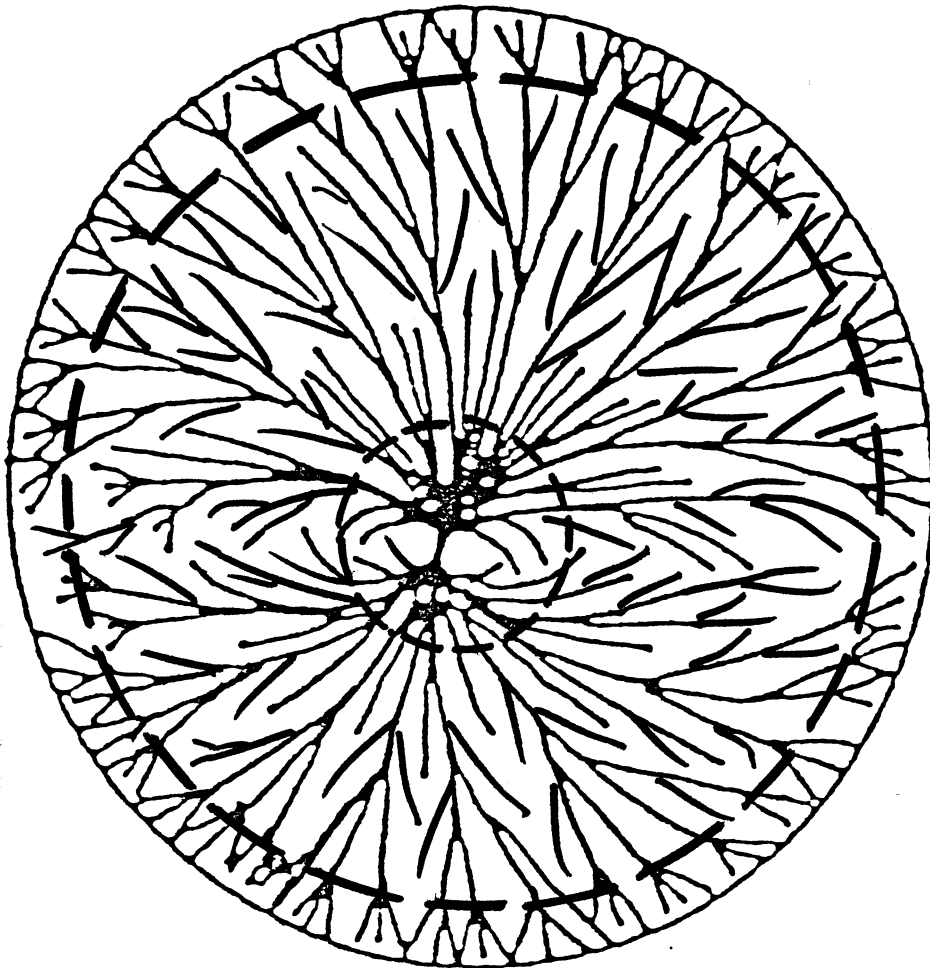


Fig. 3. Diagrammatic depiction of a carrot extract biocrystallogram with the central, median and peripheral zones.

image is inherently not a texture. Consequently it would be desirable to include structural aspects into the texture analysis, taking into consideration the zonal structure of a biocrystallogram. Since the centre location affects the zonal structure, it may be hypothesised that reducing the within- group variation in centre location in different groups of images will affect the image analysis parameters, and thereby lead to an improved classification.

2. Objective

The objective of the present study was to develop procedures for image analysis of biocrystallogram images, based on texture analysis including parameters from first- and second-order grey level statistics. The objective was specifically to examine the distribution of textural information, relevant for classification purposes, over the biocrystallogram image, thereby taking into consideration the zonal structure of a biocrystallogram.

The effects of applying various types and sizes of ROI on the image analysis parameters, and thereby on classification, were examined. Furthermore the effects of reducing the within-group variation in centre location were examined.

A so-called degradation series including seven groups of images, representing discrete 'treatment levels', was applied as the basis of the image analysis.

3. Materials and methods

3.1. Carrot extract degradation series

In biocrystallization investigations degradation series have been applied in connection with assessing the quality of agricultural products. A plant extract is produced on the basis of e.g. a carrot sample, and stored at 6°C, thereby exposed to a non-specific autolytic and microbial degradation. When producing biocrystallograms over a number of successive days systematic progressive changes in morphological features are observed, specifically a dissolution of the original co-ordinated structure. This is a parallel to the chemical degradation of macromolecules such as proteins into smaller molecules which can be followed analytically. The speed with which this dissolution takes place is regarded a qualitative parameter, analogous to storage ability (Engqvist, 1970).

In the actual series 59 biocrystallograms were produced on the basis of a carrot extract over 7 consecutive days. From each day nine biocrystallograms were available, except for day 6 with only five. For each biocrystallogram was applied 2.0 ml 10% aqueous extract, 0.8 ml of a 10% aqueous solution of dihydrated CuCl_2 , and distilled water added up to a total volume of 6.5 ml/biocrystallogram. The crystallisation underlay was supplied by circular glass plates, thickness 2 mm, diameter 90 mm. The combined evaporation and actual crystallisation took place over approximately 17 h. For a detailed description of techniques and experimental conditions applied, see Andersen et al. (1998).

The biocrystallograms contained 1–3 distinct centres. The centres were located anywhere between the geometric centre and the periphery. As the image analysis procedures were designed exclusively for single-centred biocrystallograms, out of the 59 biocrystallograms produced 26 were multi-centred and were discarded, leaving 33 for further investigation. From each of the 7 days 3, 8, 5, 6, 4, 3 and 4, biocrystallograms were available. It was presupposed that the number and location of the centres held no discriminative information concerning the samples.

In order to specify subsets based on a reduction in the within-group variation in centre location, the smallest distance in mm from the crystallisation centre to the periphery of the image along the vertical axis was measured, and the average distance was computed. From each of the 7 days the three images having a distance closest to the average were selected for a subset of 21 images. Due to the availability of only three single-centred biocrystallograms from day 1 and 6, these subsets were identical to the corresponding total sets.

On the basis of visual examination the biocrystallograms from the first degradation day could be divided into three zones: the central, the median zone, and the peripheral zone. The centre area exhibited a relatively large variation in size and morphological features within that day, some centres being characterised by a cross-like formation of crystals, others being masked by a dense structure of more or less curved crystal formations. The overall co-ordination of the median zone decreased gradually from day 1 to 7, and the outer part of the zone gradually dissolved, whereby an expanding crystal-free transition zone appeared between the median and the peripheral zone, see Fig. 4a–c. As a result the radial extension of the combined central and median zones gradually decreased from approximately 90% at day 1 to approximately 80% at day 7, and the extension of the transition zone increased correspondingly. No marked differences in morphological features were observed in the peripheral zone (approx. 90–100%) during the degradation series.

3.2. *Image analysis software*

The applied image analysis software termed ACTA (Applied Crystallisation Texture Analysis) was developed in co-operation between the Department of Mathematical Modelling at The Technical University of Denmark, and the Departments of Mathematics and Physics, and Agricultural Science, at The Royal Veterinary and Agricultural University. Procedures from BMDP Statistical Software (Dixon, 1985) were included in connection with the classification procedures. The hardware employed was a PC (Pentium 100 MHz, 64 MB RAM, 2 GB harddisk) running under the Linux operating system. For a detailed software and application instruction see Henriksen (1998).

The image analysis procedures consisted of the following five steps:

1. Preparation of biocrystallograms for analysis.
2. Specification of ROI.
3. Calculation of first-order grey level statistics.
4. Calculation of second-order grey level statistics.

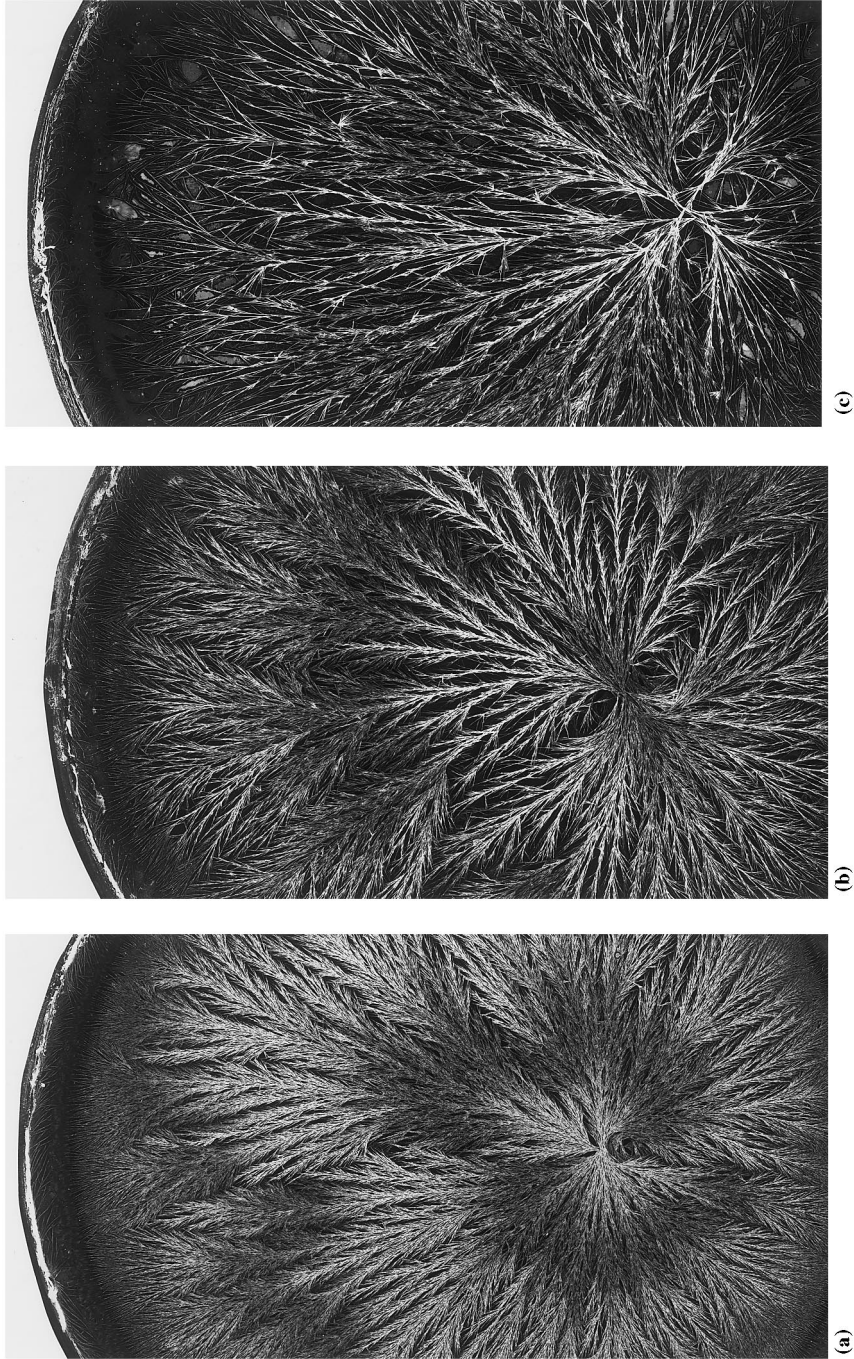


Fig. 4. (a–c). Enlarged photographic sections from biocrystallograms from the carrot extract degradation series representing (from l. to r.): (a) Day 1; (b) Day 4; and (c) Day 7.

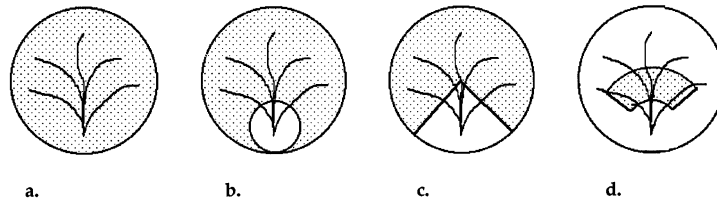


Fig. 5. (a–d). Diagrammatic depictions of the four main types of ROI: (a) Circle; (b) Circle–circle; (c) Sector; and (d) Segment. The dotted areas represent the regions-of-interest.

5. Classification of images.

3.3. Preparation of biocrystallograms for analysis

The biocrystallogram was scanned in 600 dpi 8-bit greyscale, by means of a dias-scanner.¹ The biocrystallogram was placed with the crystal-covered surface downwards, the crystallisation centre located so that this would appear in the lowest position on the image. The scanned image was cut to the size of 2112×2112 pixels, so that the top, bottom, right and left represented the framing of the image. The crystallisation centre was marked interactively.

The image was converted to HIPS format and reduced to a float image of half the size by low-pass filtering followed by sub-sampling. The specified ROI (see below) was normalised by a histogram match to a Gaussian distribution, in order to ‘calibrate’ the images. The filtering and sub-sampling procedure was then repeated four times, without the histogram match, to produce a total of five images with different resolutions. The five images are referred to as scale 1, 2, 4, 8 and 16 images (in the following abbreviated S1–S16).

3.4. Specification of ROI

The geometric centre and the crystallisation centre serve as reference points for defining the image spatially, corresponding to the origin in a Cartesian system of co-ordinates with four quadrants. Thereby it is possible to specify different types of ROI. The part of the image above the horizontal axis passing through the geometric or the crystallisation centre, corresponding to first and second quadrant, is referred to as the upper part of the image, and the part below the axis as the lower part.

See Fig. 5a–d for a diagrammatic depiction of the following four main types of ROI which can be specified for analysis:

1. Circle. Thus Circle–80 is a circle of 80% radial extension around the geometric centre. Here 0% equals the geometric centre, and 100% equals the periphery of the image.

¹ This device scans projected images of slides (transparencies) produced from photographs.

2. Circle–circle. Thus Circle–circle–80–20 is a circle of 80% radial extension around the geometric centre, minus a circle of 20% radial extension around the crystallisation centre, divided by the vertical axis. In the latter percentage measure 0% equals the crystallisation centre, and 100% the periphery of the image on the upper vertical axis.

3. Sector. Thus Sector 80–90 is a circle of 80% radial extension around the geometric centre, minus a radial section of 90° in the lower part of the image, divided by the vertical axis.

4. Segment. Thus Segment 30–60–90 is a circle segment specified relative to the crystallisation centre, in the upper part of the image. The first two numbers specify a radial interval on the upper vertical axis. Here 0% equals the crystallisation centre, and 100% the periphery of the image on the upper vertical axis. The third number represents the angle at the crystallisation centre, divided by the upper vertical axis.

The first three main ROIs are referred to as circular.

3.5. Calculation of first-order statistical parameters

The so-called histogram is a calculation of the grey level distribution of all pixels in the ROI. Depicted graphically, the first axis represents grey levels (0–255), the second axis the number of pixels as a function of the specific grey level, see Fig. 6.

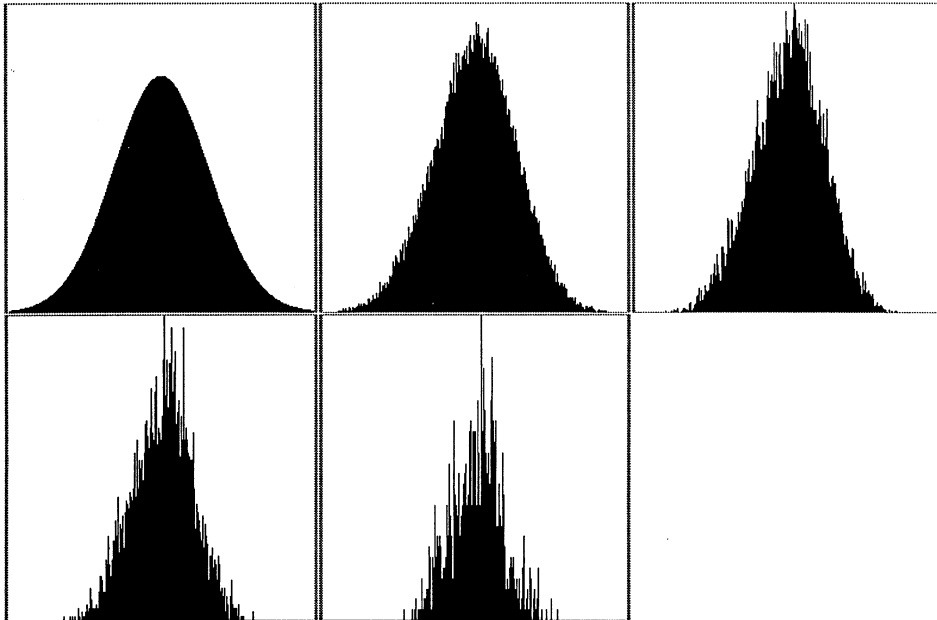


Fig. 6. Histograms of a carrot extract biocrystallogram image. From left to right and upper to lower: scale 1, 2, 4, 8 and 16.

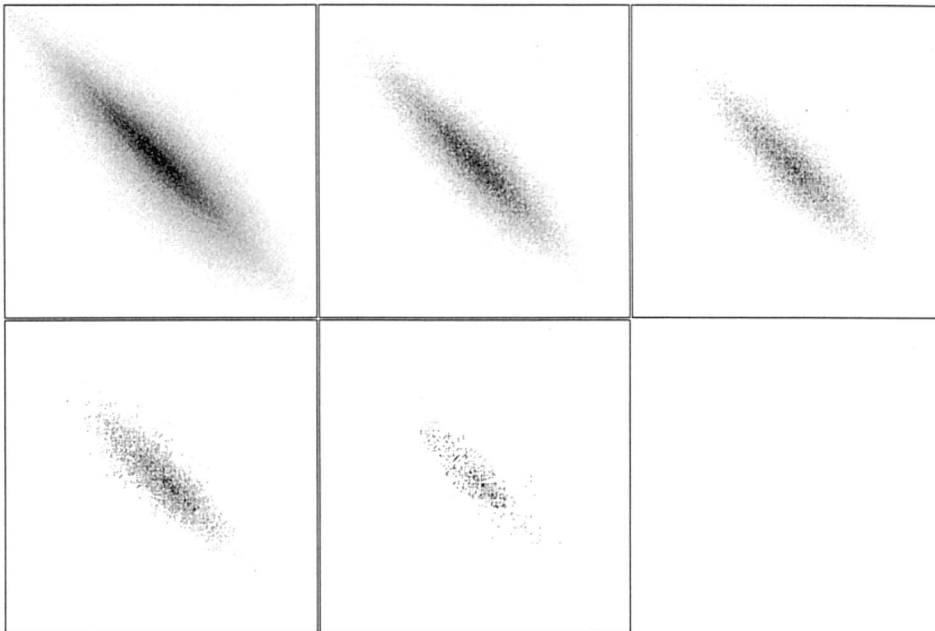


Fig. 7. GLCM (grey level co-occurrence matrix) of a carrot extract biocrystallogram image. From left to right and upper to lower: scale 1, 2, 4, 8 and 16.

From each histogram a total of eight first-order statistical parameters are derived: Mean, Variance, Covariance, Skewness, Kurtosis, Median, Entropy and Energy. These parameters are computed for the four resolution scales S2–S16, giving a total of 32 observations for each image. For computational definitions, and a closer description of the parameters, see Carstensen (1992, 1993).

3.6. Calculation of second-order statistical parameters

The so-called grey level co-occurrence matrix (in the following abbreviated: GLCM) is a calculation of the neighbour relationships between pixels in the ROI. Depicted two-dimensionally, the first axis represents the grey levels of pixels, the second axis the grey level of the corresponding neighbouring pixels. Fig. 7 gives an overall impression of the frequency with which a first axis grey level pixel corresponds to a given second axis grey level neighbouring pixel, by the 'blackness' of an area. A total of 15 parameters were derived from each GLCM: Energy, Entropy, Maximum probability, Correlation, Diagonal moment, Kappa, Difference energy, Difference entropy, Inertia, Inverse difference moment, Sum energy, Sum entropy, Sum variance, Cluster shade, and Cluster prominence. The parameters were computed for the five resolution scales, giving a total of 75 observations for each image. For computational definitions, and a closer description of the parameters, see Carstensen (1992, 1993). When combining the histogram and GLCM parameters a total of 107 observations was available for each image.

3.7. Classification of images

By means of stepwise discriminant analysis, based on the 107 observations, a specific set of histogram and GLCM parameters was selected which provided the best discrimination relative to the original 7 degradation days. The selection criterion, by which a parameter was included in the final set of parameters, was an *F*-test comparing the two sets which, respectively, included and excluded the parameter in question. An *F*-value of 4.0 was applied, which can be regarded a non-restrictive ‘standard’ value.

Two types of classification were performed:

1. An ordinary classification in which the individual image was classified on the basis of all images included in the analysis.
2. A jack-knifed classification in which the image was classified on the basis of all images except the image in question.

The posterior probabilities with which an image belonged to each of the original groups were computed. The individual image was then classified as belonging specifically to that original group to which it had the highest posterior probability of belonging. By matching the number of correct classifications made in each of the groups with the number of images in the original groups, a classification score was computed for each group. Subsequently the summarised classification score was computed, involving all correct classifications relative to the total number of images in the original groups.

In the following the summarised score, if not otherwise specified, is listed as ordinary/jack-knifed, e.g. 74.5/82.6 (in%). Comparing the ordinary and jack-knifed classification scores gives a general impression of the degree to which the image in question contributes to the group characteristics, and thereby an impression of the homogeneity of the images in a specific group.

3.8. Classification of images

A total of 18 sub-ROIs from the four main ROIs were specified, in order to examine which zones contained relevant information for classification purposes, and to examine the respective effects of including and excluding the centre area. Four recurring radial extensions of 70, 80, 90 and 100% were defined on the basis of the geometric centre, corresponding approximately to the radial extensions of the zones, including the transition zone. These radial extensions were applied for Circle, for Circle-circle-xx-15, and for Sector-xx-90. From visual observation it was estimated that subtraction of a minor circle of 15% and a sector of 90° would effectively remove the morphological characteristics of the centre area.

Segment-30-60-90, with a radial extension from 30 to 60% and a 90° angle, approximately corresponds to the area reported by some authors to contain the most consistent information for classification purposes (Engqvist, 1970). Using this ROI as a basis, the following ROIs with equal radial extensions were applied: Segment-00-30-90, Segment-30-60-90, and Segment-60-90-90. To test combinations of these segments, Segment-00-60-90, Segment-30-90-90, and Seg-

ment–00–90–90 were additionally included. The total set as well as the subset were analysed and classified.

In addition to the classifications performed by means of stepwise discriminant analysis, for a limited number of parameters, showing a monotonic relationship with degradation day number, linear regressions were performed. Furthermore a number of multiple linear regressions were performed. These were based on a set of parameters, selected on the basis of an *F*-test, comparing the two sets which respectively included and excluded the parameter in question. For the selection criterion, by which a parameter was included in the final set of parameters, an *F*-value of 4.0 was applied.

4. Results

4.1. Circle, circle–circle and sector

See Table 1 for classification scores, and the approximate sizes of the 12 circular sub-ROIs. When comparing the classification scores based on the total set we observe the highest score in Sector 70–90: 84.8/60.6. A jack-knifed score of 60.6 is also found in Sector–100–90. When comparing the scores based on the subset we observe the highest score in Circle 100: 100.0/81.0.

When comparing the average ordinary scores based on the total set and the subset, from the three main types of ROI, the total set scores are, respectively, 42, 44 and 16% lower than the subset scores. In the case of Circle and Circle–circle, when comparing the average jack-knifed scores based on the total set and the subset, the total set scores are, respectively, 44 and 43% lower than the subset scores. In the case of Sector the average jack-knifed score based on the subset is 9% higher than the total set score.

When comparing the average ordinary and jack-knifed scores based on the total set, from the three types of ROI, the ordinary scores are, respectively, 52, 51 and 35% higher than the jack-knifed scores. When comparing the scores based on the subset the ordinary scores are, respectively, 50, 52 and 71% higher than the jack-knifed scores.

The following four parameters were used for classification of the subset of Circle–100:

S2 Histogram energy, S4 Histogram median, S8 Histogram mean and S16 GLCM Diagonal moment. No single one of these parameters was able to discriminate more than two blocks of days, and correspondingly none of the parameters showed a strong linear relationship to degradation day number.

In Circle–70 the parameter S16 GLCM Difference energy shows an overall monotonic relationship with degradation day number, see Fig. 8a. A linear regression was performed with the following estimated formula and R^2 :

$$\begin{aligned} \text{S16 GLCM Difference energy} &= 2.99 \cdot 10^{-3} \cdot (\text{degradation day no.}) \\ &\quad - 2.25 \cdot 10^{-2}; R^2 = 0.81. \end{aligned}$$

Table 1
Information concerning biocrystallogram images from the carrot extract degradation series^a

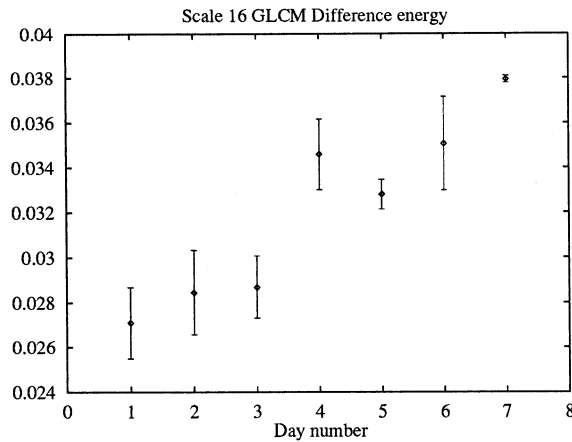
ROI	Classification score for the total set		Classification score for the subset		Approx. size of ROI at scale 1 (in pixels)
	Ord. (in%)	Jack. (in%)	Ord. (in%)	Jack. (in%)	
<i>Circle</i>					
-70	69.7	45.5	85.7	52.4	429 000
-80	54.5	36.4	85.7	47.6	560 000
-90	54.5	42.4	95.2	71.4	709 000
-100	78.8	45.5	100.0	81.0	875 000
Average score:	64.4	42.5	91.7	61.1	
<i>Circle-circle</i>					
-70-15	63.6	36.4	95.2	71.4	390 000
-80-15	57.6	45.5	81.0	47.6	521 000
-90-15	57.6	45.5	90.5	61.9	670 000
-100-15	72.7	39.4	95.2	57.1	836 000
Average score:	62.9	41.7	90.5	59.5	
<i>Sector</i>					
-70-90	84.8	60.6	71.4	52.4	321 000
-80-90	48.5	27.3	90.5	38.1	420 000
-90-90	60.6	48.5	66.7	38.1	532 000
-100-90	72.7	60.6	81.0	52.4	657 000
Average score:	66.6	49.3	77.4	45.3	
<i>Segment</i>					
-00-30-90	54.5	30.3	57.1	33.3	39 000
-30-60-90	45.5	33.3	42.9	38.1	117 000
-60-90-90	60.6	54.5	81.0	52.4	195 000
-00-60-90	66.7	48.5	66.7	47.6	156 000
-30-90-90	57.6	36.4	61.9	52.4	312 000
-00-90-90	42.4	27.3	85.7	61.9	351 000
Average score:	54.6	38.4	65.9	47.6	

^a Type of ROI; classification scores for the total set of 33 images, and the subset of 21 images, including ordinary and jack-knifed classification scores, and average scores; approximate size of ROI at scale 1.

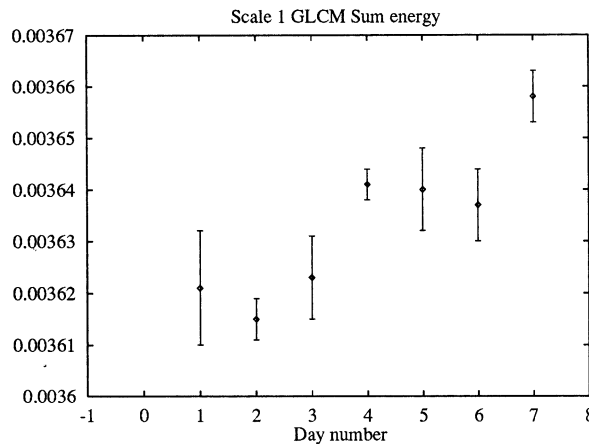
In Circle–80 the parameter S1 GLCM Sum energy shows a monotonic relationship to degradation day number, discriminating three groups of days: 1–3, 4–6, and 7, see Fig. 8b.

For results from linear and multiple linear regressions for the four Circles, see Table 2. It is seen that the R^2 values are markedly higher when applying multiple linear regression. In the case of Circle–90 a strong correlation is found ($R^2 = 0.98$), based on eight parameters, compared to a relatively weak correlation when using only a single parameter ($R^2 = 0.71$).

For the parameters applied for multiple linear regression in connection with Circle–70, 80, 90 and 100, see Table 3.



a.



b.

Fig. 8. (a–b). Two image analysis parameters applied for classification of the 21 image subset plotted against degradation day number. a. S16 GLCM Difference energy, applied for Circle–70; and (b) S1 GLCM Sum energy, applied for Circle–80.

Table 2

Information concerning linear and multiple linear regressions performed, based on image analysis parameters versus degradation day number^a

ROI Circle	Linear regression parameters applied	R^2	Multiple linear regression no. of parameters applied	R^2
-70	S16 GLCM difference energy	0.81	4	0.92
-80	S1 GLCM Entropy	0.74	2	0.83
-90	S16 GLCM sum energy	0.71	8	0.98
-100	S1 GLCM diagonal moment	0.69	4	0.86

^a Specification of size of circular ROI; parameters applied for linear regression; R^2 -values; number of parameters applied for multiple linear regression; R^2 -values.

4.2. Segment

See Table 1 for the classification scores, and the approximate sizes of the six Segments.

When comparing the scores based on the total set we observe that Segment-00–60–90, a combination of the lower and middle segment, shows the highest score: 66.7/48.5. When comparing the scores based on the subset we observe that Segment-00–90–90, a combination of all three equal radial Segments, shows the highest score: 85.7/61.9.

When comparing the average ordinary scores based on the total set and the subset, the total set score is 21% lower than the subset score. When comparing the average jack-knifed scores based on the total set and the subset, the total set score is 24% lower.

Table 3

Parameters applied for the classification of Circle-70, -80, -90 and -100

Circle-70	Circle-80
S16 GLCM Difference energy	S1 GLCM Entropy
S16 GLCM Inverse difference moment	S8 GLCM Correlation
S4 GLCM Inertia	
S16 GLCM Sum entropy	
Circle-90	Circle-100
S1 GLCM Cluster shade	S1 GLCM Diagonal moment
S16 GLCM Difference energy	S4 Histogram skewness
S2 GLCM Maximum probability	S4 GLCM Diagonal moment
S1 GLCM Cluster prominence	S2 GLCM Diagonal moment
S16 Histogram entropy	
S8 GLCM Sum entropy	
S16 Histogram energy	
S8 GLCM Inverse difference moment	

When comparing the average ordinary and jack-knifed scores based on the total set the ordinary score is 42% higher than the jack-knifed. In the case of the subset the ordinary score is 38% higher.

5. Conclusion

It was demonstrated in the present study that by means of the developed image analysis procedures termed ACTA, it was possible to classify correctly the applied 21 biocrystallogram images, originating from a carrot extract degradation series including seven groups.

The applied images represent biocrystallograms typically found in biocrystallization investigations of agricultural products reported in the literature, concerning the range in morphological features found in the seven groups, and concerning the range in morphological features found in each group. Concerning the number of replicates in each group, i.e. three, this represented the lowest possible number as a basis of classification, performed by means of stepwise discriminant analysis.

Earlier studies have been based on a circular ROI, with no restrictions on the number of centres or their location. The high classification scores seen in the present study, based on single-centred images with a restriction on the variation in centre location, emphasise the fact that a biocrystallogram is inherently not a texture which can be perceived as having a spatially homogeneous distribution of morphological features. Indeed the textural information relevant for classification purposes is dependent on the location of the centre from where the ramification structure expands in all directions. Thus when applying the subset of 21 images, in which the within-group variance is reduced, markedly higher classification scores are found. Thus out of the 72 scores listed, six scores ranging from 90 to 100% are all found in the subset. For future studies it may therefore be recommended to impose a restriction on the variation in centre location, since this variation is a phenomenon which may be expected to appear in biocrystallograms from all agricultural products.

The relatively large pairwise differences in scores between the ordinary and jack-knifed scores in favour of the ordinary, in both the total set and the subset, indicate inhomogeneity among the images in the individual groups. However a detailed interpretation is not possible on the basis of the present study.

We found that discriminative information was present in all zones. Thus the experience of some authors applying visual classification, that the most consistent and discriminative information is found in the upper median zone, corresponding approximately to Segment–60–90–90, was not confirmed. When comparing the four main types of ROI we found that the three circular ROIs generally showed higher scores than Segment, i.e. the larger ROIs performed better than the smaller ones. In fact we found the single 100% score in Circle–100, and no general increase in score was found when excluding the centre.

The procedure of applying five resolution scales, in combination with a total of 23 first- and second-order parameters, appears relevant, and there are no indications of marked superiority of individual parameters in classifying the images. We observe that five out of eight histogram parameters, and 11 out of 15 GLCM parameters are applied for classification in connection with the few results listed here. Similarly parameters are applied from all five scales, in the case of GLCM Diagonal moment indeed four out of five scales are applied.

The developed image analysis procedures can provide reproducible data in connection with biocrystallization investigations of agricultural products, which can be statistically correlated to other experimental data. Thereby they may increase our understanding of the qualitative information obtained from biocrystallization investigations of the quality of agricultural products, concerning e.g. the effects of farming systems, fertilisation, storage ability and processing.

The fact that only single-centred images can be applied represents a limitation. However out of the 26 initially discarded multi-centered images, the majority were two-centred, and again the majority of these contained one dominant centre, i.e. the second centre only occupied a minor part of the overall ramification structure. Including these images, based on relevant restrictions on the location and the relative dominance of the second centre, may increase considerably the number of images available as a basis of classification.

For future studies it may be recommended to introduce polar, non-Cartesian systems of co-ordinates which to a higher degree reflect the zonal structure of biocrystallograms. Besides the type of classification presented here, whereby the image in question and the images constituting the classification groups originated from the same investigation, other types of classification might be developed. Thereby individual images originating from separate investigations might be classified and compared relative to a set of standard scales, each containing a relevant number of discrete groups. Thus images originating from a carrot sample might be classified relative to scales, produced on the basis of carrot samples distinguished by experimental factors such as physiological maturity, post-harvest degradation, and destructive processing by means of e.g. irradiation. Actual structure analysis procedures, whereby the crystal structure is analysed relative to a crystal-free background, will no doubt improve the discriminative and classificatory power of the image analysis. Finally the progress in hardware technology, with ever increasing computational capacity and speed, may prove it advantageous to include colours without unduly time-consuming procedures.

Acknowledgements

We thank the Danish Agricultural Ministry for the fundings which made the present study possible. We thank Jens Michael Carstensen, Department of Mathematical Modelling, The Technical University of Denmark, for valuable advice and comments throughout the study.

References

- Andersen, J.-O., Laursen, J., Kølster, P., 1998. A refined biocrystallization method applied in a pictomorphological investigation of a polymer. *Elemente Natur*. 68, 1–20.
- Balzer, U., Balzer, F., 1993. Picture-developing methods. Effect of three farming systems (bio-dynamic, bio-organic conventional) on yield and quality of beetroot (*Beta Vulgaris* L.Var. *Esculenta* L.) in a seven year crop rotation. *Acta Horticult.* 339, 11–31.
- Carstensen, J.M., 1992. Description and simulation of visual texture. Ph.D. thesis. Department of Mathematical Modelling, Technical University of Denmark.
- Carstensen, J.M., 1993. Co-occurrence feature performance in texture classification. In: Høgda et al. (Eds.), *Proceedings of the Eight Scandinavian Conference on Image Analysis*. University of Tromsø, Tromsø, Denmark, pp. 831–838.
- Connors, R., Trivedi, M., Harlow, C., 1984. Segmentation of a high-resolution urban scene using texture operators. *Comput Vis. Graphics Image Proc.* 25, 273–310.
- Dixon, W.J. (Ed.), 1985. *BMDP Statistical Software*. University of California Press, Berkeley, CA.
- Engqvist, M., 1970. *Gestaltkräfte des Lebendigen*. Klostermann, Frankfurt am Main.
- Engqvist, M., 1989. *Qualitätsprüfung an Gemüse durch die Kupferchlorid-Kristallisationsmethode*. Järna.
- Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybernet.* 3 (6), 610–621.
- Henriksen, C.B., 1998. *ACTA 1.0. User's Manual*. Internal Report, Department of Agricultural Science, The Royal Veterinary and Agricultural University. Copenhagen.
- Kleber, W., Steinike-Hartung, U., 1959. Ein Betrag zur Kristallization von Kupfer(II)-Chlorid-Dihydrat aus Lösungen. *Zeitschr. Kristallogr.* 111, 213–234.
- Lê Gia, V., Teisseron, G., Michel, M.C., Cauffet, G., 1996. Application of texture analysis for the crystallisation image characterisation. In: Grossmann, H. et al. (Eds.), *Proceedings from the Third European Research Symposium: Image Analysis for Pulp and Paper Research and Production*, Stockholm 1995.
- Lê Gia, V., 1996. *Extraction de parametres de texture d'images de croissance cristalline dendritique*. Doctoral thesis. Lime, UJF/IUT de Grenoble.
- Neuhaus, A., 1957. Die Kristallographischen Grundlagen der Diagnostischen Kristallizationen. In: Selawry, A., O. (Eds.), *Die Kupferchlorid-Kristallization in Naturwissenschaft und Medizin*. Fischer Verlag, Stuttgart, pp. I–XXXV.
- Parkkinen, J., Selkäinaho, K., Oja, E., 1990. Detecting texture periodicity from the co-occurrence matrix. *Pattern Recogn. Lett.* 11 (1), 43–50.
- Pfeiffer, E., 1931. *Studium von Formkräften an Kristallizationen*. Naturwissenschaftliche Sektion am Goetheanum, Dornach.
- V. Hahn, F.V., 1962. *Thesigraphie*. Franz Steiner Verlag, Wiesbaden.