STATISTICAL PROCESSING OF METEORIC SNAPS USING MATLAB TOOLBOXES

Eliška Anna Kubičková

University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics

Abstract

Mathematical statistics is a strong tool for processing of digital images. Numerous image analysis and image enhancement techniques are based on probabilistic and statistical methods. This paper deals with the use of statistical functions of MATLAB's Image Processing Toolbox for performing image analysis and image enhancement. Further the search for meteors in astronomical images on the basis of cluster analysis methods using MATLAB's Statistics Toolbox is described.

1 Statistical Processing of Meteoric Images

This paragraph will describe some statistical functions, which can be performed by Image Processing Toolbox, applied to processing of digital meteoric images. Statistical functions enable to perform a range of image processing and image enhancement functions, for example deblurring, colour enhancement, and contrast enhancement. One of the most important statistical functions is displaying of the image histogram. MATLAB's function *imhist* displays an image histogram above a greyscale colorbar. The number of bins is 256. Histograms of various meteoric snaps are in Figure 1. Function *histeq* enhances the image contrast using histogram equalization. It enables to transform the intensity values so that the histogram of the output image approximately matches a specified histogram. By default function histeq tries to match a flat histogram with 64 bins. This function works on the entire image. An example of the use of histeq is shown in Figure 2. Function *adapthisteq* represents contrast-limited adaptive histogram equalization (CLAHE). This function enhances the contrast of the greyscale image by transformation of the brightness values using CLAHE. It operates on small regions in the image, which are called *tiles.* Contrast of each tile is enhanced so that the histogram of the output region approximately matches the histogram specified by the 'Distribution' parameter; see [3] for details. The neighbouring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous area, can be limited to avoid amplifying any noise that might be present in the image. The adapthisteq function is applicable to a greyscale image and also to a colour image. We can see examples of the use of this function in Figure 3.

Other statistical functions from Image Processing Toolbox can be used to explore image matrices features. Function *corr2* performs the computation of the correlation coefficient between two matrices or vectors of the same size by the equation (1),

$$r = \sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B}) / \sqrt{\sum_m \sum_n (A_{mn} - \overline{A})^2} \sqrt{\sum_m \sum_n (B_{mn} - \overline{B})^2} \tag{1}$$

where $\overline{\mathrm{A}}$ and $\overline{\mathrm{B}}$ are the means of elements in the matrices A and B. As corr2 can be used for matrices of the same size only, it can be used for the computation of the correlation coefficient between greyscale and binary image. Function *mean2* computes the mean or average of the matrix elements. MATLAB's function *std2* computes the standard deviation of the matrix elements. Results of the use of mean2 and std2 functions are shown in Table 1. Other functions enable to compute correlation coefficient for vectors and matrices, variance for vectors, maximum and minimum of elements, median, and mode (most frequently value in array). Functions *std* and *mean* can be used also for vectors. Detailed description of statistical functions from Image Processing Toolbox, their syntax and possibilities of the use are brought in [2].

Ing. Bc. Eliška Anna Kubičková, University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics, Univerzitní 22, 30614 Plzeň, email: eliskaak@kky.zcu.cz, elize@volny.cz

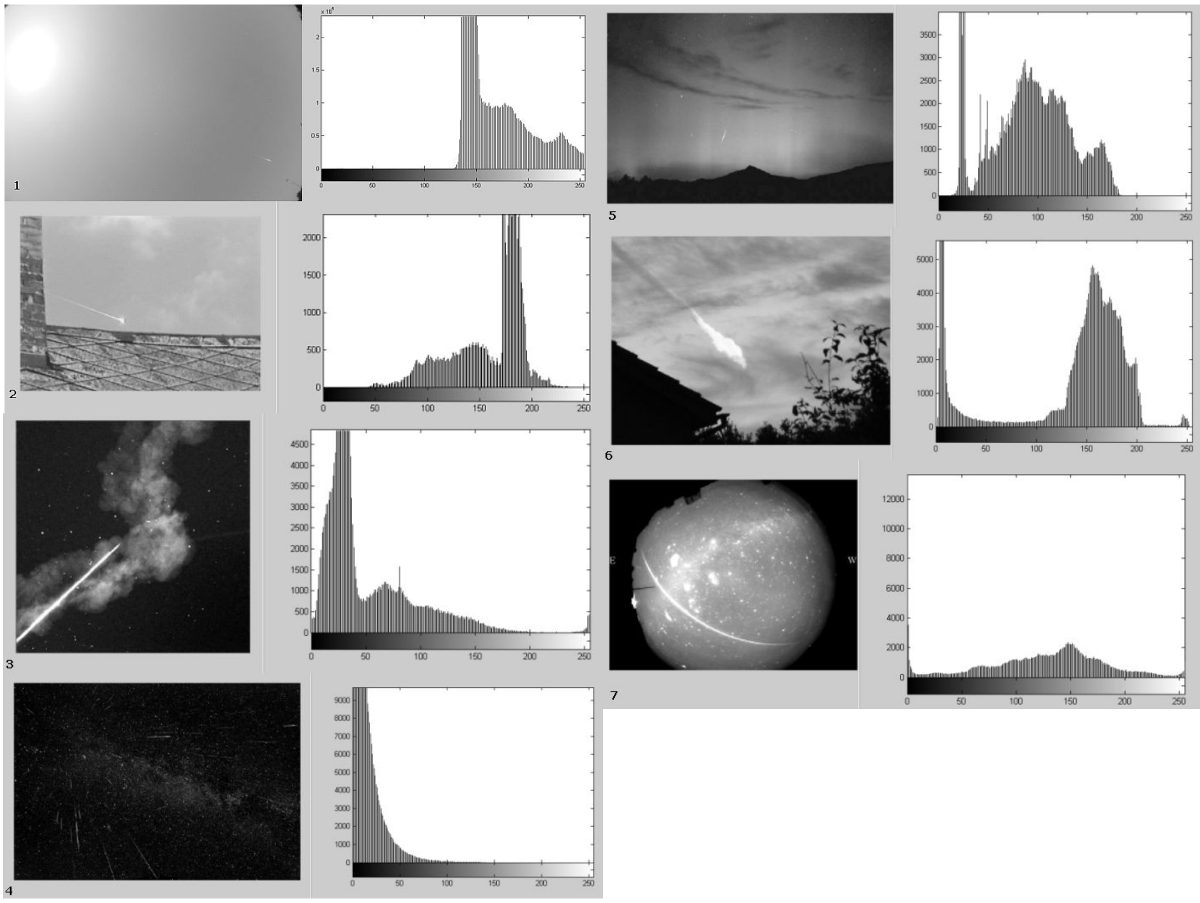


Figure 1: Histograms of various meteoric snaps

Table 1: APPLICATION OF STATISTICAL FUNCTIONS TO THE IMAGES IN FIGURE 1

Figure	mean2	std2	
1	178.6562	out of memory	
2	168.9044	51.0843	
3	52.5223	43.1879	
4	17.6674	20.4814	
5	95.0617	46.2088	
6	123.8498	67.6433	
7	91.6280	73.7833	

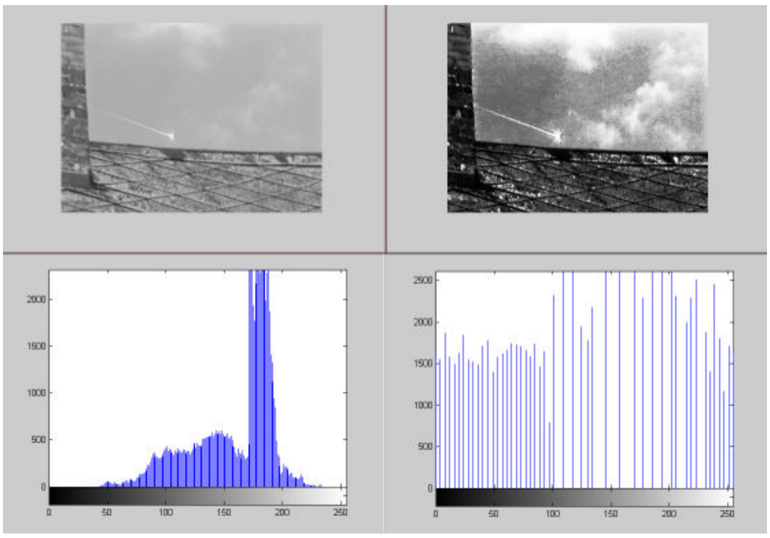


Figure 2: Using of the histeq function; left column: original image and its histogram, right column: equalized image and its histogram

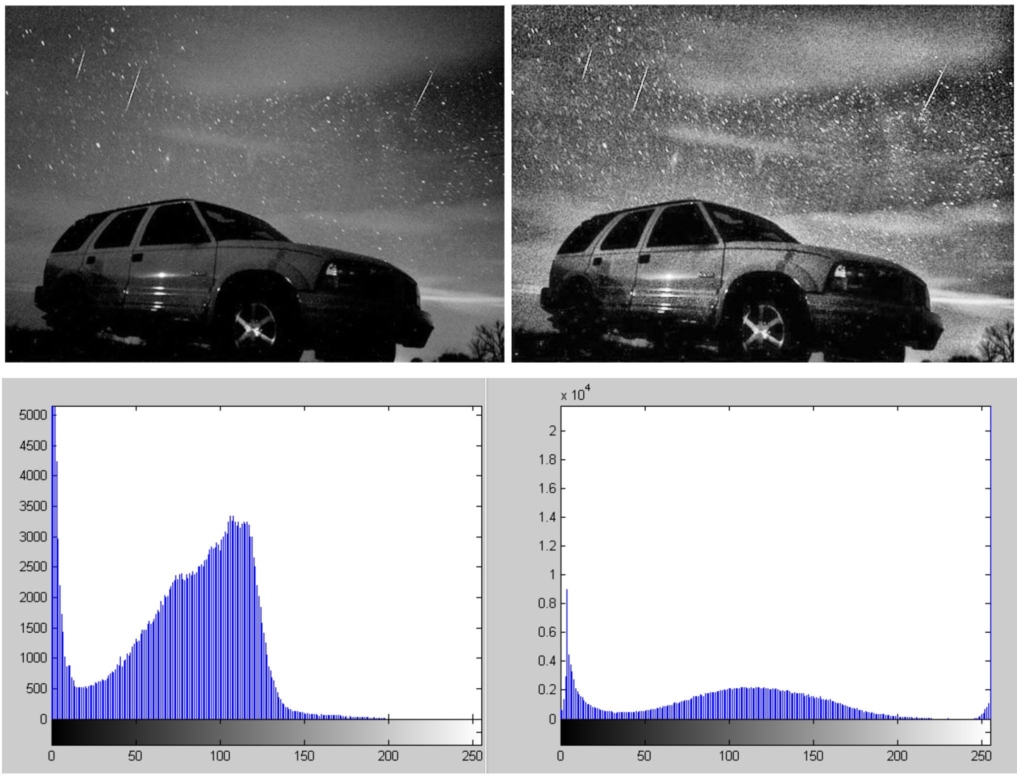


Figure 3: Using of the adapthisteq function; left: the original image and its histogram, right: the image processed by adapthisteq and its histogram

Function *imadjust* from Image Processing Toolbox performs adjusting of image intensity values or colormap. It maps the intensity values in greyscale image to new values such that 1% of data is saturated at low and high intensities of an original image. This process increases the contrast of the output image. This function is equivalent to use imadjust after stretchlim (see [3]). It can be used also for a colour image. The use of imadjust is shown in Figure 4.

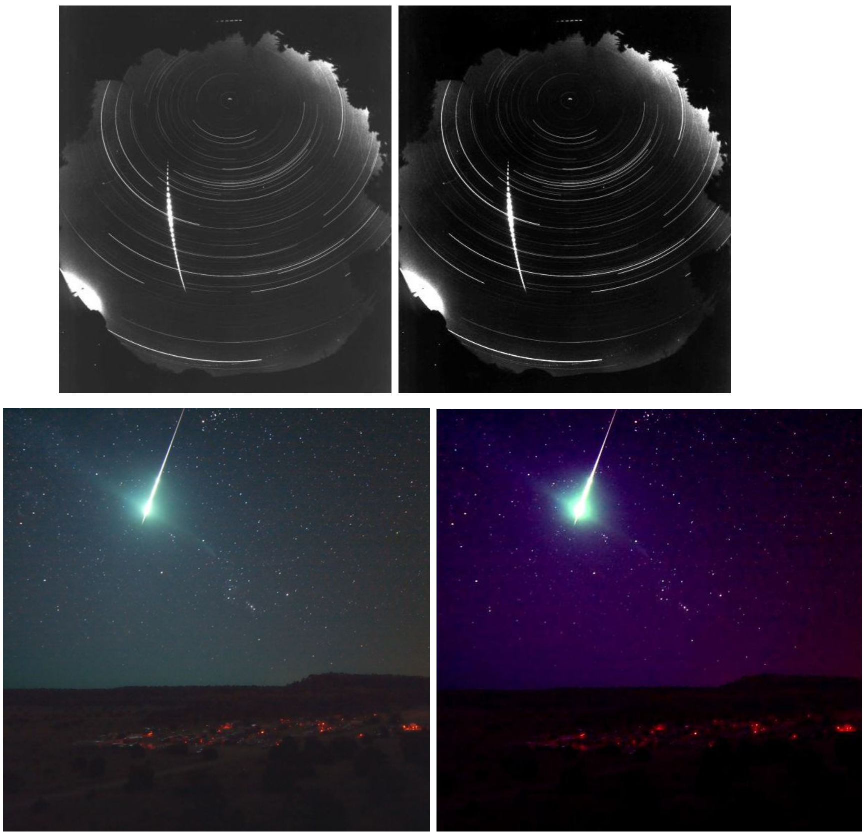


Figure 4: The use of imadjust function; left: an original image, right: a new image

Function *decorrstretch* performs a decorrelation stretch that is applied to a multichannel image. This function enhances the colour separation of an image with significant band-band correlation. The exaggerated colours improve visual interpretation, which enables feature discrimination easier. An example of the use of decorrelation stretch is in Figure 5.

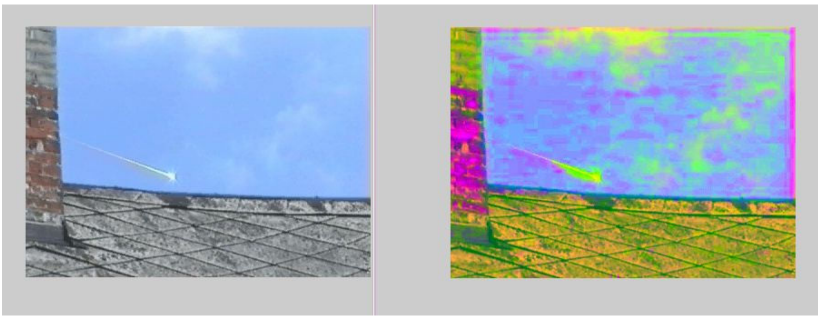


Figure 5: The application of decorrstretch function

Function *stretchlim* finds limits to contrast stretch in an image. This function enables to add a linear contrast stretch after the decorrelation stretch. Adding the linear contrast stretch enhances the resulting image by further expanding the colour range. It is possible to apply a linear contrast stretch as a separate operation after performing a decorrelation stretch, using stretchlim and imadjust. This alternative can give inferior results for uint8 and uint16 images, because the pixel values must be clamped to [0..255] or [0..65535]. The use of stretchlim is shown in Figure 6.

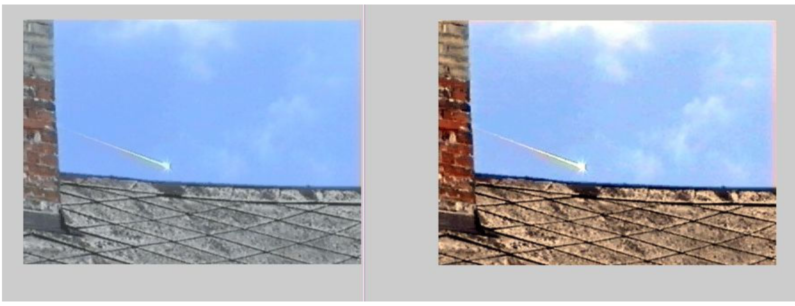


Figure 6: The application of stretchlim function

2 Search for Meteors in Astronomical Snaps Using Statistics Toolbox Functions

2.1 Theoretical Introduction into Cluster Analysis

MATLAB's Statistics Toolbox contains numerous functions, which perform cluster analysis tasks. Cluster analysis (clustering) is a large group of different methods, which try to separate objects from a given dataset into individual groups on the basis of similarity and dissimilarity among these objects. There are numerous definitions of cluster analysis in literature. The first definition of cluster analysis by Robert C. Tryon (1939) is the following: "Clustering is a general logical process defined as a procedure for objective grouping of individuals based on their similarities and dissimilarities." [4] As it is difficult to find an exact definition of cluster analysis as it is not easy to find unified division of clustering methods. The most current division of clustering methods is division into two main groups – hierarchical and non-hierarchical methods. The first group is usually further divided into two groups – divisive and agglomerative methods. Hierarchical methods produce a nested series of partitions. Partitional clustering produces the only one partition. Agglomerative clustering technique begins with each pattern in a distinct (singleton) cluster, and merges clusters together until a stopping criterion is reached. Divisive method begins with all patterns in a single cluster and performs splitting until a stopping criterion is met. Hierarchical clustering algorithms produce a dendrogram representing the nested grouping of patterns and similarity levels, at which grouping change. The dendrogram can be broken at a different level. The most used hierarchical clustering algorithms are the Single linkage and Complete linkage methods. These algorithms use as a clustering criterion distances between elements of individual groups. Single linkage measures the distance between two clusters as the minimum of the distances between all pairs of patterns measured from the two clusters – one pattern is from the first cluster, the other from the second cluster. Complete linkage measures the distance between two clusters as the maximum of the distances between all pairs of patterns measured from the two clusters. The term of distance is very important in cluster analysis, because it often is used as similarity or dissimilarity criterion. More information about types of distance measurements is found, for example, in [6]. The typical representative of partitional clustering methods is *k*-means algorithm. This method choices *k* clusters centres to coincide with *k* randomly chosen patterns or *k* randomly defined points inside the hyper volume containing the pattern set. Each pattern is assigned to the closest cluster centre. Then the cluster centres are recomputed using the current cluster membership. Previous two steps are repeated until a convergence criterion is met.

Mathematical description of clustering is based on the definition of cluster.

Definition: Let X = $\{x_1, x_2, ..., x_n\}$ is a set of items. Let *D* is a dissimilarity coefficient. Then cluster is a subset *A* of the set *D*, which complies the following inequality:

$$\max D(x_i, x_j) < \min D(x_k, x_l), \text{ where } x_i, x_j, x_l \in A, x_k \notin A. \tag{2}$$

Clustering is the process, which divides the given set X into clusters.

More details about clustering theory are in work [1].

2.2 The Use of Cluster Analysis in Research of Meteoric Images

Digital images are represented by image matrix, which contains values of brightnesses in each image point named pixel. Clustering methods are used in image processing for example for image restoration, image retrieval, segmentation, and other image operations (see [1]). Grouping in image data can be based on colour, brightness, or coordinates of the image points. Here two clustering methods mentioned in the previous paragraph – single linkage and *k*-means algorithms are used. As a similarity criterion are used brightnesses of a greyscale image. The MATLAB's Statistics Toolbox contains functions *clusterdata* and *kmeans* to perform these clustering methods. Meteors represent areas with the highest values of brightness in the image. For that reason we can suppose that image areas containing meteor (or more meteors) could represent areas with very bright image points. These points could be grouped by clustering.

Before clustering itself some aided functions for estimation of cluster number and for computation some image statistics were used. These functions are named *dendrogram, cophenet,* and *silhouette.* Theoretical background of these functions is found in [5]. The dendrogram function displays a tree diagram representing the nested structure of the partitions. Numerical values associated with each stage usually represent the distances between two clusters. The dendrogram is shown in Figure 7. The function cophenet [5] computes the cophenetic correlation coefficient that compares accuracy of two partitions. Silhouette statistics enables to estimate the number of groups in a dataset. Observations with a large silhouette are well clustered. A small silhouette width indicates that values are scattered among clusters. Silhouette statistics we can see in Figure 8. The number of groups in hierarchical clustering can be determined by *upper tail rule.* This formula is described in [5] and its application to meteoric snaps is shown in [1]. Functions clusterdata and kmeans were applied to various types of meteoric snaps and it was tested their ability to find meteors. The function kmeans was used with two centroids, function clusterdata was used with five clusters.

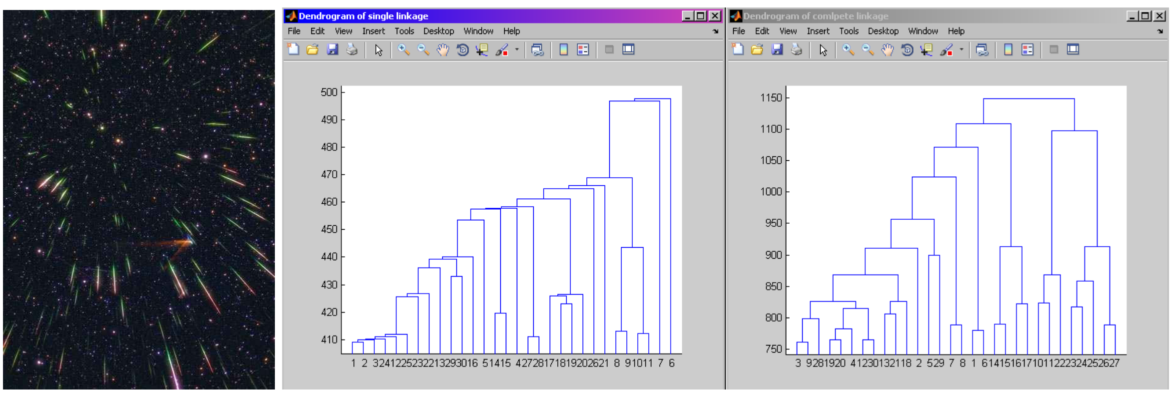


Figure 7: The dendrogram for Leonid meteor shower; left: single linkage, right: complete linkage

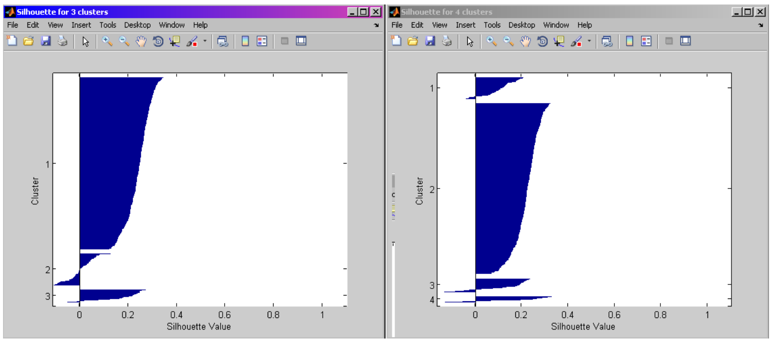


Figure 8: Silhouette plots for 3 and 4 clusters

Table 2: APPLICATION OF K-MEANS CLUSTERING TO THE ASTRONOMICAL IMAGES

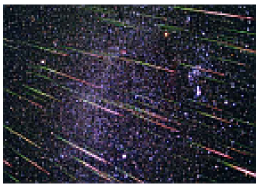
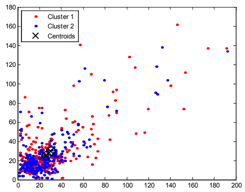
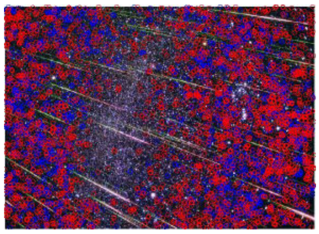
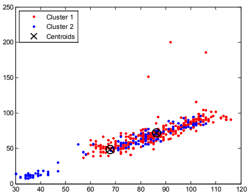
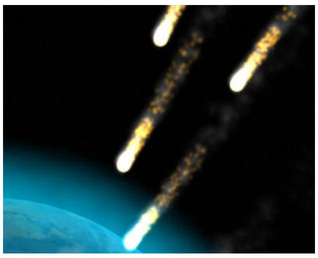
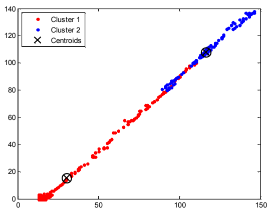
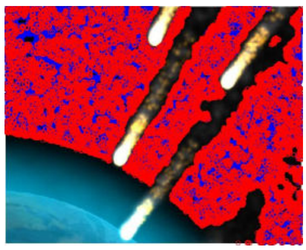
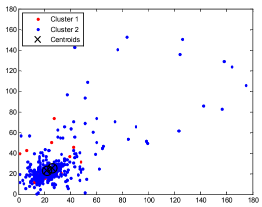
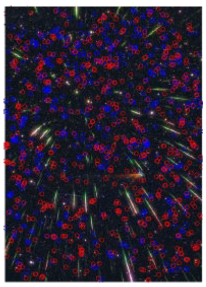
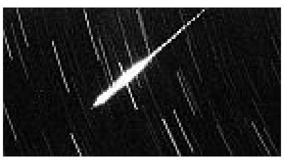
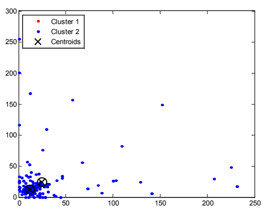
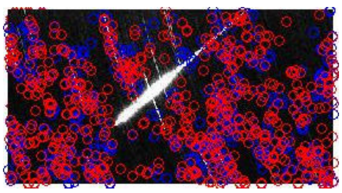
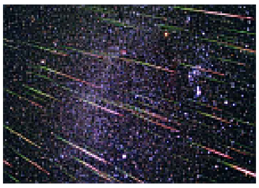
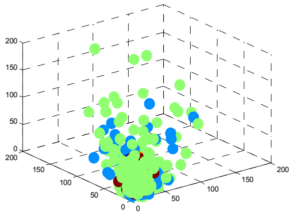
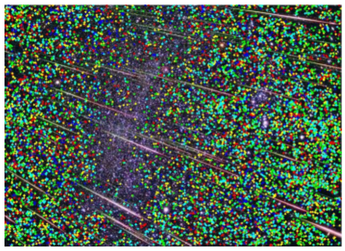
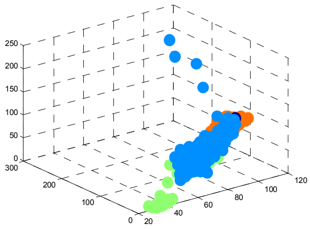
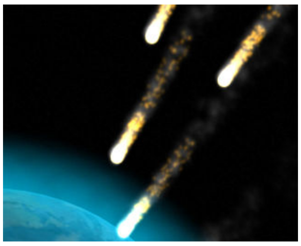
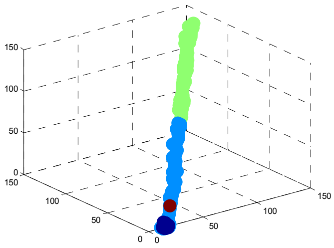
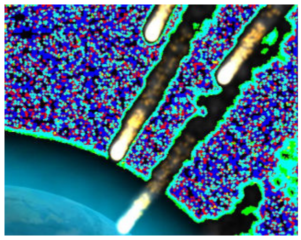
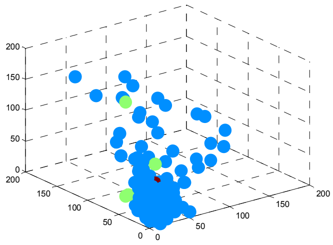
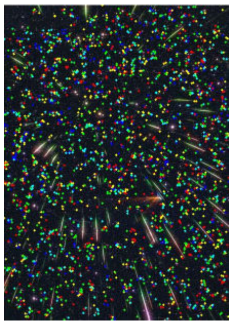
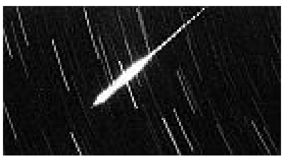
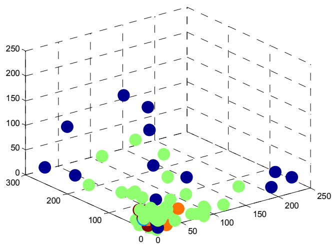
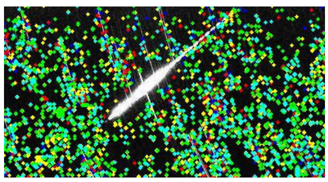
 1	 
 2	 
 3	 
 4	 
 5	 

Table 3: APPLICATION OF SINGLE LINKAGE CLUSTERING TO THE ASTRONOMICAL IMAGES

 1	 
 2	 
 3	 
 4	 
 5	 

3 Conclusions

The first part of this paper brought an overview of probabilistic and statistical based functions, which are included in MATLAB's image processor named Image Processing Toolbox. These functions enable to perform statistical tasks as displaying image histogram and computing mean or maximal element of matrix, standard deviation, or correlation of two matrices. Other functions enabling to perform image enhancement operations were described as well. Examples of equalization of greyscale image, enhancement of separate colours in multichannel image, finding limits to contrast stretch in an image, and adjusting of image intensity using MATLAB's functions applied to astronomical snaps are included. This section can serve to design other image analysis functions. The second part of this article contains theoretical introduction into cluster analysis methods and their use for research of astronomical snaps. The hierarchical single linkage method and partitioning *k*-means method were used for searching of meteors in meteoric images. MATLAB's Statistics Toolbox was used for application of these methods. Description of some aided functions as displaying dendrogram, computation of correlation coefficient to compare partition accuracy, and estimation of number of clusters, and results of their application to meteoric snaps are included. Results of the use of clustering methods themselves are synoptically shown in Table 2 and 3. We can see that both methods prove good image segmentation features (items 2 and 3 in Tables). Clustering appears as a suitable method for image segmentation and search for large objects containing image pixels with the same brightnesses or colours. The items 1 and 4 show that clustering is not a suitable technique for detection of scattered objects with faint brightness as meteors in meteoric streams. Finally we can summarize that cluster analysis could find its use in research of astronomical snaps as a method for searching of very bright individual meteors (bolides).

Acknowledgement

This work has been supported by the grant of The University of West Bohemia: "Intelligent methods of machine vision and understanding", Project No. SGS- 2010-054.

References

[1] E.A. Kubičková. *Cluster Analysis in Meteor Detection.* University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics, Pilsen, 2011.
[2] E.A. Kubičková. *Probability and Statistics in Computer Vision.* University of West Bohemia, Faculty of Applied Sciences, Department of Mathematics, Pilsen, 2011.
[3] *Image Processing Toolbox, User's Guide.* The MathWorks, Inc., 2008.
[4] A. Lukasová, J. Šarmanová. *Metody shlukové analýzy.* SNTL, Praha, 1985
[5] W.L. Martinez, A.R. Martinez. *Exploratory Data Analysis with MATLAB.* Chapman & Hall/CRC Press UK, London, 2005.
[6] K. Spohnerová. *Shluková analýza.* Diplomová práce. Univerzita Palackého v Olomouci, 2010.