Object classification, segmentation and parameter estimation in multichannel images by classifier learning with clustering of local parameters

Vladimir V. Lukina*, Nikolay N. Ponomarenkoa, Alexander A. Zelenskya,
Karen O. Egiazarianb, Jaakko T. Astolab

aNational Aerospace University, 61070, Kharkov, Ukraine;
bTampere University of Technology, Institute of Signal Processing,
P.O. Box-553, FIN-33101, Tampere, Finland

ABSTRACT

In different applications, it is often desirable to retrieve useful information from multichannel (color, multispectral, dual or full-polarization) images. On one hand, multichannel images are potentially able to provide a lot of useful information about sensed objects (terrains). On the other hand, the task of its reliable extraction is very complicated. And there are many reasons behind this like inherent noise, lack of a priori information about object features, complexity of scenes, etc. Therefore, numerous different approaches based on various functional principles and mathematical background have been already put forward. In majority of them, image classification and segmentation are common operations that precede estimation of object parameters. However, practically all methods are far away from completeness and/or perfection since they suffer from different drawbacks and application restrictions. Recently we have proposed methods based on learning with local parameter clustering that were rather successfully applied to image locally adaptive filtering and detection of objects with certain properties. This paper is an attempt to extend this approach to image classification, segmentation and object parameter estimation. A particular application of substance quantitative analysis from color images is considered. The proposed approach is shown to solve the aforementioned task quite well and to have a rather high potential for other applications.

Keywords: multichannel images, classification, learning, clustering

1. INTRODUCTION

Multichannel mode of image forming is nowadays widely used in numerous applications. One can easily mention spaceborne and airborne multi- and hyperspectral remote sensing for various purposes1-3, radar observation of Earth surface4,6, product inspection7, color imaging8,9, etc. An obvious advantage of multichannel imaging is that due to multi-dimensionality of provided data it becomes possible to retrieve useful information more reliably and to estimate object parameters with better accuracy by means of simultaneous exploiting similarity and differences in component (band) images. At the same time, increasing the number of bands (channels) results in radical increasing the complexity of any kind of procedures of multidimensional data processing like preliminary analysis, registration, calibration, filtering, classification, segmentation, interpreting, object parameter estimation, compression, etc.5,14 This explains the fact that there exist many different strategies and approaches to object classification, segmentation and parameters estimation7,13,14

Image segmentation and classification are, in general, separate and different operations although they can be interrelated. There are different approaches and methods to both segmentation and classification of multichannel data. In particular, it is possible to, first, segment an image and then to employ classification of segmented areas with taking into account averaged statistical characteristics for them as well as, e.g., textural features. The authors7,15 state that there exist four basic groups of segmentation methods: edge-based, neighborhood-based, histogram-based and cluster-based ones. All of them possess their own advantages and drawbacks. For example, for edge-based methods it is often difficult to obtain closed-loop contours especially if images are subject to rather intensive noise6,16. Segmentation of texture is also a

* lukin@xai.kharkov.ua  tel./fax +38 057 3151186
difficult task\textsuperscript{13} and it requires special attention\textsuperscript{17}. Multichannel image classification is a complicated task as well and in spite of considerable efforts and intensive research in this area there is still a lot of work to be done. Bayesian classifier techniques, Independent Component Analysis and Principal Component Analysis based approaches, neural network (NN) and recently appeared support vector machines – this is only a brief enumeration of various means that have been tried for solving a task of multichannel image classification\textsuperscript{14,18-22}. Although these methods are rather different, majority of them in one or another way rely on multi-dimensional clustering and/or classifier learning. This is because in any case one has to somehow separate features in feature space in order to perform data classification.

There are many unsolved problems in multichannel image classification. To name a few, we can mention a used feature set selection, immunity of these features with respect to noise and other distortions, non-gaussianity of feature multi-dimensional distributions and their intersecting for different classes, learning (training) data sample forming, learning acceleration, etc. We have run into similar problems in design of locally adaptive filters\textsuperscript{5,23-25}. Many local parameters (calculated within a pixel neighborhood) have been tried in order to provide this pixel (local) recognition as belonging to image homogeneous region, edge, small sized object, texture, etc. Finally, an approach to locally adaptive filtering based on learning with clustering has been proposed\textsuperscript{26} that allowed to select the most informative local parameters (activity indicators) automatically and to easily combine them using a tree of simple logic operations of local parameters comparison to automatically pre-determined thresholds. In this way it is also possible to determine non-informative and/or noise-sensitive local parameters (features) that are the so-called fictive and are not exploited in pixel (local) recognition. Note that a drawback of such pixel-by-pixel local recognition\textsuperscript{23,24} as well as pixel-by-pixel classification is appearance of misclassified separated pixels. This shortcoming can be partly alleviated by either applying special post-processing misclassification correction procedures\textsuperscript{27,29} or by a multichannel image careful pre-filtering before its classification\textsuperscript{14,22} that is especially efficient in case of rather intensive noise. Also note that many researchers insist on expediency of using both spectral and spatial features in segmentation and classification of multichannel and color data\textsuperscript{7,30,31}.

In automatic design of locally adaptive filters based on learning with clustering\textsuperscript{26} traditional image processing quantitative criterion (MSE, PSNR) was exploited. However, recently it has been demonstrated\textsuperscript{32} that other criteria dealing with the final task of image processing (e.g., probabilities of small sized object correct detection and false alarm rate) can be used in locally adaptive pre-filter design as well. In this paper, we further advance this approach and extend it to segmentation and classification of multichannel (color) images. In Section 2, we discuss image and noise properties one can run in practice in general and that we have encountered in our particular application. Assumptions concerning image and noise characteristics are given and requirements to multichannel image segmentation and classification are discussed. Section 3 describes a proposed approach to multichannel data processing, some aspects of its automation are considered. Experimental results analysis is presented in Section 4. Then, the conclusions follow.

2. IMAGE/NOISE PROPERTIES AND REQUIREMENTS TO IMAGE CLASSIFICATION

There are several obstacles (factors) that can prevent reliable segmentation and classification of multichannel images in practice. Let us briefly consider the basic ones among them. First, usually images one deals in practice are corrupted by noise in less or larger degree. This noise can be of different nature and origin, it might have different statistical and spatial properties. Moreover, in multichannel images noise often is characterized by non-identical basic parameters like variance in different component (band) images, in some cases noise can be even of different type in different channels\textsuperscript{7,6}.

Thus, before starting to solve image segmentation and classification task, it is desirable to establish noise type and to estimate its corresponding basic characteristics. In particular, this can be needed in edge detection, for selecting a proper filter for data pre-processing, in order to forecast how strongly noise can influence image classification, etc. Since we, in general, consider multichannel data and a number of channels can rise up to hundreds, it is expedient to perform a task of noise type identification and parameter estimation (NTIPE) in an automatic (blind) manner. One can argue that for many already operating imaging systems a priori information on noise type and basic parameters can be available. If this is really so and the corresponding parameters are quite stable (do not considerably change from one observation to another), then the task simplifies and the stage of NTIPE can be skipped. However, quite often some noise characteristics like additive noise variance or multiplicative noise variance can in some degree vary from one observation mission to another or they might depend upon imaging conditions. Then, noise parameter estimation for each set of multichannel images can be recommended. Fortunately, nowadays there exist automatic means for NTIPE (see the papers\textsuperscript{10,33,34} and references therein).
In our particular case we dealt with color RGB images of specially pre-processed substance surface (namely, cement). A sample image is presented in Fig. 1.a. Our experiments using the methods 10,33,34 allowed to establish that noise in all three components was of additive nature and it had Gaussian distribution. Fortunately, its variance was practically the same for all RGB components and approximately equal to 8 (for conventional 8-bit representation). The variance did not vary a lot from one to another image from the considered set of images (more than 30 in aggregate). Before continuing, note that a practical task of ours was to estimate percentages of four components in this substance, namely, alite, belite, celite and interstices by automatic analysis of offered images. Such percentages characterize material quality and technological process used in its manufacturing. Certainly, this task can be solved in different ways, but, in fact, we had to classify the aforementioned four types of objects (components), to determine areas (squares) occupied by them and then to calculate the corresponding percentages using the numbers of accordingly classified pixels.

![Fig. 1. An example of original image with four classes of objects (a), enlarged fragment illustrating blurred object edges (b), enlarged fragment illustrating artifacts (yellow color objects inside black objects (c)](image)

Noise can be i.i.d. or spatially correlated. The latter case is more unfavorable since then any filtering less efficiently removes noise. For the considered images, noise is spatially correlated. The width of noise high correlation area is about 3 pixels. Note that following the way of image processing automation, it could be nice to have blind procedures for evaluation of noise spatial correlation. Then it would be possible to perfectly adjust filter type and parameters, e.g., scanning window size. We stress here that within multichannel data classification procedure, image pre-filtering is a desirable operation that commonly improves classification results 5,14,16,22,35.

Blur is one more factor influencing image segmentation and classification. At segmentation stage, image blur reduces object edge detection performance. At classification, blur presence can lead to appearing falsely detected classes in pixels forming contours between different objects 14. In our application, we had some blur that is seen by visual analysis of an enlarged fragment presented in Fig. 1.b. One can try coping with blur by applying some image deconvolution techniques36. But image reconstruction (deconvolution, deblurring) is a rather complex operation especially if blind deconvolution is to be done. Moreover, the use of deconvolution can result in specific changes of image statistics that is undesirable. Because of this, in our case we decided to follow another way of coping with blur. A multidimensional edge sharpening filter 37 that helped a lot in other image classification applications 30,12 has been exploited. The use of such filter has been hindered by one more phenomenon we observed for our color images, namely, color component small misalignment. Fig. 2 shows an example of image pre-processing by the filter 37.
There are also some other factors that prevent reliable image segmentation and classification. They are heterogeneity of objects, artifacts like that one demonstrated in Fig. 1.c, etc. Heterogeneity is clearly demonstrated by histograms represented in Fig. 3 that are taken from training (learning) data samples. Analysis of these histograms shows that clusters corresponding to each class of objects might have few peaks (consist of few sub-clusters). This is an obstacle that might prevent successful application of Bayesian classifiers. The histograms also demonstrate that classes are not easily separable by such criteria as minimal distance to a cluster center. Other simple classification algorithms that can be based on using the rules like “if for a given pixel $L<H<U$, then refer it to the $k$-th class (where $L$ and $U$ are the lower and upper thresholds and $H$ is the component value, respectively)” or their combinations are unable to separate classes. In our case, the use of such rules results to the fact that many pixels are not referred to any of four classes.

Taking into consideration the results of image and noise property analysis, let us formulate requirements to image classification. First, it is desirable to make classifier as insensitive to noise and degradations as possible. This means that since we intend to employ training samples in our classifier learning, then being learnt for some images or their fragments, the classifier has to be successfully applied to images for which it has not been trained. Third, classifier accuracy should be appropriate; this can be verified by comparison to classification results obtained in another way, for example, by a qualified expert. Available a priori information is to be taken into consideration. In our case, we have two types of such information. First, segmented objects should have, at least, ten pixels neighboring each other (characterized by compactness). Second, there is no need in absolutely correct localization of objects. This is because later for determination of component percentages we would anyway count pixels corresponding to four types of objects and small errors in localization would “average” for an entire image of rather large size.
3. PROPOSED APPROACH TO IMAGE CLASSIFICATION

As it follows from analysis in previous Section, in our application we deal with supervised classification in the sense that the number of classes is known in advance (four). For original RGB representation of images, centers of classes can be approximately determined by using typical fragments of certain class objects in images and averaging the obtained values. These class centers for R, G, and B components are given in Table 1 (data are obtained for four images). However, this information is not too valuable because of aforementioned properties of data statistics for the considered classes.

Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Alite</th>
<th>Belite</th>
<th>Celite</th>
<th>Interstices</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>115</td>
<td>117</td>
<td>174</td>
<td>69</td>
</tr>
<tr>
<td>G</td>
<td>141</td>
<td>114</td>
<td>182</td>
<td>78</td>
</tr>
<tr>
<td>B</td>
<td>94</td>
<td>68</td>
<td>103</td>
<td>43</td>
</tr>
</tbody>
</table>

Based on carried out analysis and previous experience, classification technique should take into account not only the pixel values in R, G, and B components, but also other features. In particular, the ratios of component values could be useful. Pre-filtered image values can carry useful spatial information and be less sensitive to noise. Image representation in other, not RGB, color system can be more expedient and worth exploiting. Different local activity indicators like local variance can serve for indicating object edges valuable for segmentation and classification and so on. In other words, while starting to think what features are the most informative it is difficult to decide in advance.

Because of this, on one hand, a classifier should be able to process a rather large number of input features of different origin. On the other hand, a classifier has to “ignore” features that occur to be not informative or too “noisy”. Such features can be considered as fictive or redundant. For several types of classifiers the presence of such fictive features can considerably complicate classifier design and learning. For other types like evidential reasoning and neural network classifiers the presence of fictive input features is not crucial although it might lead to more time required for learning.

![Fig. 4. Block diagram of classifier learning for image fragment classification](image-url)
For our novel classifier based on learning with clustering of a set of features we assumed that per-pixel classification would be performed (Fig 4). At the same time, we kept in mind that for artifact removal and misclassification correction some special post-processing of classification data using special morphological operations could be used (optionally).

According to the proposed approach, learning is iterative and oriented on minimization (maximization) of a given criterion of clustering quality \(^26\). This approach can be used for both efficient classification of image fragments and selection of methods for further processing (analysis) of these fragments, e.g., their filtering with the aim of noise suppression \(^26\).

In the process of classifier tuning (learning) a set of sample (training) fragments manually pointed by a user is exploited; an example of such sample fragments is depicted in Fig. 5.

![Training sample for classification of the component “Alite” (fragments are shown by white contour frames)](image)

A training sample should contain fragments for all types of objects to be classified. These fragments can be marked with small errors, i.e., some fragments can partly include a small part of another type of object. However, commonly decreasing of such errors lead to better final classification.

<table>
<thead>
<tr>
<th>Feature number</th>
<th>Feature description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>The pixel values of components R, G, and B, respectively</td>
</tr>
<tr>
<td>4-7</td>
<td>The distances (in RGB space) from a given pixel vector to the centers of clusters that correspond to Alite, Belite, Celite and Interstices. The distances are calculated as ( L = \left( \frac{(R_y - R_c)^2 + (G_y - G_c)^2 + (B_y - B_c)^2}{3} \right) ), where ( R_y, G_y, B_y ) denote the values of R, G, and B components for the given pixel, ( R_c, G_c, B_c ) are the values of R, G, and B for the corresponding cluster (see Table 1).</td>
</tr>
<tr>
<td>8-10</td>
<td>The values of Y, Cb, Cr for a given image pixel in color space YCbCr</td>
</tr>
<tr>
<td>11-13</td>
<td>The output values in RGB space of 5x5 mean filter in a given image pixel</td>
</tr>
<tr>
<td>14-16</td>
<td>The output values in RGB space for the vector filter (^37) in a given image pixel</td>
</tr>
<tr>
<td>17-19</td>
<td>The output values in RGB space of 11x11 mean filter in a given image pixel</td>
</tr>
</tbody>
</table>

Therefore, feature bank included 19 different features some of which can be considered as color ones and other features in one or another manner take into account spatial information (different filter outputs) (see Table 2) although a feature
set is formed for each image pixel and classification is applied pixel-wise. As the result of classifier learning, three features have been not used at all. In other words, they have been considered as non-informative (fictive). Namely, these are features 4, 7, and 8. The fact that the features 4 and 7 occurred to be fictive one more time stresses the problems of classifiers based on calculation and analysis of distances to the cluster centers. The feature 8 is not used since intensity Y in YCbCr space is not sufficient for separating object classes.

Consider now features that are not fictive and contribute to classification. Table 3 gives the values of “relative importance” of features that have been used by the learnt classifier. By “relative importance” of a feature we mean a percentage of test image pixels for which this feature has been used in classification in the obtained classification tree. As seen, there are, at least, 8 features in the feature bank that are often used and their relative importance is about or over 5%. The most informative are the features 14-16, i.e. the RGB output values of the vector filter that sharpens edges. This confirms the statement that spatial features are very important in classification and outputs of nonlinear filters can be used as them.

Table 3. Relative importance of features

<table>
<thead>
<tr>
<th>Feature</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>4.9 %</td>
<td>2.4 %</td>
<td>2.4 %</td>
<td>-</td>
<td>0.8 %</td>
<td>15.8 %</td>
<td>-</td>
<td>-</td>
<td>2.7 %</td>
<td>2.5 %</td>
</tr>
<tr>
<td>Feature</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Importance</td>
<td>1.0 %</td>
<td>13.2 %</td>
<td>0.6 %</td>
<td>13.5 %</td>
<td>7.4 %</td>
<td>15.5 %</td>
<td>7.3 %</td>
<td>1.6 %</td>
<td>8.4 %</td>
<td></td>
</tr>
</tbody>
</table>

4. EXPERIMENTAL RESULTS ANALYSIS

Classifier learning has been performed using training fragments for which the training feature samples have been calculated from four images. Then the learnt classifier has been used for 25 other images that have not been used in learning. According to our estimates, the RMSE of component percentage determination using classification results is about 0.5%. Minimal percentage of a particular component that has been met in material (substance) samples is about 5% (it corresponds to celite and varies in the limits from 5% to 11%). For other components of the considered material the percentages are larger. For belite the percentage varies within the limits from 8% to 31%, for alite within the range from 22% to 51% and from 22% to 35% for interstices. With taking this into account, we consider that the obtained accuracy (RMSE) of component percentage determination by segmentation and classification is sufficient for the considered application.

This conclusion coincided with the opinion of experts. They have classified and determined the component percentages manually. The largest difference between component percentages determined by experts and evaluated automatically from image segmentation and classification data was not more than 1.7%. Note that component percentage can be not the only information set characterizing a considered material. Spatial shapes and size characteristics of fragments corresponding to different object classes can be important information as well that gives additional useful information to a specialist. In this sense, image segmentation data might offer useful information to an expert. Using segmented data, it is possible to automatically estimate spatial and size characteristics from segmented and classified image.
Fig. 6 demonstrates an example of image segmentation into fragments corresponding to different types of objects. The data have been obtained by the learnt classifier. Colors in Fig. 6b correspond to cluster center colors presented in Table 1. Note that we exploited hard classification. In general, soft classification is also possible using the proposed approach.

5. CONCLUSIONS

In this paper the earlier proposed approach to designing the image processing algorithms based on learning with clustering has been extended to multichannel (color) image classifier design. The obvious advantages of the proposed approach are the following. First, it does not need preliminary analysis of feature information importance and its possible contribution into improving classification. The informative and non-informative (fictive) features are separated automatically at learning stage. Then there is no need to calculate the fictive features at all while performing image segmentation and classification. Second, the proposed classifier is able to easily incorporate spectral (color) and spatial features where the latter ones are taken into account by means of considering linear or nonlinear scanning window filter outputs. For the considered application of substance (cement) quantitative analysis the proposed classification procedure provided practically acceptable final results.

REFERENCES
