JOURNAL OF CHEMOMETRICS, VOL. 11, 131–139 (1997)

CLASSIFICATION OF SILVER HALIDE MICROCRYSTALS VIA *K*-NN CLUSTERING OF THEIR SHAPE DESCRIPTORS

VOLODYMYR V. KINDRATENKO,[1] BORIS A. TREIGER[1] AND PIERRE J. M. VAN ESPEN[1]
[1] *Micro- and Trace Analysis Center (MiTAC), Department of Chemistry, University of Antwerp (UIA), B-2610 Wilrijk, Belgium*

SUMMARY

A method for the classification of tabular grain silver halide microcrystals according to their shape is presented. Various approaches of shape analysis and recognition and their applicability for the given problem are discussed. Shape descriptors obtained from Fourier power spectra are used to describe the shape of microcrystals. The classification of the shapes is based on nearest neighborhood algorithms. Results of the classification by four different algorithms are compared. The fuzzy four-nearest-neighbor classifier was found to be the most appropriate one. © 1997 by John Wiley & Sons, Ltd.

Journal of Chemometrics, Vol. **11,** 131–139 (1997) (No. of Figures: 4 No. of Tables: 3 No. of Refs: 12)

KEY WORDS shape description; image processing; *K*-NN clustering

INTRODUCTION

Tabular grain silver halide microcrystals are used as light sensitive material in modern photographic emulsion. They are produced in a reactor vessel by the precipitation reaction of Ag^+ with Cl^-, Br^- or I^- ions. Their shape, size and composition are determined by the growth (precipitation) conditions and in turn determine the properties and quality of the photographic material. Knowledge about the size and shape distribution of these microcrystals is important for the optimization of the precipitation process and for the study of the photographic properties of light sensitive films.[1] This information can be obtained through analysis of *individual microcrystals*, namely (i) recognition of individual microcrystals according to their shape and (ii) counting differently shaped microcrystals. Therefore in this paper we consider the problem of classification of tabular silver halide microcrystals according to their shape.

In the literature only a few publications deal with this specific problem.[2, 3] The methods described are based on the analysis of geometrical parameters of microcrystals obtained from their scanning electron microscope (SEM) images. The disadvantage of this approach is that it is very difficult to derive geometrical parameters from SEM images with sufficient accuracy. It is, however, not necessary to base a classification on geometrical parameters, since other suitable descriptors can be used as well. It has been shown[4] than many artificially generated and real shapes can be characterized by their Fourier power spectra. Fourier coefficients have been applied successfully to the shape description of tabular silver halide microcrystals.[5] In this work we discuss the problem of using the calculated Fourier coefficients for the purposes of automated classification. Four different algorithms for the classification of microcrystals were tested: the crisp and fuzzy *K*-nearest-neighbor and the crisp and fuzzy one-nearest prototype classifiers.

Correspondence to: Pierre J. M. Van Espen

CCC 0886–9383/97/020131–09 $17.50

Received May 1995
Accepted August 1996

METHOD

Description of the shape of microcrystals

A general problem of shape quantification is the development of a variable that unambiguously describes (and separates) shape characteristics of different objects independently of magnitude and orientation. A large number of approaches can be found in the literature,[6] the most commonly applied being Fourier coefficients, chain coding, fractal dimensions and dynamic shape factors.

Fractal analysis is an appropriate tool for the characterization of complex shapes. This is not the case for tabular silver halide microcrystals, since they have quite regular shape which can be described in the frame of Euclidean geometry. The dynamic shape factors introduced by Medalia[7] are based on the representation of an object as an ellipsoid with equivalent radii of gyration about the central principal axes. For the shape characterization of tabular silver halide microcrystals this method is not selective enough: different shapes result in very similar parameters.

Both methods mentioned above attempt to condense all the details of the shape into a single number. There can, however, be an unlimited number of visually different shapes with the same fractal dimension or similar shape factors. At the other extreme it is possible to use a large number of parameters to preserve all shape information in enough detail, e.g. using Fourier descriptors or chain coding. Chain code shape representation is an effective method which in combination with Fourier analysis proves to be useful in a surprising number of cases.[6] In fact, chain coding is a shape-unrolling technique producing a list of numbers which can be interpreted as the changes in slope versus position along the particle's silhouette edge. In the case of silver halide microcrystals such a description is more difficult to use and interpret than a Fourier method, which is described below.

Fourier shape description is done in the following way. The particle's silhouette edge is described in polar co-ordinates as a function of the radius R versus the swept angle θ from the center of gravity of the particle.[6] It can be expanded in a *Fourier series*

$$R(\theta)=R_0+\sum_{n=1}^{\infty}[A_n\cos(n\theta)+B_n\sin(n\theta)]$$

where R_0 is the radius of a circle having the same area as the profile generated by the particle's silhouette edge, θ is the polar angle measured from an arbitrary reference line and A_n and B_n are the Fourier series coefficients. Each term of this series is called a *harmonic*. The harmonic amplitude $\sqrt{(A_n^2+B_n^2)}$ indicates the contribution of the nth component to the overall shape,[8] where n is the harmonic order. In practice the Fourier series coefficients are calculated as

$$A_n=\frac{1}{\pi}\sum_{\theta}R(\theta)\cos(n\theta),\qquad B_n=\frac{1}{\pi}\sum_{\theta}R(\theta)\sin(n\theta)$$

where $\theta\in[0;\,2\pi)$.

The sequence of harmonic amplitudes $\{\sqrt{(A_n^2+B_n^2)}\}$ forms a vector (e.g., vector x) which can be considered as the *Fourier shape descriptor* of the shape of the given particle. Alternatively, the squared harmonic amplitude $A_n^2+B_n^2$ is often used as a Fourier-based shape descriptor.

This method causes problems if the shape is re-entrant so that the radial vector is multivalued.

However, in our application this only occurs for certain overlapping silver halide microcrystals and this fact can be used for their immediate recognition.

Classification of microcrystals via their shape descriptors using *NN* algorithms

Let $z_1, z_2, \ldots, z_c$ be a set of c prototype vectors (e.g. the centers of classes in which each member is represented by a vector obtained in the way described above) representing the c classes. The *crisp one-nearest prototype classifier* assigns an input sample vector y, which is of unknown classification, to the class of its nearest-neighbor prototype.[9]

Let $x_1, x_2, \ldots, x_q$ be a set of vectors representing the q labeled samples from the c classes. The *crisp nearest-neighbor classifier* assigns a input sample vector y, which is of unknown classification, to the class of its nearest neighbor. This idea can be extended to the *K nearest-neighbors*. However, when more than one neighbor is considered, the possibility that there will be a tie among classes with a maximum number of neighbors in the group of K nearest-neighbors exists. The usual way of handling this problem is to assign the sample vector to the class, of those classes that tied, for which the sum of distances from the sample to each neighbor in the class is minimal.

The idea of one-nearest prototype classifier can be extended to the *fuzzy one-nearest prototype classifier*. This classifier assigns a membership of a sample vector y to the ith class ($i = 1, 2, \ldots, c$) rather than assigning vector y to a particular class i. The basis of the algorithm is to assign membership $u_i(y)$ as a function of the vector's distance from all prototypes:

$$u_i(y)=\frac{1}{\|y-z_i\|^{2/(m-1)}}\left(\sum_{t=1}^{c}\frac{1}{\|y-z_t\|^{2/(m-1)}}\right)^{-1}$$

The *fuzzy K-nearest-neighbor classifier* assigns a membership value of a sample vector y in the class i as a function of the vector's distance from its K nearest-neighbors. Let $u_i(y)$ be the assigned membership of the vector y (to be computed) and u_{ij} be the membership in the ith class of the jth vector ($j=1, 2, \ldots, q$) of the labeled sample set. u_{ij} can be computed in the following way based on the K-nearest-neighbor rule. The K^* (not K of the classifier) nearest-neighbors to each jth sample (e.g., a sample from class t) are found and their membership in each ith class is assigned according to the equation

$$u_{ij}=\begin{cases}0{\cdot}51+0{\cdot}49\ n_i/K^*, & t=i\\ 0{\cdot}49\ n_i/K^*, & t\neq i\end{cases}$$

where the value n_i is the number of neighbors found which belong to the ith class. Finally, $u_i(y)$ is computed according to the equation

$$u_i(y)=\sum_{k=1}^{K}\frac{u_{ik}}{\|y-x_k^*\|^{2/(m-1)}}\left(\sum_{k=1}^{K}\frac{1}{\|y-x_k^*\|^{2/(m-1)}}\right)^{-1}$$

where $\{x_k^*\}$ ($k=1, 2, \ldots, K$) is the set of K nearest-neighbors of the given vector y. The parameter m determines how heavily the distance is weighted when calculating each neighbor's contribution to the membership value. As m increases, the neighbors are more evenly weighted and their relative distances from the point being classified have less effect.

EXPERIMENTAL

Backscattered electron images of tabular silver halide microcrystals were acquired with an integrated microscope and X-ray microanalyzer (IMIX) system on a Jeol JSM-6300 electron microscope using an electron energy of 20 keV, a beam current of about 1 nA and typical magnification of about 3000×. Images of more than 100 different microcrystals were collected.

The images were binarized with a correlation-criteria-based technique.[10] Contours of microcrystals were extracted from their binary images by the contour-following technique[11] and stored as a list of co-ordinates. The center of mass of the contour of each microcrystal was calculated as the sum of the corresponding contour's coordinates x and y divided by the number of points on the contour. The 360 radii measured from the center of mass to the contour points with step 1° were obtained for each microcrystal and normalized to the largest radius in order to obtain comparable Fourier shape descriptors. If some radial vectors are multivalued (i.e. the shape is re-entrant), the corresponding object is considered as belonging to the class of overlapping microcrystals.

Tabular grain silver halide microcrystals have pronounced triangular or hexagonal shape (see Figure

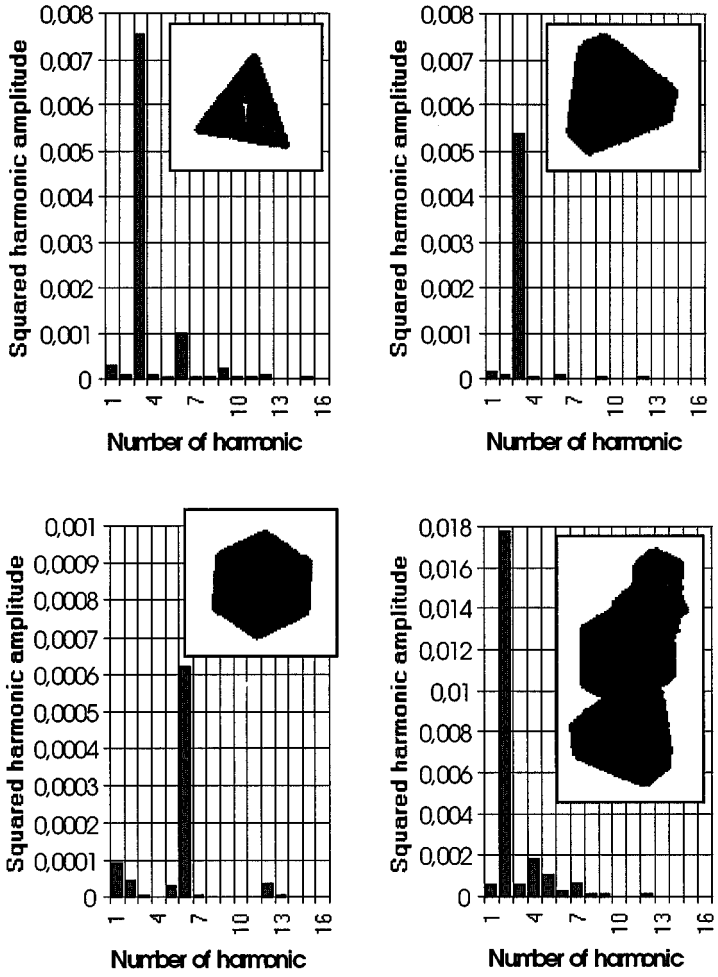

Figure 1. Differently shaped microcrystals and corresponding squared harmonic amplitudes

Table 1. Minimal, maximal and average values of the second, third and sixth squared harmonic amplitudes obtained from experimental data set. The set of labeled samples (prototypes) for the crisp and fuzzy four-nearest-neighbor classifiers was obtained by combining only minimal and maximal values. Average values of the second, third and sixth squared harmonic amplitudes were used as the set of prototypes for the crisp and fuzzy one-nearest prototype classifiers.

	Squared amplitudes of 2nd harmonic			Squared amplitudes of 3rd harmonic			Squared amplitudes of 6th harmonic		
Microcrystals	Min.	Max.	Ave.	Min.	Max.	Ave.	Min.	Max.	Ave.
Triangles	~0	0·0004	0·0002	0·0069	0·0081	0·0075	0·0004	0·0010	0·0007
Truncated triangles	~0	0·0004	0·0002	0·0008	0·0067	0·0037	0·0000	0·0004	0·0002
Hexagons	~0	0·0002	0·0001	~0	0·0004	0·0002	0·0003	0·0006	0·0005
Overlapping microcrystals	0·0044	0·0293	0·0168	0·0001	0·0045	0·0023	~0	0·0005	0·0003

1) which is directly related to their growing conditions. Previously it has been shown[5] that such microcrystals can be well described only by the second, third and sixth squared harmonic amplitudes: differently shaped microcrystals have essentially different squared amplitudes for the second, third and sixth harmonics. This fact is illustrated in Figure 1, where the first 16 squared harmonic amplitudes of the Fourier coefficients for the four types of microcrystals are shown. Thus our classification algorithms use only these three squared harmonic amplitudes.

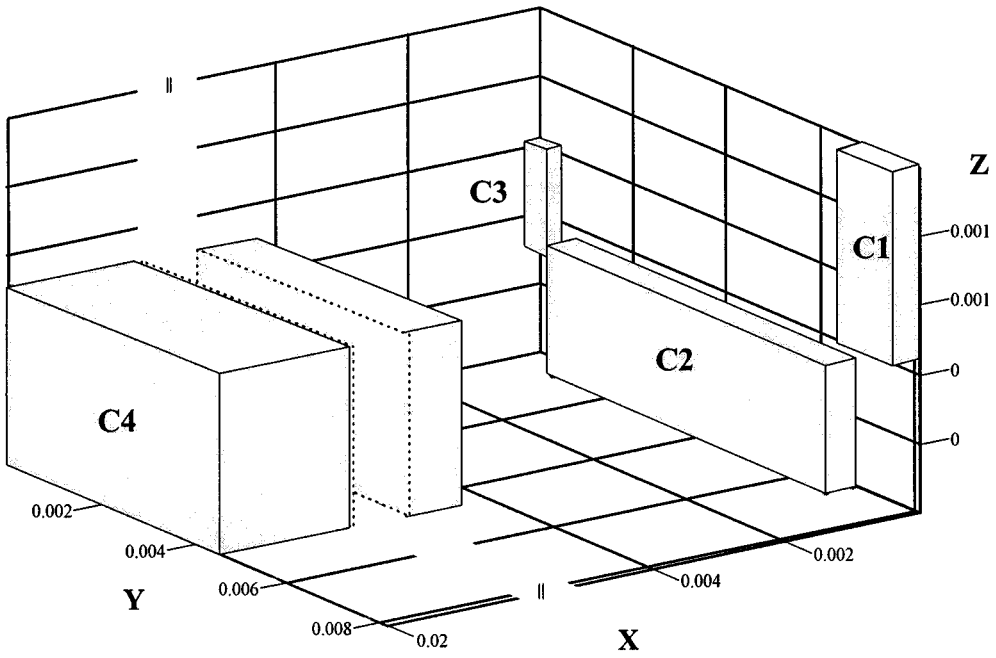

Figure 2. 3D space formed by squared amplitudes of second (*X*-axis), third (*Y*-axis) and sixth (*Z*-axis) harmonics. The corners of the parallelepipeds correspond to the eight prototypes for each class of microcrystals (C1, triangular microcrystals, C2, truncated triangular microcrystals; C3, hexagonal microcrystals, C4, overlapping microcrystals). The 3D space is divided by these four parallelepipeds into four separated 'class subspaces'

Table 2. Summary of classification of 65 microcrystals by crisp four-nearest-neighbor and one-nearest prototype classifiers.

Microcrystals	Total number of microcrystals	Correctly classified by crisp four-nearest-neighbour classifier	Correctly classified by crisp one-nearest-prototype classifier
Triangles	12	100%	100%
Truncated triangles	25	100%	32%
Hexagons	16	100%	100%
Overlapping microcrystals	12	92%	33%

The *K*-NN classification algorithms require a set of labeled samples (or prototypes) which represent the different classes. This set was generated by analyzing the data obtained from the set of collected microcrystals. Minimal and maximal values of the squared amplitudes of the second, third and sixth harmonics were determined for each class of microcrystals (see Table 1). Combining these values, the eight prototypes for each class of microcrystals were obtained (Figure 2). One additional prototype for each class of microcrystals was obtained by using the average values of the minimal and maximal squared amplitudes of the second, third and sixth harmonics. Thus each class of microcrystal shapes is represented by nine points in a 3D space formed by the squared amplitudes of the second, third and sixth harmonics. Such artificially generated labeled samples allow us to divide the 3D space into 'class subspaces' which are bounded by the parallelepipeds formed by the first eight points for each class (see Figure 2). The use of the original Fourier shape descriptors as labeled samples gave irreproducible results.

For the crisp *K*-nearest-neighbor and fuzzy *K*-nearest-neighbor classifiers all nine points of each class were used as the set of labeled samples. For the crisp one-nearest-prototype and fuzzy one-nearest prototype classifiers only the central points, i.e. the average values of the minimal and maximal squared amplitudes, were used as the set of labeled prototypes.

Agglomerates of overlapping microcrystals can be segmented by applying a segmentation technique.[12] After such segmentation, single microcrystals are processed in the way described above, starting from extraction of their contours.

RESULTS AND DISCUSSION

Results of the classification of microcrystal shapes by the crisp *K*-nearest-neighbor and crisp one-nearest-prototype classifiers are given in Table 2. We decided to use $K=4$ as the most suitable value after thorough analysis of all cases for $K=1-9$. We observed that in most cases the first four nearest-neighbors of any point (in the 3D space of the squared amplitudes of the second, third and sixth harmonics) corresponding to the microcrystal shape descriptors belong to one class.

The crisp four-nearest-neighbour classifier gave good results. At the same time the use of the crisp one-nearest prototype classifier results in a large number of misclassifications, especially of truncated triangles and overlapping microcrystals. These results are expected, because triangles and hexagons form small, compact clusters in the 3D space of the squared amplitudes of the second, third and sixth harmonics whereas truncated triangles and especially overlapping microcrystals show a large variety of shapes and therefore form large clusters.

Although the crisp four-nearest-neighbor classifier gives good results, the advantage of using fuzzy *K*-nearest-neighbor and fuzzy *K*-nearest-prototype classifiers can be understood by examining a microcrystal which has a truncated trapezoidal shape (see Figure 3). The crisp classifier will classify it in one of the four classes even if the truncated trapezoid does not belong to them. At the same time

the fuzzy four-nearest-neighbor classifier ($K=4$, $m=3$) gives values of the membership for the truncated trapezoid in every class. These values, however, are not high enough for making the decision about its belonging to any class (see Figure 3). Moreover, the fuzzy classifier allows us to do more precise classification of triangles. The triangular microcrystals are defined as microcrystals whose length of the short edges is smaller than 10% of that of the long one.[1] According to human observation it is not always obvious to classify some microcrystals in the class of triangles or truncated triangles. The fuzzy four-nearest-neighbor classifier gives in these cases values of membership in the class of triangles lower than 0·9. This fact can be used for recognizing such intermediate microcrystals.

The fuzzy one-nearest prototype classifier has the same disadvantage as the crisp one-nearest prototype classifier, i.e. it often misclassifies truncated triangles and overlapping microcrystals.

Differently shaped microcrystals	Results of classification by the crisp		Membership values, obtained by the fuzzy 4-NN classifier				Membership values, obtained by the fuzzy one-N prototype classifier				Finally classified as
	4-nearest neighbor classifier	1-nearest prototype classifier	C1	C2	C3	C4	C1	C2	C3	C4	
	triangle	triangle	1	0	0	0	0.88	0.07	0.04	0.1	triangle
	triangle	triangle	0.88	0.12	0	0	0.8	0.12	0.06	0.02	(inter-mediate) truncated triangle
	truncated triangle	truncated triangle	0.12	0.88	0	0	0.34	0.46	0.15	0.05	truncated triangle
	truncated triangle	hexagon	0	0.88	0.12	0	0.12	0.31	0.53	0.04	truncated triangle
	hexagon	hexagon	0	0	1	0	0.03	0.06	0.89	0.01	hexagon
	overlapping micro-crystal	overlapping micro-crystal	0	0.22	0	0.78	0.21	0.3	0.34	0.16	overlap-ping micro-crystal
	overlapping micro-crystal	hexagon	0	0.42	0.03	0.54	0.17	0.29	0.44	0.1	unknown

Figure 3. Comparison of results of classification by different algorithms. The column 'Finally classified as' was obtained from classification by the fuzzy four-nearest-neighbor classifier using the additional rule for the classification of truncated triangles described in the text. The same abbreviations as in Figure 2 are used.

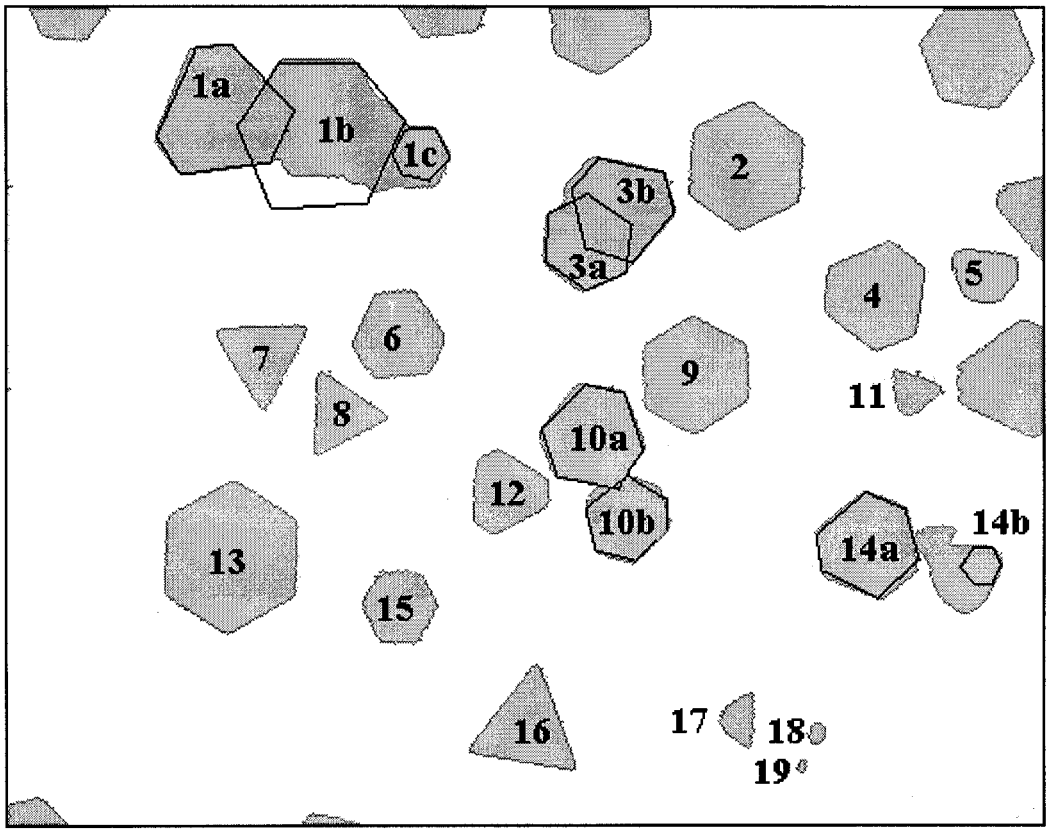

Figure 4. Binarized backscattered electron image. Labeled microcrystals are classified according to their shape as shown in Table 3

Table 3. Results of classification of silver halide microcrystals shown in Figure 3

#	Classified as	#	Classified as
1a	Truncated triangle	10a	Truncated triangle
1b	Incorrectly reconstructed	10b	Truncated triangle
1c	Too small	11	Too small
2	Hexagon	12	Truncated triangle
3a	Hexagon	13	Hexagon
3b	Truncated triangle	14a	Truncated triangle
4	Truncated triangle	14b	Too small
5	Unclassified	15	Hexagon
6	Truncated triangle	16	Triangle
7	Triangle	17	Too small
8	Triangle	18	Too small
9	Hexagon	19	Too small

For the experimental data set considered, the fuzzy four-nearest-neighbor classifier gave highest membership values of all overlapping microcrystals in their own class and only for one overlapping microcrystal did it give a high value of membership for the class of hexagonal microcrystals (0·81). Since this class is represented by a compact small cluster it is possible to specify the lowest value of membership of belonging to this class as 0·9. Actually, because we have some advance knowledge about differences between the classes of microcrystals, we can specify the lowest values of membership of belonging for the remaining classes as well. For triangles and hexagons this value is 0·9. For truncated triangles and overlapping microcrystals this value is 0·7. A microcrystal is unclassified if it has membership values in every class lower than the specified lowest values. One additional rule is used to classify a microcrystal as an (intermediate) truncated triangle: it must have a membership value in the class of triangles between 0·7 and 0·9 and a membership value in the class of truncated triangles larger than 0·1. An example of such a situation is given in Figure 3.

Thus the accurate classification of the shape of tabular grain silver halide microcrystals is possible only with the fuzzy four-nearest-neighbor classifier. The three other classification algorithms examined in this work, produce misclassifications. An example of results of the classification by the procedure described above is shown in Figure 4. In this example the shape of the overlapping crystals was reconstructed using the method described in Reference 12.

REFERENCES

1. I. Geuens, 'Development of micro and surface analysis methods for the quantitative characterization of silver halide microcrystals', *Ph.D. Dissertation*, University of Antwerp (1993).
2. B. Nys, 'The development of image processing and analysis techniques for applications in chemistry and biomedicine', *Ph.D. Dissertation*, University of Antwerp (1991).
3. E. Dougherty, J. McCabe, E. Thurston and J. Laskin, *Imag. Sci. Technol.* 47 (1990).
4. B. Raeymaekers, P. Van Espen and F. Adams, *Microchim. Acta (Wien)*, 1984 II, 437–454.
5. V. Kindratenko, P. Van Espen, B. Treiger and R. Van Grieken, *Microchim. Acta (suppl.)* **13**, 355–361 (1996).
6. J. Russ, *The Image Processing Handbook*, 2nd edn, CRC Press, Boca Raton, FL (1995).
7. A. Medalia, *Powder Technol.* **4**, 117 (1970–1971).
8. J. Beddow, G. Philip and A. Vetter, *Powder Technol.* **18**, 19 (1977).
9. J. Keller, M. Gray and J. Givens Jr., *IEEE Trans. Syst., Man, Cyber.* **SMC-15**, 580 (1985).
10. A. Brink, *Pattern Recogn. Lett.*, **9**, 335 (1989).
11. A. Rosenfeld and A. Kak, *Digital Picture Processing*, Vol. 2, Academic, London (1982).
12. V. Kindratenko, B. Treiger and P. Van Espen, *Microsc., Microanal., Microstruct.* submitted.

