HOUGH TRANSFORM APPROACH TO IDENTIFICATION OF FLARE STARS IN MULTI-EXPOSURE PLATE IMAGES*

Dimo Dimov, Katya Tsvetkova, Milcho Tsvetkov, Alexander Kolev,
Ognyan Kounchev

ABSTRACT. The paper discusses the progress in the automation of the detection of flare stars through astronomical observations with plates obtained by the multi-exposure "chain" method. In spite of the fact that the astronomical plate observations are considered obsolete in modern astronomy, the computer processing of these images is an actual problem, due to the considerable number of wide-field plates in our country, in Europe and throughout the world. The solution of the problem is implemented in a sequence of three stages: (i) detection of the form of the chains, (ii) location of the chains in the input plate image and (iii) detection of chains representing "flare stars". The paper is focused on the second stage, the implementation of which requires the use of Hough transform (HT). A theoretical and experimental analysis of the suggested approach connected with HT is presented. The "bottlenecks" in the development are discussed on the basis of a comparative analysis with a competitive approach—the "cumulative" statistic approach to chains in the given image. At the end, a combined approach is offered, uniting the various advantages, but again based on HT. Some experimental results and illustrations are presented. The perspectives for future work, connected with an implementation of a whole software system designed to solve the problem, are described.

1. Introduction. The problem addressed is segmentation of star chains obtained by a method specific to astronomical plate observations—multi-exposure of chain images obtained on one plate of the observed sky region. The exposures (a number of $4 \div 6$ is usually chosen) are with equal time duration (e.g., about $\sim 5 \div 10$ min each), and the relative (mechanic) shifts of the plate after every exposure are also equal and in a fixed selected direction. The only exception is the first shift, which is bigger than the remaining ones, thus indicating the chain's beginning [1, 2].

No direct astronomical software for segmentation of the plate chain stellar images is available for use in the project Astroinformatics at the Bulgarian Academy of Sciences [3]. The astronomical plate observations are already obsolete but the existence of more than 2 400 000 wide-field plates world-wide, stored in various observatories as scientific heritage in time domain astronomy, requires the respective software approach. Generally said, the purpose is the automatized (or even automatic) segmentation/localization of the chains for all (or for as many as possible) star objects in a given plate image, after which (rough or more precise) comparative analysis can be done on the chain set for detection of non-ordinary chains or chains of different shapes including so-called *"flare stars"* [1, 2].

The use of Hough Transform (HT) is proposed for detection of the slope α of the axes of "chains" in the image, by analogy with the approach to detection of the slope of text lines in text images [4]. This natural idea - to automatically segment the chains by a Hough/Radon transform in order to determine the direction and length of the chains – has not been exploited, at least not before 2002 [5].

However, the use of HT has many "hidden obstacles". This is why we discuss in parallel a "cumulative" approach to the localization of chains in an image. The paper presents the comparative analysis, both theoretic and experimental, of the two approaches. Finally, an effective combined approach is proposed for problem solving.

2. A brief description of the astronomical method of stellar chains. The *method of stellar chains* is a photographic multi-exposure method applied with wide-field telescopes (Schmidt or Maksutov type), where after a single exposure, usually of 5 to 10 minutes, the telescope is moved along the Right Ascension coordinate and again a new exposure is made with the same duration—up to 6 such single exposures on one plate, so the multiple exposure plates mostly have 6 exposures with a duration of a single exposure of 10 minutes or less. If the star shows brightness variations with time resolution less than 10

minutes, the images in the stellar chains will not be equal. So this method is not only good for discovery of brightness variability in the stars, comparing the plates obtained in the same region in different days and even in different years, but is also especially suitable for the discovery of quick brightness variability.

Such quick brightness variability exists in the flare stars—a very quick increase of brightness and a decline from maximum to minimum for some minutes up to some hours. The flare stars belong to the population of young stellar clusters and associations; in fact they are their significant fraction. This is due to the fact that flare activity is a common characteristic in the early evolution of all red dwarf stars.

Flare star searches in stellar clusters and associations were done within the frames of the international co-operative programme of several astronomical observatories—Tonantzintla, Byurakan, Asiago, Abastumani, Konkoly, Rozhen, etc. We had access to such material from the monitoring flare star observations in Rozhen Observatory of the Bulgarian Academy of Sciences, done with the 50/70/172 cm Schmidt telescope in the period 1979 ÷ 1995.

When the plates are compared with a blink-comparator, a selection effect appears due to the 10-minute duration of the single exposure, which is a low time resolution. Thus it may preclude the detection of energetic, but low-amplitude flare events on bright stars. So the flare star surveys with the help of the flare star monitoring photographic method of multi-exposure images, each of 10-minute duration, usually discover the most chromospherically active stars. The less active stars could be missed. If the star is very faint, it will be impossible to observe small flares.

An automated flare search method was applied for some southern stellar aggregates of different age [1, 2]. This method removed partially the selection effect and increased the number of discovered flares (mainly slow flare events, flares on bright stars and on In-type variables) by 50%, compared to visual inspection by a blink-comparator.

3. The main problem to solve. Our aim in future will be the automation of the method [2] for new flare star detection, as well as for new flare-ups with the help of chain images and probable improvement of the results achieved by this method.

The purpose of the presented paper is to describe the theoretical bases and the experimental development of a computer system for solving the problem in the following sequence of three stages:

- Detection of the chains' common shape.

- Localization of all chains in the plate image.
- Detection of chains representing "*flare-ups or new flare stars*".

In order to illustrate these problems, Fig. 1 shows a plate image with stellar chains, taken in the region of the Pleiades stellar cluster. Fig. 1 (a, b, c) show 3 typical configurations of chains.

In most cases we might consider that the *first problem* is *a priori* solved, i.e., that the chain form is known *a priori*: a fixed number (up to 6) of stellar images, located along a straight line, the first image taken is shifted at a distance about twice longer than the others, which are as a rule equally removed from one another. Most often it turns out that this approximation of the description of the form is insufficient for the correct solution of the next two problems. Besides, as far as we work with archival materials, it is appropriate for the system to contain a module for solving the problem of the chain form, at least for verification of the description given *a priori.*

Many approaches exist in astronomy to solving the *third problem*, i.e., when the chains generated by the astronomical objects considered are known and are present in the star catalogues used. Moreover, the astrometry approach of chains has appeared mainly with the idea to trace the dynamics of astronomical objects whose location is *a priori* known [1]. Nowadays an additional idea arises: to use this archival material (wide-field chain plates) for discovery of other potentially interesting astronomical objects, for example unknown flare stars [2, 3].

Thus in this project we consider the *second problem* as a basic one. We propose two approaches for its solution, an approach realized by HT and a "cumulative" approach that we will both discuss in a comparative analysis.

4. Characteristic of the plate images with chains. The required file format, used in Astroinformatics to store and transfer astronomical images, is FITS (Flexible Image Transport System) [6]. The FITS file format enables storing the necessary image details of the original photographic plate as well as other astronomical data.

The main parameter of scanning and acquiring as output file in FITS format (from our point of view) is the resolution of the plate scanning, which can be up to 2400 dpi, and most often is 1600 dpi, ensuring 10 μm, respectively 16 μm per pixel, for a conventional plate of square size of $16{\times}16$ cm^2. In this way, volumes of plate images in FITS format of the order of 0.5 GB are not at all rare.[1] On the other hand, conventional personal computers, which are the main
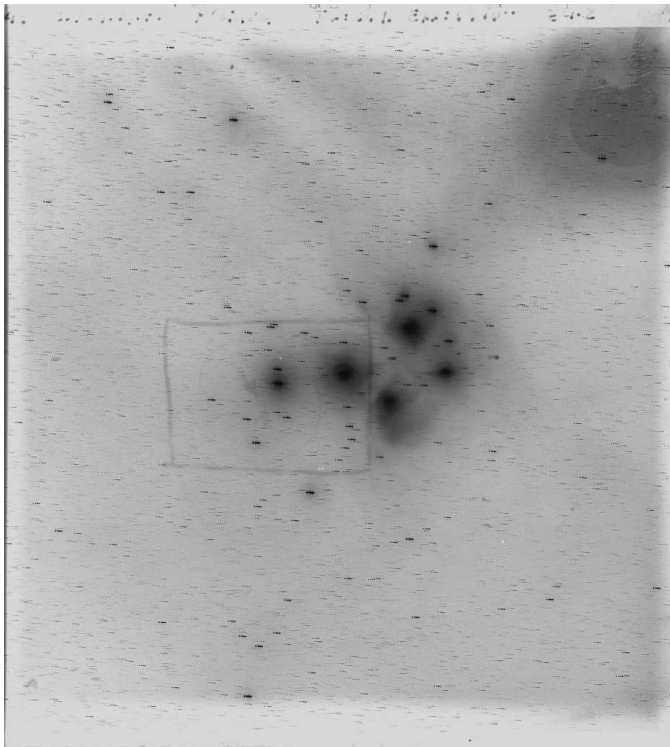
---

[1] http://vo.aip.de/plates/pot-cdc_description.html

Fig. 1(a). A routine chain
of a given star

Fig. 1(b). Chains of
bright stars

Fig. 1. Original image of "chains" type – a wide field plate
of size $16 * 16$ cm^2, covering $4 * 4$ square degrees in the
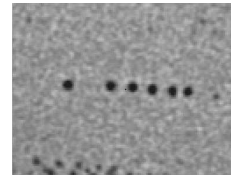area of the Pleiades stellar cluster
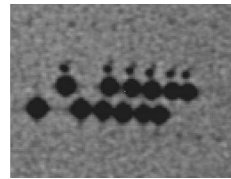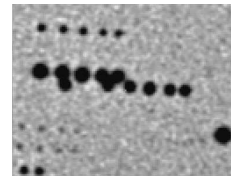
Fig. 1(c). Overlapping
chains

computing power in this project, are still inappropriate for processing similar data volumes. This is why at this stage we shall work with the so called "preview" plate images, in the classical BMP-Gray format, whose volume is smaller by an order of magnitude, i.e., ≤ 50 MB. In the progress of the discussion hereinafter it will be demonstrated that this (or an even larger) reduction in accuracy is entirely acceptable for adequate localization of chains.

The chain images we work with in the frames of Astroinformatics project [3] are stored in the Wide-Field Plate Data Base (WFPDB)[2]. Many of these images are characterized by a relatively high level of noise.

The photographic images of the traditional astronomic objects are most often not pinpoint but rather circles with a relative surface, even for the faint stars. The searched phenomena in the case of "chains" are characterized with

---

[2] http://www.skyarchive.org

possible brightness variability in time: flare stars, pulsating stars, double stars, or are astronomical objects with variable space location (comets, asteroids), etc., [1], caused non-even stellar images or subsequent images with (eventually) different slope in the respective chain.

It is usually accepted that the radial stellar profiles can be modeled by Gaussians (Normal distribution profiles), taking into account the instrumental and atmospheric influence upon it [1, 2, 5, 7]. The size of the stellar photographic image depends on the stellar magnitude, the characteristics of the used telescope and optics giving different aberrations, the Earth's atmosphere, caused diffraction, and the reaction of the emulsion used. Also, various artifacts can be found on the plate images: signs and/or notes most often written close to the plate frame, emulsion damages caused by the passage of time, etc. In other words, it is almost always necessary to reduce the noise of the astronomical images, i.e., to outline the star objects (stars, galaxies, etc.) with respect to the unnecessary background there.

The most frequent approach used for preliminary suppression of image noise is the so-called *binarization* of the images. Details about the binarization procedure, necessary for the chain images considered, have been given in [8]. Only some of them will be pointed here, referring mainly to the following aspect:

5. Classical approaches to localization of stellar objects in plate images. In the practice of astronomy and astrometry the SExtractor program [7] is quite popular as part of a system of software tools[3]. SExtractor saves in a text file the star coordinates (as well as the star value and other parameters) of the discovered stellar objects in the input astro image, which are then verified by the information in a star catalog.

However, SExtractor cannot be directly used to detect flare objects into chain plate images. Having in mind the approach herein suggested, SExtractor could be used only for preliminary processing of the images, instead of binarization, accounting the risk of all the resulting inconveniences, despite the high precision of the result [8].

SExtractor provides the coordinates of the centers of the stellar objects detected in the plate image, precisely registered (attached) to a given star catalog. But in order to detect chains, such attachment is not necessary. For example, the (ρ, θ)-HT procedure, which is proposed, is to a great extent invariant with respect to the coordinate system used. Indeed, for dimensions of the plate less than a few square angular degrees, see Fig. 1, the coordinate differences between

---

[3]http://www.astromatic.net/software/sextractor

the sky coordinate system (of a given catalog) and the coordinate system of the image may be reduced to translation, scaling and rotation. And these linear geometric transforms do not influence the configurations of the searched sets of objects – chains of stars, located along parallel lines of fixed but unknown slope.

A possible non-linearity of "barrel-like type" of the optics distortion of the telescope used can be ignored for the moment. Some diffraction disturbances of the optics of the telescope, mostly expressed in brighter star objects, will be expressly isolated as artifacts at the stage of preliminary binarization.

SExtractor is available on the shell, but it requires specific setups of the control [7]. Certainly, a bitmap image of the underlined objects (discovered by SExtractor) can be generated also by the text output of SExtractor with the help of appropriate external software, after which the ideas about HT offered above will be applied. Similar external software can be written to implement the desired HT directly by the given output, for example by the output text file of SExtractor. But these actual possibilities have the following shortcomings:

• The known implementations of SExtractor work in Linux environment only, which hampers its use in another OS, such as Windows;

• In all cases, if the implementation will be indirect, i.e., composed of two successive and independent software modules, their setups must be compatible. This of course is not bad, if Linux is used to further develop the toolbox of Astromatic.net, but our experience with HT is for the moment concentrated in a Windows environment [4, 8];

Thus, when setting the natural requirement for processing automation, the development of a specific algorithm (and software) is necessary for the complete solution of the problem of chain localization.

Fig. 1 illustrates a preview of a plate image with "chains" taken in the region around the stellar cluster Pleiades. The specific image noise is well visible in Fig. 1 (a, b, c), where three typical types of chains are shown: a relatively well localizable (normal) chain (Fig. 1(a)), a chain of a particularly bright star, where the so-called "halo" effects are seen (Fig. 1(b)), and one common case of overlapping chains, caused by the closeness of the corresponding stars and/or by inappropriately chosen shift direction of the photo plate (Fig. 1(c)). In fact, the surrounding "granular" noise, which appears in the figures as well, can be regarded as a structure of overlapping chains of very close stars of low brightness.

6. Formulation of the problem of chain localization. The following parameters of the chains are of importance here:

• Length N of the chains, i.e., number N of the components shaping the chain of a given star object. It is obvious that due to the specifics of the

astronomic method considered, the length of all the chains is the same and is equal to the number of exposures of a given photo plate. I.e., N may be considered a known constant, a natural number, usually $N = 4 \div 6$. Regardless of that, it is good if the approach of chain localization gives an estimate of N for verification.

   • The coordinates (x_i, y_i) of the centers C_i of the corresponding components O_i, $i = 1, 2, \ldots, N$ (i.e., exposures O_i) of a given star object O, forming the chain O for this object.

   • Average brightness V_i of the components O_i of a given chain O:

$$(1) \qquad V_i = \mathbf{E}(P_i(x, y), \ (x, y) \in O_i), \quad i = 1, \ldots, N$$

where $P_i(x, y)$ is the brightness density of the component in pixels (x, y), $(x, y) \in O_i$. (Further in the text, will be used to denote the operator of average value over the respective definition area, in the above case — O_i).

   • The averaged shifts (i.e., translations) Δ_i between the successive pairs of centers (C_i, C_{i-1}), $i = 2, \ldots, N$, along all the chains in the image (or in its representative part):

$$(2) \qquad \Delta_i = \mathbf{E}(|C_i^{(j)} - C_{i-1}^{(j)}|, \ j = 1, \ldots S), \quad i = 2, \ldots, N,$$

where S is the number of all chains in the image.

   The searched *flare stars* will be identified by their chains at:

   • Considerable difference ΔV_i with respect to brightness (above a given feasible limit ε_V, $\varepsilon_V > 0$):

$$(3) \qquad |\Delta V_i| > \varepsilon_P, \quad \Delta V_i = V_i - V_{i-1}, \quad i = 2, \ldots, N,$$

from one component to another, where V_i are defined by (1);
and also at:

   • Considerable difference ΔC_i with respect to translation (above a given feasible limit ε_C, $\varepsilon_C > 0$):

$$(4) \qquad |\Delta C_i - \Delta_i| > \varepsilon_C, \quad \Delta C_i = |C_i - C_{i-1}|, \quad i = 2, \ldots, N,$$

of the centers C_i from one component to another, with respect to the corresponding average shifts Δ_i, defined by (2).

   Considerable deviations in the shifts ΔC_i are typical rather for mobile objects, the detection of which is not considered a purpose in the astro images considered. In the *flare stars* that are looked for, relatively small deviations in ΔC_i might appear in a few tolerances around the common direction α of the shifts.

The discovery of flare stars in chain plate images is considered an astronomical rarity, i.e., for the greater part of the chains it is expected that they enter the given, somewhat rough, feasible limits ε_C and ε_V. Thus at an already given resolution of the input astro image one expects to obtain a mainly equal average brightness of the components in the chain, as well as uniform translations between the components, i.e., close to the preliminary known (or averaged) ones.

That is why, having in mind probable omissions by criteria (3) and (4), we define a more precise criterion for the detection of flare stars, namely, at an insignificant (under given feasible limits ε_C and ε_V) difference in the translation ΔC_i and also in the average brightness ΔV_i from one component to another, but at a significant:

• Average difference ΔP_i in brightness density (above a given feasible limit ε_P, $\varepsilon_P > 0$):

$$(5) \qquad \Delta P_i = \mathbf{E}(|P_i(x,y) - P_{i-1}((x,y) + \Delta C_i)|, \ (x,y) \in O_i \cap O_{i-1}),$$

from one component to another, $i = 2, \ldots, N$; where $P_i(x,y)$ is the density of brightness (intensity) of the component O_i, $(x,y) \in O_i$. For definiteness, the area O_i of the corresponding component (i.e., the mask, defined by a binarized image) is formally expanded by zeroes within the interval $(-\infty, +\infty)^2$.

7. Comment on the parameters defined. Of course, in criteria (3), (4) and (5), the differences ΔV, ΔC and ΔP, which are $N-1$ in number, may be defined instead on a chain of components, on all the pairs of these components, which are $N(N-1)/2$ in number, and then the difference taken between the maximum and the minimum. However, this refinement is similar for the two approaches, i.e., it does not contribute to the comparative analysis between them, hence we are skipping it.

On the other hand, it is desirable for the abovementioned feasible limits (threshold values) ε_C, ε_V and ε_P to be currently evaluated by the approach selected. We think that this estimate is relatively easy to implement, for example, searching for the "optimal (maximal) jump" in the statistics for the differences ΔV, ΔC and ΔP on all the chains discovered. That is to say, this estimate will be done after the localization of the chains, equally for both approaches, hence we will not consider it herein either.

The defined parameters of the chains correspond to a great extent to the ones routinely used in astronomy, but all the acceptable simplifications contribute to the solution of the main problem here – automatic segmenting (localization) of the chains. For example: the estimation of the centers and areas of the stars

(chains components) is with an accuracy up to a pixel; the brightness density in these areas is with an accuracy up to an intensity quantum (2^{16} quanta per pixel), divided by the respective area, etc. This enables the use of images with reduced resolution instead of the respective complete FITS files, which increases the share of available computer resources directly for the problem considered. The corresponding complete FITS files may be regarded at the final stages, for precise computations on the already discovered (localized) chains, as well as for the discovery of new "irregular" chains, visible only at higher and/or maximal resolution.

At this stage, when the computer resources are limited with respect to operating memory, the problem with the most detailed (i.e., complete FITS) image could also be solved by parts, with all the troubles of border synchronization and without any significant advantages in chain localization. Thus, and without loss of generality, instead of the complete FITS file of the astro image, we shall use its preview in a BMP format, of dimensions $< (10^4 * 10^4)$ pixels, with 1B for intensity per pixel, i.e., with a volume $< 50 \div 100$ MB.

Fig. 2(a) illustrates a chain of $N = 6$ components, five (successive) shifts ΔC_i, $i = 2, \ldots, 6$, are shown, while Fig. 2(b) shows the alteration in the brightness density of the components, as a section along $L \equiv \{l \in (C_i \rightarrow C_{i-1}),\ i = 1, \ldots, 6\}$.

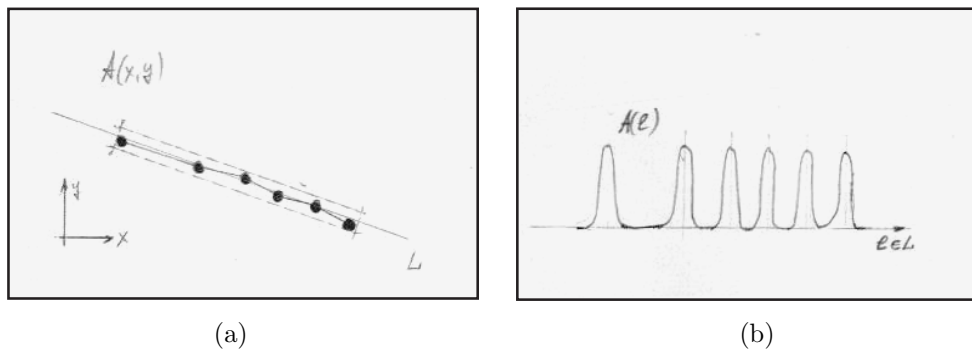(a)   (b)

Fig. 2. (a) A scheme of a chain of $N = 6$ components (exposures) with illustration of: the minimally enveloping and tilted rectangle for the chain, the average direction L and the line string from one component to another; (b) The alteration in component brightness along the section with the line string

Since the cases of *flare stars* are rare, we may accept that the ideal astro image is comprised mainly of patterns of the type of Fig. 2(a), their number S corresponding to the stars in the sky region observed, and the total number of stars (components of chains) in the image is K, $K = NS$. Obviously here

some peripheral phenomena are ignored—near by the image frame, where a set of "false" flare stars could be expected, caused by cutting off by the frame.

8. Approaches to solving the problem of chain localization.

At this stage of the development we define two main approaches for solving the problem of discovery of chains:

(A) a projective approach by HT for the detection of "linearly elongated" chains and

(B) a cumulative approach with *a priori* evaluation of the chains shape.

The two approaches differ considerably, both in the part of interpretation of the shape (form) of the chains and in the part of their localization.

In the first approach it is assumed that the chains are *"linearly elongated"* (Fig. 2(a)) rather than scattered on the plane (Fig. 4(a)). In this way HT can be used to find the average slope α of the chains' axes in the image, by analogy with the approach to determining the slope of text lines in images of text and/or handwriting [4]. In practice, the so-called (ρ, θ)-HT, which is similar to the right Radon transform (known in computer tomography), presents all the necessary projections of the input image in accumulating space, where the chains with collinear axes are transformed into well expressed maxima, easy to discover. Thus separate rows of chains, i.e., *HT-rows*, may be isolated to assist at the next stage the isolation of the chains along all HT-rows found. Meanwhile, the centers C_i and the average brightnesses V_i can be evaluated, as well as the respective areas of the components O_i, $i = 1, 2, \ldots, N$, of every chain O found, which enables direct (and preliminary) selection of the potential collisions (i.e., *flare stars* that are to be found).

However, the classical efficient implementations of H require the input image to be *a priori* (and appropriately) binarized [8]. In practice the binarized image can be regarded as a binary mask, its "white" areas indicating "stars" (i.e., chains' components), whereas the "black" area, which usually (and preferably) predominates, indicates lack of interesting objects and/or suppressed noise in the original astro image. Traditionally in image processing a "black" area is encoded by "0" and a "white" one by the code "$\neq 0$". In this way the presence of predominating "black" (zero) area improves the speed of the HT used.

With this convention, the binarized image is accepted as a "positive" image; while in astronomy mainly the original "negatives" are used. That is why we shall consider that the original astro image at the input is inverted here for the purpose of the processing, and the result obtained is inverted back at the output. These two inverting operations have negligible complexity related to the basic computations (made by the two approaches).

In the second approach it is also assumed that the input astro image is *a priori* binarized. This enables the efficient discovery of all (or a sufficient number of) star objects in the image by a simple (linear) search of the binary mask. Thus, it is considered that the coordinates $(x_i, y_i)^{(j)}$ of the corresponding centers $C_i^{(j)}$ of the components $O_i^{(j)}$, $i = 1, 2, \ldots, N$, forming the chain for a given star $O^{(j)}$, $j = 1, \ldots, S$, where S is the number of all chains in the image, have already been computed. Or for the number K of all the components (stars in the image), it is obtained that $K = NS$. Then if we translate the input image (K times) up to the center of each "star" (i.e., component of a chain) and then sum up the images obtained, we might expect the summed image to be a "cumulative" configuration of no more than $1 + N(N-1)$ local maxima, plus the surrounding noise of significantly lower average (and/or maximal) level.

Thus if the chains of the input astro image are of the type of Fig. 2(a) (i.e., linearly elongated) and if for the shifts $\Delta C_i = C_i - C_{i-1}$ from component O_{i-1} to component O_i , $i = 2, \ldots, N$ of the chains the following is satisfied:

(6) $$(|\Delta C_i| \approx |\Delta C| = cte, \; i = 3, \ldots, N) \& (|\Delta C_2| \approx 2|\Delta C|),$$

then the number of the existing maxima in the cumulative configuration discussed will be limited to $2N + 1$. These maxima (or rather areas of a single significant maximum) are located in a symmetric expansion of the initial "linear elongation", compare Fig. 2(a) with Fig. 3(a).

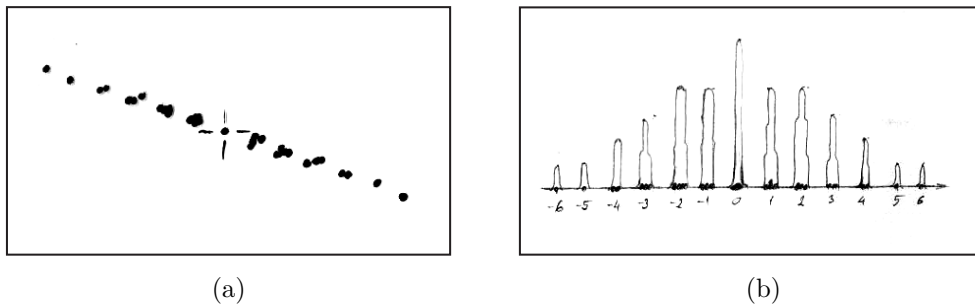(a)   (b)

Fig. 3. (a) "Cumulative" configuration for the chains of the type illustrated in Fig. 2(a), for $N = 6$ there are $2N + 1 = 13$ explicit maxima; (b) Alteration in the brightness of the corresponding "cumulative" configuration

Naturally, for the detection of these areas of "a single significant maximum", it might be necessary to perform a preliminary smoothing of the cumulative image with the help of an integrating filter with a relatively small basis (and its value may be *a priori* evaluated).
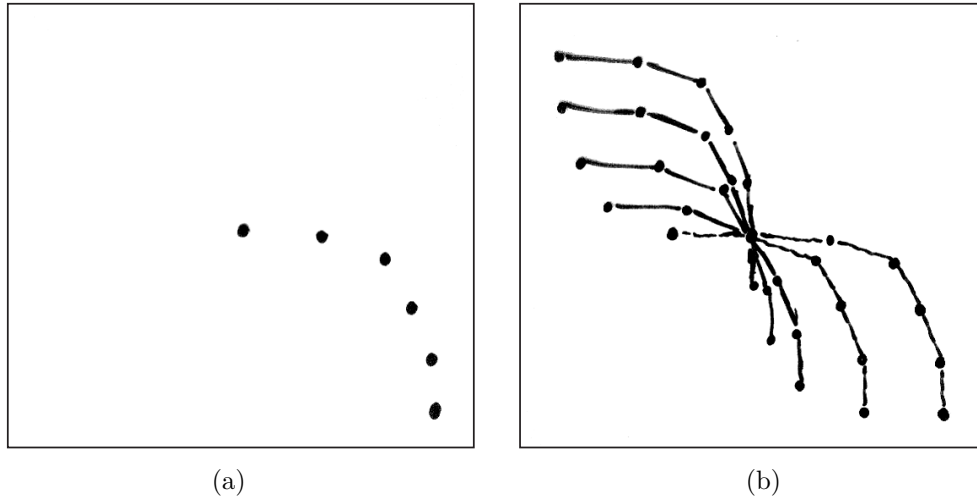
(a)   (b)

Fig. 4. (a) A more general scheme of chains of $N = 6$ components (exposures); (b) The respective "cumulative" configuration contains $1 + N(N - 1) = 31$ explicit maxima. The ribs in (b) are given as an illustration of the separate translated chains. The "initial" star in the center of (a) corresponds to the central "accumulated" star in (b)

In addition to this, for a great part of the images with chains from WF-PDB, we have $N = 4 \div 6$. Thus for the values $V(k_i)$ of the explicit maxima, i.e., in the corresponding points k_i , $i = -N, \ldots, N$, it can be written:

$$
\begin{aligned}
&V(k_0) = vSN, \\
&V(k_{-2}) = V(k_{-1}) = V(k_1) = V(k_2) = vS(N - 2), \\
&V(k_{-k}) = V(k_n) = vS(N - n), \ n = 3, \ldots N, \\
&V(k_{-N}) = V(k_{-(N-1)}) = V(k_{(N-1)}) = V(k_N) = vS;
\end{aligned}
$$
(7)

where v is the average brightness of the stars (chains components) in the initial astro image:

$$
\begin{aligned}
v &= \mathbf{E}(V^{(ji)}, (ji) = 1, \ldots K) \\
&= \mathbf{E}(P^{(ji)}(x, y), (x, y) \in O^{(ji)}, (ji) = 1, \ldots K), \quad K = NS.
\end{aligned}
$$
(8)

In the general case of cumulative configuration, see Fig. 4(b), the following can be written:

$$
\begin{aligned}
&V(k_0) = vSN, \\
&V(k_{-i}) = V(k_i) = vS, i = -N, \ldots, -1, 1, \ldots N.
\end{aligned}
$$
(9)

Table 1. Comparative analysis of the two approaches discussed

	Projective algorithm (A)		Cumulative algorithm (B)
A1	Preliminary processing of the input image—binarization.	B1	Preliminary processing of the input image—binarization.
A2	Hough transform of the binarized image in order to find the slope α of the chains. For higher speed, we work with the masks (spots) of the stars in the binarized image.	B2	Localization of all the stars, for example, by making contours of their "spots" in the binarized image. Determining the "gravity center" C_i of each spot $\langle i \rangle$, $i = 1, 2, \ldots, K$.
A3	Localization of HT-rows of chains and localization of the separate stars (components of chains) in every row. Determining the "gravity center" C_i of the stars, $i = 1, 2, \ldots, K$.	B3	Obtaining the "cumulative" image from the original input image.
		B4	Determining significant local maxima by a simple projection approach (along horizontals and verticals) or by an approach similar to steps B1 and B2 herein.
A4	Computing the distances between the stars in every HT-row and acquiring statistics of the distances obtained. Processing the statistics, considering that $N = \mathrm{const}$ and relations (6) for defining the average shifts Δ_i, $i = 2, \ldots, N$ between the components in the chain.	B5	Combinatorial approach to determine the average shifts Δ_i, $i = 2, \ldots, N$ between the chains' components. For the case discussed of (almost) evenness of distances and $N = 6$, see Fig. 3(), the expressed local maxima must number $2N + 1 = 13$, and the respective values have to satisfy (7), i.e., the combinatorial approach mentioned is significantly simplified.
A5	Determining the (potential) chains in every HT-row.		
A6	Checking the chains for probable "breaking up" between neighboring HT-rows (as a result of step A3). "Gluing" the disbanded chains.	B6	Determining the (potential) chains in the whole image.
A7	Checking all discovered chains for regularity, in order to find any collisions (flare stars).	B7	Checking all discovered chains for regularity, in order to detect collisions (flare stars).
A8	Visualization of the results in a form convenient for the user astronomer.	B8	Visualization of the results in a form convenient for the user astronomer.

The relations (7), (8) and (9) are valid under the condition that the durations of the separate exposures of an astro image are approximately equal, as mentioned in the introduction. Otherwise, the majority of chains should be expected to be irregular in their brightness, meaning that their irregularity pattern has to be discovered additionally (and beforehand) for the next use.

The above relations are given without any strict proof, because the purpose of the present paper is to evaluate only the efficiency (i.e., software complexity and processing speed) of a computer implementation, following the two approaches considered.

Table 1 presents in two parallel columns brief descriptions of the two algorithms, so that their qualitative analysis may be simplified.

9. Comparative analysis of the two approaches discussed. Advantage of the projective approach relative to the cumulative approach.

● **Consideration 1:** The two approaches require *a priori* binarization. With greater noise in the original astro image the binarization, in the sense of isolation of the useful objects (chain components), might not be perfect. HT is to a great extent invariant with respect to the noise occurring in a binarized image, because this noise is usually obtained by isotropic projection features (even in all directions), in contrast to the chains, which if prevailing, determine a stable projective maximum in the evaluation of their slope. Similar noise resistance is expected for the cumulative approach, in case the noise does not dominate over the localized components of chains (as integral intensity). Despite this invariance, it is always preferable that the *a priori* binarization maximally suppress the noise in the images. More details about this matter can be found in [8].

● **Consideration 2:** In both approaches, we can use the fact (6), for evenness of the distances between the components in a given chain.

Still, in practice these distances are not entirely equal. For a part of the plates (especially the older ones) these distances are set manually (without precise mechanics). In this sense the estimation of these distances is easier in the projection approach, where the chains are *a priori* distributed in HT-rows in the image.

The cumulative approach gives an advantage in this aspect, but only when the shape of the chains differs much from the linear one, or when it is linear, but with significantly different distances, so that the accumulated maxima, cf. (7), (8), (9), are easily separated from one another. The cumulative approach gives good results in the ideal case (6), when the accumulated maxima are localized in

$2N + 1$ points evenly distributed along the chains' direction, see (7), but not in the real cases of unevenness herein discussed.

● **Consideration 3:** The selected (ρ, θ)-HT operates effectively only with a linear ("elongated") shape of the chains that are searched in the image. Of course, many known implementations of HT detect non-linear shapes [9], but for each of them the shape must be *a priori* defined (for example, by a "reference" curve), which could re-direct us again towards the cumulative approach. We ignore this possibility, because in the practice of WFPDB no patterns with distinguished non-linear shape of the chains are available.

One disadvantage of the projective approach is that the close chains are localized in the same HT-row, i.e., a procedure is necessary for additional "sticking" of the parts of the same chain, which have appeared in several neighboring HT-rows.

On the other hand, the cumulative approach is not influenced by the shape of the chains; one of its objectives is to discover it, for example, if the sequence of plate shifts is arbitrary (chaotic), but in the case considered this does not apply. Moreover, in this case the detection of the chain shape directly from the cumulative image is quite a difficult task, due to the small deviations in the distances along the chain, which neutralizes the abovementioned shortcoming of the projective approach.

● **Consideration 4:** The complexity of (ρ, θ)-HT is of a cubic order: $(x_{\text{size}} y_{\text{size}} \theta_{\text{size}})$, where x_{size} and y_{size} are the dimensions of the input image and θ_{size} is the number of the projective directions. The preliminary binarization leads to serious acceleration of the HT procedure, about $(x_{\text{size}} y_{\text{size}})/A$ times, where A is the total area of the star spots in the binarized image. In spite of that, the duration of HT execution on chains images remains relatively big, in the range of minutes (tens of minutes for the image considered in Fig. 1). In order to speed up the process of evaluation of the chains' slope, it is suggested that HT is performed not on the initial astro image, but on a reduced, of smaller dimensions, cumulative image of the chains; by this the total running time is diminished to the range of seconds. Of course, appropriate *a priori* binarization will be necessary here as well, but it is defined more easily than the basic binarization at the beginning steps.

As for the complexity of the cumulative approach, as a whole it could be estimated as quadratic. But as already mentioned the specific noise in the images reduces the results of this otherwise efficient approach to nothing.

● **Consideration 5:** The last 2 steps of the two algorithms (approaches) are of the same type and they can be excluded from the comparative analysis.

In addition, after recognition of the common shape (form, mask) of the chains and localization of a great part of them, a final search in the original astro images can be accomplished (with the help of this mask, shape, form) for precise retrieval of all the chains by both approaches. Since the mask dimensions are usually much smaller than the dimensions of the whole image, the complexity of this search is (almost) quadratic, i.e., it can be performed directly, without the use of frequency (Fourier) approaches for acceleration [10]. This final step is similar for the two approaches and also excluded from comparative analysis.

10. Uniting the advantages in a combined algorithm. As a result of the comparative analysis and having in mind mainly considerations 2 and 3, we propose the following combined algorithm C for chain localization:

C1. Pre-processing of the input image: binarization and conversion into positive—white chains against the background of black sky.

C2. Localization of all the stars, for example by outlining the contours of their "spots" in the binarized image. Determination of the "center of gravity" C_i of each spot $\langle i \rangle$, $i = 1, 2, \ldots, K^*$, where K^* is the estimate obtained for the number of all stars (chains components) in the image. It is assumed that the competitive classical approach for localization of spots by two accumulating projections (horizontally and vertically) will cause complications of the performance due to the large number of stars expected to appear in the binarized image.

C3. Obtaining a reduced "cumulative" image (RCI) from the original input image. RCI is a square with a side size of Q_{size} that can be evaluated by the formula:

$$(10) \qquad Q_{\text{size}} = \min \left\{ C_{\text{size}}(2N + 1), x_{\text{size}}, y_{\text{size}} \right\},$$

where $N = \text{const}$ is the number of components per chain, and C_{size} is the mean size of the square area assigned to a star:

$$(10a) \qquad C_{\text{size}} = 1.25 \sqrt{(x_{\text{size}} y_{\text{size}} / K^*)},$$

where K^* is the estimate for the number of stars, i.e., their spots in the binarized image of dimensions x_{size}, y_{size} that coincide with those of the original image. The coefficient 1.25 is introduced for more precise evaluation of the value w at the following step **C4**.

C4. H of RCI for finding the slope α of the chains.

Not only for speed, but for accuracy of the result as well, a preliminary binarization of RCI is necessary, where the corresponding binarization threshold

t could be simply evaluated, accounting for possible deviations from the ideal (refer to (6) and (7)) shape of the chains, as follows:

$$(11) \qquad\qquad t = (V(k_0) - w)/2N + w,$$

where $V(k_0)$ is the maximal intensity found in RCI, and w is the average brightness of RCI. Unlike the value v, see (7), w is easy to compute beforehand. For adequacy of this estimate of t, at step **C3**, a reserve coefficient of 1.25 is used while estimating C_{size}.

C5. Localization of the HT-rows of chains in the binarized input image. Localization of the separate stars (chains components) in every HT-row and determining the chains in this row. "Sticking" some chains, possibly "disbanded" in neighboring HT-rows.

C6. Check of all incomplete chains (with a number of components $< N$) in every HT-row by overlaying (matching) the row on the original input image. A part of the components of such chains may be omitted during the binarization at step **C1**.

C7. Check of all discovered chains for regularity, in order to find probable collisions (flare stars and/or others).

C8. Visualization of the results in a form convenient for the user astronomer.

End of the algorithm.

11. Experiments. A software system was developed to experiment with the considered approaches to chain localization. The system is designed as a standard Windows XP/Vista/7 application, written in C/C++ in Borland C Builder 5.0 development environment. The experiments were run on an IBM-compatible PC: Intel Pentium 4 CPU 2.8GHz, MM $2 \div 4$ GB.

The astro image used for experiments is *a priori* converted from FITS in Windows BMP format (1B per pixel) and is shown in Fig. 1. To save computer resources only the central part of this image is used (see Fig. 5(a)), with dimensions: $x_{\text{size}} = 3148$, $y_{\text{size}} = 3131$.

The number of components of each chain in the image is $N = 6$.

The processing of the proposed algorithm is illustrated in Figs $5 \div 9$:

$\Diamond$ Step **A1** is illustrated in Fig. 5: The input image (in Fig. 5(a)) and its binarization (in Fig. 5(b)) have the same dimensions. In Fig. 5(b), the stars (i.e., chain components) are represented by white spots against the predominating black background. The (almost) isotropic noise permeated through the used adaptive binarization procedure is well visible [8].
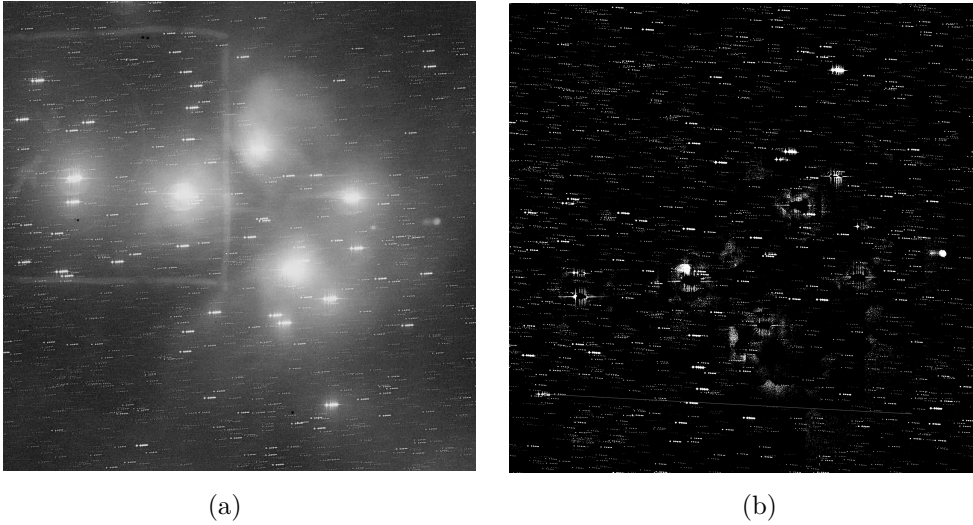
(a)   (b)

Fig. 5. (a) The central part of Pleiades (a positive, $3148 * 3131$), and (b) Its binarization. A few areas of permeated (almost) isotropic noise are clearly seen. In addition, an example of a HT-row is also shown

$\diamond$ Step **A2** is illustrated in Fig. 6 directly by the result panel of HT carried out on the binarized image from Fig. 5(b). Evaluation of the chains' slope α in the input image is the general result here, $\alpha = -3.0°$. The "exact (ρ, θ)-HT" applied here is performed via the so-called "*Basic construction*", described in [4].

$\diamond$ Step **A3** is illustrated also in Fig. 6 (above, right) with the vertical HT-histogram, for $\alpha = -3.4°$, on which we localize the HT-rows.

$\diamond$ Step **A4** is illustrated in Fig. 7 through the histogram of distances among the stars (chain components) already segmented in all the HT-rows in the input image.

Following (5), after processing the histogram, the initial offset $|\Delta C_2|$ of the chains is estimated as $|\Delta C_2| = 24$, and for the remaining offsets as $|\Delta C_i|$, $i = 3, \ldots, 6$; $|\Delta C_i| \approx |\Delta C| = 12$. At this stage, the offsets are estimated with a precision of one pixel, i.e. assuming $\varepsilon_C = 1$, cf. (4), which is quite acceptable for the chain segmentation.

The probable increase of precision, primarily through artificial scale of the input image size, leads to proportional growth of the possible values in the histogram's abscissa (Fig. 7) and respectively to quadratic increase in the execution time of either the current step **A4** or the previous step **A2**, which is unacceptable (especially for **A2**).
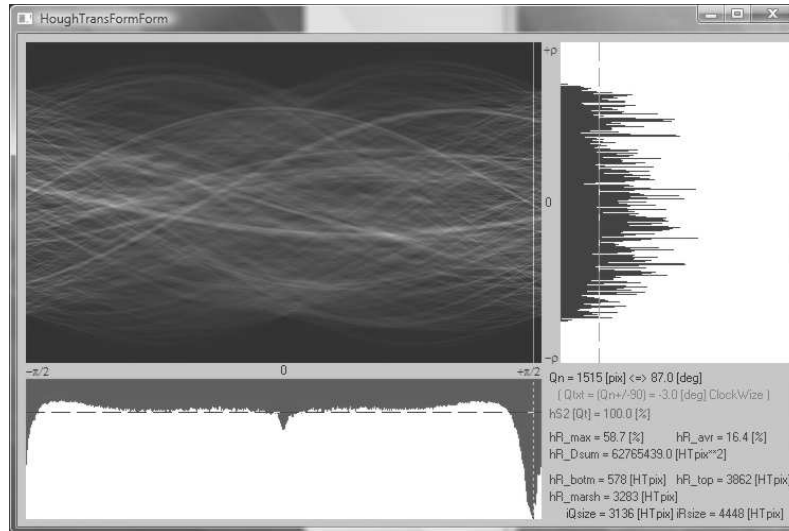
Fig. 6. Result of the operation of HT-module on the image of Pleiades (from Fig. 5(b)): (above, left): Image of the accumulated HT space (with dimensions $3136*4448$); (below, left): The estimation of the chains' slope, $\alpha = -3.0°$; (above, right): Accumulated penetrating projection of the input image for the found slope ; (below, right): Panel with other (service) results
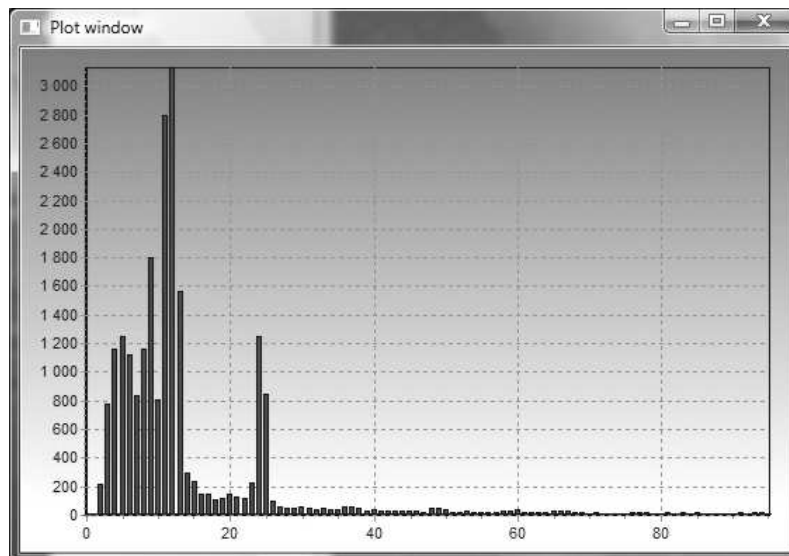

Fig. 7. Statistics of all the distances between (spots of) stars in the binarized image in Fig. 5(b), along all the encountered HT-rows of slope α, $\alpha = 3.0°$. The mean shifts per chain are computed as: $|\Delta C_2| = 24$ and $|\Delta C_i| \approx |\Delta C| = 12$, $i = 3, \ldots, 6$, by the simple rule (6)
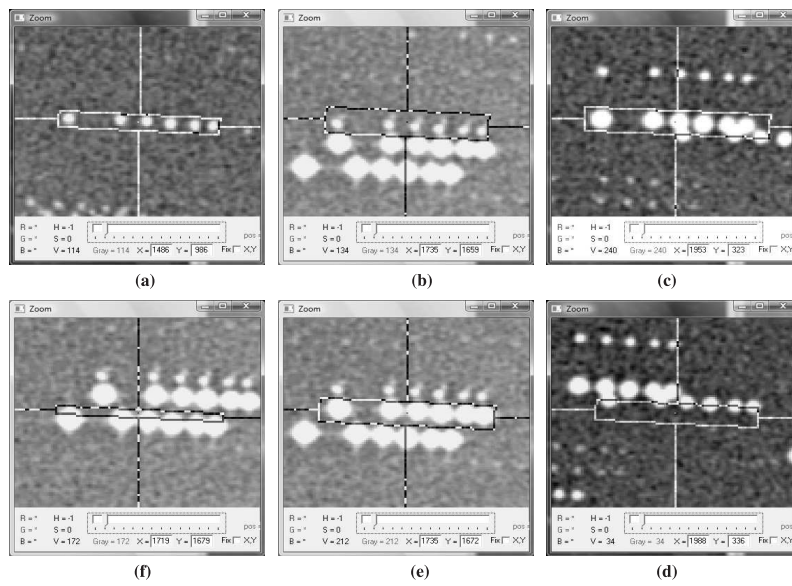
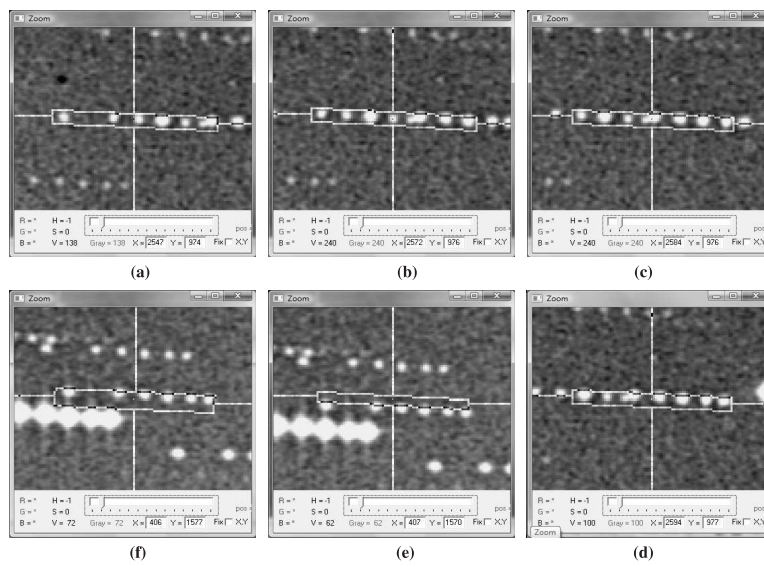Fig. 8. Localized chains from Fig. 1 (a), (b), (c)

Fig. 9. Other cases of localized chains: (a) ÷ (d) 4 overlapping chains; (e), (f) a "disbanded" chain between two neighboring HT rows

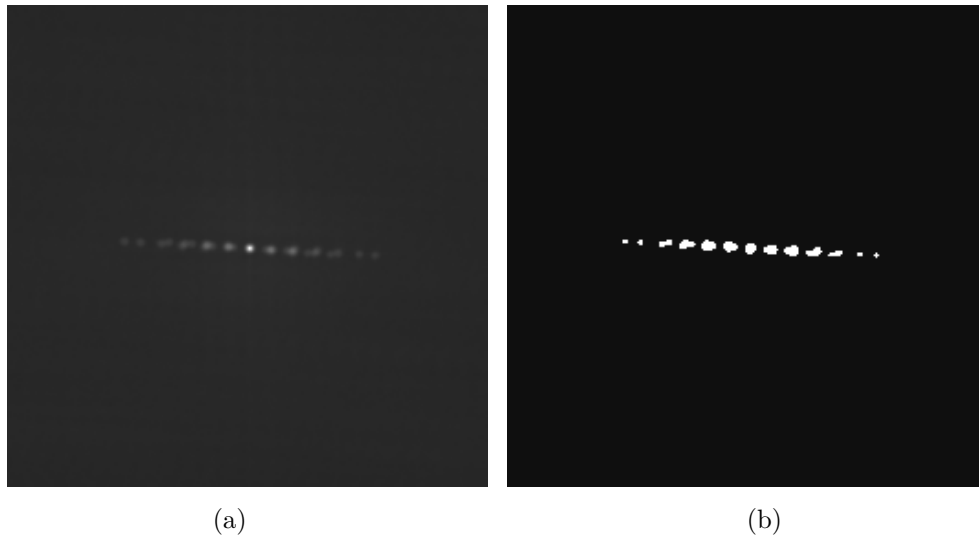(a)   (b)

Fig. 10. (a) The reduced cumulative image generated from the binarized image in Fig. 5, and (b) Its binarization. Note that (a) and (b) are positives with dimensions $264 * 264$
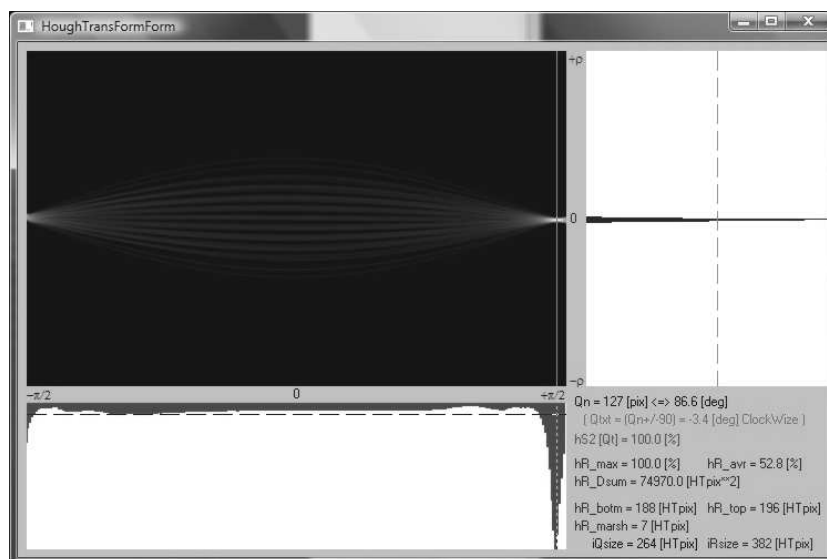

Fig. 11. The result of HT-module applied on the binarized cumulative image from Fig. 10(b). Here the evaluation of the chains' slope gives $\alpha = -3.4°$, which is similar to the result in Fig. 6

◇ Step **A5** is illustrated in Fig. 8 and Fig. 9, where several cases of segmented chains are shown. The segmentation is defined with the minimal surrounding and tilted rectangle (MSTR) for each chain.

As a rule, the chains' elongation (the long side of MSTR) is localized very precisely. Exceptions to the rule can appear at chains whose endmost components are overlapped by components of neighboring chains, see Fig. 8(c) and Fig. 9(a). Meanwhile, the short side of MSTR often exceeds the real "width" of chains, see Fig. 8(b) and Fig. 9(f). These discrepancies of the short side evaluation are due to overlapping of chains along the respective HT-row (of slope α) that is a specificity of the approach.

For the same reasons, a part of the segmented chains occur longitudinally cut off by the respective MSTR, see Fig. 8(d), (f) and Fig. 9(e), (f). The adjacent parts of these chains are localized by other MSTRs belonging to neighboring HT-rows; see Fig. 9(e), (f). A conjugation of these chain parts, following the MSTRs' neighborhood, is envisaged for the step **A6**.

Assuming a (relatively) uniform chain distribution of the input astro image, we can estimate the number of cases of imprecise evaluation of long sites of MSTRs as proportional to the average width C_{size} of the HT-rows, see (10a). At the same time, the number of cases of imprecise short sites of MSTRs and/or cases of "breaking up" chains—as proportional to the average length of the HT-rows, i.e., approximately to the image diagonal, $\mathrm{diag}(x_{size},\ y_{size})$. Obviously, $C_{size} \ll \mathrm{diag}(x_{size}, y_{size})$, i.e., discrepancies in the MSTR length occur much less often than in the MSTR width. Hence, all these discrepancies (longitudinal and/or transverse ones) will be perfected at the future step **A6**.

Fig. 10(a) illustrates the reduced cumulative image (RCI) generated from the input astro image in Fig. 5(b). The RCI is the result of steps **B2** and **B3** (of the algorithm **B**), and respectively of steps **C2** and **C3** (of the combined algorithm **C**). For the average size C_{size} of the virtual "free space" per chain component, we have: $C_{size} = 25$, see (10a), whence for the size Q_{size} of the RCI we obtain: $Q_{size} = 264$, according to (10).

Fig. 10(b) illustrates the binarization carried out on the RCI to lighten the (eventual) next HT.

Fig. 11 illustrates the HT on the binarized RCI. The result for the chains' slope here is: $\alpha = -3.4°$. The discrepancy of $0.4°$ towards the estimation from Fig. 6 can be explained with current inaccuracies of RCI generation, with natural imprecision caused by RCI binarization, and most of all with an insufficiently precise performance of the "exact (ρ, θ)-HT". These shortcomings will be neutralized at subsequent stages of the development.

At this stage, the following is experimented upon:
- Projective algorithm A (without steps A6 and A7);
- Cumulative algorithm B (without steps B5, B6 and B7);
- The combined algorithm C that is proposed, now under development.

The total number of HT-rows encountered at step $A2$ is $R = 443$. An example of an HT-row, namely, number 373, is shown over the binarized image in Fig. 5(b). Two "normal" chains are (only) detected in the MSTR of this HT-row.

For the purposes of the analysis, we can preliminarily estimate the total number K of stars (chain components) in the image observed (Fig. 5(a)) via the SExtractor program: $K = 17280$. Thus, the total number S of chains expected in the image can be evaluated as $S = K/N = 2880$.

The total number (K^*) of the stars discovered at step $A3$ by algorithm A is: $K^* = 25174$, which is quite unexpected: $K < K^*$, and is explained by the noise due to binarization at step $A1$ causing emerging of false stars.

The total number S^* of the localized "normal" chains at step $A5$, i.e., such that do not satisfy the criteria (3) and (4), is $S^* = 2784$, which is quite expected ($S^* \approx S$).

Thus the number Q of localized stars which are not tied in chains due to some omissions at step $A1$ (the binarization of input plate image) is $Q = K^* - NS^* = 8470$. The searched flare stars will be found among them. The chains in which these stars are included will be localized at steps $A5$ and $A6$, which are now under development.

Table 2. Execution time (processing of the image in Fig. 5(a)) in steps of algorithm A

Steps	A1	A2	A3	A4	A5	A6	A7	A8	Total
Time [sec]	4.3	1342.5	69.7	0.01	0.03	–	–	0.1	1417.0
Input: [MB]	9.96	9.96	13.95	0.02	0.67	–	–	1.56	9.96
Output: [MB]	9.96	13.95	0.67	0.00	0.08	–	–	0.00	–
Relative time: [sec/MB]	0.4	134.8	5.0	0.5	0.04	–	–	–	142.3

Table 2 and Table 3 show the measured duration of the experimental execution by steps of both algorithms—the basic A and the combined C. The cumulative algorithm B is considered an intermediate version between A and C, and in this way not interesting any more.

It is well visible from Table 2 that the "bottle neck" of algorithm A is the step $A2$, performing an HT on the entire input image (see Fig. 5(b)). It is avoided by the algorithm C, where the necessary HT (step $C4$) is conducted on

Table 3. Execution time (processing of the image in Fig. 5(a)) in steps of algorithm C

Steps	C1	C2	C3	C4	C5	C6	C7	C8	Total
Time [sec]	4.3	0.8	2.5	0.3	68.8	–	–	0.1	76.8
Input: [MB]	9.96	9.96	11.85	0.07	9.96	–	–	1.56	9.96
Output: [MB]	9.96	1.84	0.07	0.10	0.67	–	–	0.00	–
Relative time: [sec/MB]	0.43	0.08	0.21	4.28	6.91	–	–	–	7.71

the binarized RCI, obtained at steps **C2** and **C3**, and of a much smaller size than the input image. Thus, the general acceleration in processing speed in the current case can be evaluated as: 1342 / 77 ≈ 17 times (compare the total durations from Tables 2 and 3).

In this line of thinking, for the original astro image in Fig. 1, with size 7460×8168 ($\sim$ 59 MB), we can expect that the total execution time through algorithm C will be < 450 sec, which is quite acceptable with the computer power used at the current stage of development. The expected total time through algorithm A is in the order of hours, i.e., the comparison of processing speed is obviously in favor of the algorithm C. Besides, the critical step **A2** is non-executable in the case of Fig. 1, because the operation system Windows 7 fails to allocate the necessary main memory (at this stage).

Finally we can conclude that the effective operation of the proposed combined algorithm C is experimentally proved.

12. Conclusions. A new approach has been proposed, based on the advantages of two competitive heuristic approaches to localization of star chains in an astronomical plate image. The approach uses the so-called "exact" implementation of (ρ, θ)-HT, well known in classical image processing [4]. The approach is based on a preliminary statistical estimation of the chains' shape in the input astro image, through the so-called reduced "cumulative" image, which is effectively processed by (ρ, θ)-HT. The approach is the basic stage in the solution of automatic discovery of "flare stars" in plate images with chains.

The development of a software system designed for this purpose is at its final stage. The efficiency of the suggested approach, whose theoretic grounds are exposed herein, is also experimentally proved.

The perspectives of future work on the research discussed include:

• Development of the necessary software modules of the suggested algorithm C, more precisely **C6**, **C7** and **C8** modules, briefly defined in Section 10.

• Precise implementation of the exact (ρ, θ)-HT: At the present stage an approximate implementation, offered in [4], is used for the iterative achievement of (almost) maximal accuracy. For reasons of processing speed, the (ρ, θ)-HT currently experimented with is an acceptable compromise with the accuracy of results. Because of its being non-iterative, the precise implementation is expected to be much more speedy than the current performance of the exact (ρ, θ)-HT.

The following will be defined as possible extra problems, which are also interesting to solve:

• The possible reconstruction of the "original sky" from the plate image with chains;

• The optimal choice of the direction (and the value) of the shifts before performing the experiment with chains photographing.

The present development is within the frames of Astroinformatics project and is directed towards serving user's requests to WFPDB. In this sense a reasonable "implementation" of the approach proposed here is expected to be within the frames of the European Virtual Observatory [3].

REFERENCES

[1] Aniol R., H. W. Duerbeck, W. C. Seitter, M. K. Tsvetkov. An automatic search for flare stars in southern stellar aggregates of different ages. Flare Stars in Star Clusters (Eds L. V. Mirzoyan, B. R. Pettersen, M. K. Tsvetkov), Associations and Solar Vicinity, IAU Symp. 137, Kluwer Acad. Publish., Dordrecht-Boston-London, 1990, 85–94.

[2] Winterberg J., M. Nolte, W. C. Seitter, H. W. Duerbeck, M. K. Tsvetkov, K. P. Tsvetkova. Flares and Flashes. In: Proceedings of the IAU Colloquium No. 151 (Eds J. Greiner, H. W. Duerbeck, R. E. Gershberg), Springer Verlag, Berlin, 1995, 119–120.

[3] Kounchev O., M. Tsvetkov, D. Dimov et al. Astroinformatics: A Synthesis between Astronomical Imaging and Information & Communication Technologies. Modern Trends in Mathematics and Physics. ISBN 978-954-580-264-5, *Bulg. J. Phys,* **32** (2009), No 2, 60–69.

[4] Dimov D. T. Using an Exact Performance of Hough Transform for Image Text Segmentation. In: Proceedings. of the IEEE Conf Image Proc (ICIP), Thessaloniki, Greece, Vol. **1**, 2001, 778–781.

[5] STARCK J.-L., F. MURTAGH. Handbook of Astronomical Data Analysis. Springer-Verlag, 2002.

[6] CFITSIO – A FITS File Subroutine Library. GODDARD SPACE FLIGHT CENTER.
`http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html`, June 2009

[7] BERTIN E., S. ARNOUTS. SExtractor: Software for source extraction. *Astronomy & Astrophysics Supplement Series*, **117** (1996), 393–404.

[8] DIMOV D., A. DIMOV. Data Driven Approach to Binarization of Astronomical Images. In: Proceedings of the CompSysTech'10, June 17–18, 2010, Sofia, Bulgaria, Mini-Symposium on Astroinformatics, 2010 – ACM International Conf. Proceeding Series, Vol. **471**, ACM PRESS, NY, USA, ISBN 978-1-4503-0243-2, 478–484.
`http://portal.acm.org/citation.cfm?id=1839379.1839465&coll=DL&dl=GUIDE&CFID=4447488&CFTOKEN=73980402`

[9] ILLINGWORTH J., J. KITTLER. A Survey of the Hough Transform. *Comp. Vision, Graphics, and Image Proc.*, **44** (1988), 87–116.

[10] DUDA R. O., P. E. HART, D. G. STOCK. Pattern Classification (2nd ed.), John Wiley & Sons, Inc., 2001.

Dimo Todorov Dimov
Bulgarian Academy of Sciences
Institute of Information and Communication Technologies
Acad. G. Bontchev Str., Bl. 29-A
1113 Sofia, Bulgaria
e-mail: `dtdim@iinf.bas.bg`

Katya Petrova Tsvetkova
Bulgarian Academy of Sciences,
Institute of Mathematics and Informatics
Acad. G. Bontchev Str., Bl. 8
1113 Sofia, Bulgaria
e-mail: `katya@skyarchive.org`

Milcho Kirilov Tsvetkov
Bulgarian Academy of Sciences
Institute of Astronomy / SSADC
72, Tsarigradsko Shose Blvd
1784 Sofia, Bulgaria
e-mail: `milcho@skyarchive.org`

Alexander Assenov Kolev
National Military Academy "G. S. Rakovsky"
82, Evlogi Georgiev Blvd
1336 Sofia, Bulgaria
e-mail: `alexkolev@yahoo.com`

Ognyan Ivanov Kounchev
Bulgarian Academy of Sciences
Institute of Mathematics and Informatics
Acad. G. Bontchev Str., Bl. 8
1113 Sofia, Bulgaria
e-mail: `kounchev@math.bas.bg`