STAR CROSS-IDENTIFICATION IN BIG CATALOGS WITH SIGNIFICANT EPOCH DIFFERENCE AND WITHOUT PROPER MOTIONS

L.I. YAGUDIN
Pulkovo Observatory
St. Petersburg, Russia

Abstract. A classical task of cross-identification of stars taken from different catalogs becomes a non-trivial one in a case of very big catalogs with significant epoch difference and without proper motions. Some new concepts for this problem solution are proposed here. The method discussed was applied for the Guide Star Catalog with the Astrographic Catalogue identification. About 4.3 million stars were identified in both catalogs.

1. Intention, Problems

The concepts and algorithms described here were developed during a short stay at the Astronomishes Rechen-Institut, Heidelberg, in in connection with the scientific collaboration agreement between the Deutsche Forschungsgemeinschaft and the Russian Academy of Sciences. As a part of our common work it was necessary to identify stars in two big catalogs: in the Guide Star Catalog (GSC) of the Space Telescope Scientific Institute and in the Astrographic Catalogue (AC).

It was practically impossible to apply the usual procedures and existing software to fulfill this job. Problems are evident. These are:

- extremely large number of stars in both catalogs (about 16 million in GSC and about 4.5 million in AC) and a very big star density as a consequence;
- very big epoch difference (about 80 years) and absence of proper motions for more than 97 per cent of the stars;

I. M. Wytrzyszczak, J. H. Lieske and R. A. Feldman (eds.),
Dynamics and Astrometry of Natural and Artificial Celestial Bodies, 511, 1997.
© 1997 *Kluwer Academic Publishers. Printed in the Netherlands.*

https://doi.org/10.1017/S0252921100047035 Published online by Cambridge University Press

- absence of additional information like magnitudes or spectral classes which could help to resolve ambiguous cases of identification. There are only star brightness estimates in each catalog, but these values are very rough and non-homogeneous;
- both catalogs are the collections of photographic plates, partly overlapped. While GSC stars are identified on different GSC plates, this job is not done for the Astrographic Catalogue. So, it was necessary to identify AC with AC and AC with GSC in the same procedure.

2. Solution

Some simple geometrical concepts were developed and implemented as algorithms and FORTRAN programs for resolving a number of problems. We describe here just the main points of our method without any technical details.

2.1. IDENTIFICATION WINDOW SHAPE

Windows usually applied for star identification – circular areas on the sky – work well just for the identification of isolated stars. Even in such cases one never can be sure in his identification because other candidates can exist out of one's window very close to its border. If we have many stars close to one other, such windows do not work. Our first principal proposal is to use a *window in the form of two circles with same center.* In the algorithm, the identification rule is as follows: stars from different catalogs can be identified if they are located in the inner circle and there are no other stars in a ring between circles. The evident advantage of such window is a high assurance of every accepted identification.

2.2. IDENTIFICATION WINDOW SIZE

The question "how big can the window be?" is very important. Usually a window of some fixed size is used in accordance with a catalog's accuracy and/or epoch difference. By using a fixed size window we usually miss stars with big proper motions and stars with big errors in coordinates. Our second principal proposal is to use for identification *the window of a variable size.* What we have to fix really are the minimum and maximum allowable values for the inner circle radius and the external-to-inner-radius ratio.

The minimum value of inner circle radius can correspond to the typical image size on the plate (which is a measure of the quality of the telescope optics). Its maximum value depends on the catalogs' epoch difference and on a range of proper motions which we would like to detect. By taking a big enough upper limit we can identify stars with both slow and fast

https://doi.org/10.1017/S0252921100047035 Published online by Cambridge University Press

proper motions in the same procedure. This is the main advantage of our variable-sized-window concept.

The ratio of external and inner circle radii is a measure of identification reliability. Taking a greater ratio one gets more reliable identifications but at the same time one misses more identifications in dense fields of sky and vice versa.

Using numbered parameters, the computer program has to choose for itself the actual values of radii in any individual sky field depending on star population in this field.

2.3. IDENTIFICATION WINDOW POSITION

Traditional identification procedure takes the window centered on a star from one catalog and looks for a corresponding star in the second catalog. We intend to develop a more universal algorithm which could identify stars in more than two catalogs and/or stars from overlapped plates (as in the AC case) at the same time. So, our third principal proposal is *to center the identification window in some point*, which is a center of group of star-candidates; *which is i.e. the center of circle surrounding all star-candidates.* Two, sometimes three stars have to be used to calculate coordinates of this point.

2.4. IDENTIFICATION WINDOW CONSTRUCTION

Realization of the proposed principles is not so easy as the usual identification procedure of course. Constructing such an identification window different for every individual star (in other words, gathering stars in a groups) is the main tool of the proposed method. The algotithm developed by us is compex enough, but the purpose of this paper is to explain just the main concepts. So, we will skip technical details and tricks. In other words our program, starting with some specific star adds step-by-step neighboring stars from both catalogs to the group, thus increasing on each step the size of the identification window and changing its position. The procedure is finished when the ring between external and internal circles is empty or when the upper limit for window size is achieved.

2.5. ANALYSIS OF IDENTIFICATION WINDOW CONTENT

Three additional circumstances can be taken into account on this stage. First, the GSC is significantly denser than the AC, i.e. we can suppose with very high probability, that any AC star can be found in the GSC. That means also, that if more than one GSC star could be identified with one AC star, then the more preferable candidate is the brightest GSC star. This

https://doi.org/10.1017/S0252921100047035 Published online by Cambridge University Press

fact save us from comparing the stellar magnitudes in each catalog – which would be practically impossible because they are very non-homogeneous. Instead of this we can just compare magnitudes in the same GSC catalog, which is more reliable.

Second, two stars from the same plate cannot be the same star, evidently. So, just retaining the plate number among other star parameters helps to resolve many doubtful cases.

Third, the identification process is a sequential one of course. So, we can easily imagine a situation when, passing through some dense sky region, the program skips some star as an unresolved case and then, due to variability of the identification window size, it identifies another star which was a reason for failure in the first case. This suggests that the identification procedure must be iterative: by deleting now-identified stars from the source files and by repeating the procedure we can get new identifications.

Five different algorithms are developed for treatment of the content of the identification window corresponding to different star configurations on the sky. These configurations are: 1) single isolated star; 2) single star with faint companions; 3) multiple isolated stars; 4) multiple stars with faint companions; and 5) very dense sky region. Each algorithm works with source files independently. The logic of decision taking is more or less evident in each case. That is described below.

2.5.1. *Single isolated star*

This is the simplest case. Starting with some AC star this algorithm groups all neighbouring stars from both catalogs by constructing a corresponding identification window. Then this group of stars is identified as the same star if:

- there is only one GSC star in the inner circle;
- there is at least one AC star in the inner circle;
- all AC stars in the inner circle are from different AC plates;
- there are no stars in a ring between external and internal circles.

2.5.2. *Single Star with faint companions*

This algorithm is more complicated than the previous one. First, it is working with a wider region of the sky. Then, it tries to pre-identify the AC stars following similar principles. Finally, it identifies the brighter GSC star with some AC star if:

- there is only one AC star in the window;
- this AC star is in the inner circle;
- the brightest GSC star is located in the inner circle;
- all other GSC stars in the window are from the same GSC plate as the brightest one.

https://doi.org/10.1017/S0252921100047035 Published online by Cambridge University Press

2.5.3. *Multiple isolated stars*

This algorithm like the previous one first pre-identifies AC stars. It continues the analysis if all stars are situated in the inner circle, if there are no other stars in the ring and if the number of GSC stars is equal to the number of AC stars. The next step is very important: before making any decision, the program calculates the mean distance between the group of AC stars and the group of GSC stars and shifts one of the groups by this distance. Such a trick eliminates the galactic rotation and cancels a problem of big epoch difference and the absence of proper motions for the majority of stars. After shifting, the program identifies every star from one catalog with the nearest star from the second catalog. Identification is accepted if every star is identified only once and if the distances in pairs are not larger than some critical value.

2.5.4. *Multiple stars with faint companions*

This algorithm is very similar to the previous one. The only difference is that before identification, the program removes from the comparison the faintest GSC stars until the number of GSC stars becomes equal to the number of AC stars. At that point, any removal is accepted if there still are other GSC stars from the same plate in the identification window.

2.5.5. *Very dense sky region*

This algorithm can be applied just after applying the previous algorithms. The main trick here is to shift the AC sky to the GSC sky in such regions using already-identified stars. After shifting, the situation becomes more clear and the previous algorithms can yield new identifications.

3. Application

All the concepts described here, excluding algorithm 2.5.5, were implemented as a set of FORTRAN programs which were applied for the GSC and AC star identification. About 4.3 million stars were identified in both catalogs. About 0.2 million stars remained unidentified. Later visual checking of about 64 000 identified stars in one of the GSC zones has shown that:

1. all unidentified AC stars belong to extremely dense sky regions. A significant part of them could be identified in the framework of proposed concepts, but we had no time to develop the corresponding computer program;
2. approximately 53 400 stars were identified by algorithm 2.5.1;
   approximately 6800 stars were identified by algorithm 2.5.2;
   approximately 2800 stars were identified by algorithm 2.5.3;
   approximately 1000 stars were identified by algorithm 2.5.4;

https://doi.org/10.1017/S0252921100047035 Published online by Cambridge University Press

3. Perhaps 286 identifications can be erroneous. Among them:
   88 stars identified by 2.5.1 algorithm or 0.14% of all identifications;
   180 stars identified by 2.5.2 algorithm or 0.28% of all identifications;
   6 stars identified by 2.5.3 algorithm or 0.01% of all identifications;
   12 stars identified by 2.5.4 algorithm or 0.02% of all identifications.
   Reasons for doubts are as follows:

   - 2.5.1 algorithm: the distance between AC stars is too large. Two scenarios can be proposed: 1) there are 2 stars on the sky; only one of them was measured on one AC plate and on the GSC plates; only the second one was measured on the other AC plate. In this case the 2.5.1 algorithm is bad, but the probability of such an event is very low indeed; 2) there is only one star in the region on the sky, but its AC measurements are wrong (misprint, mispunch etc.). This is more possible.
   - 2.5.2 algorithm: faint GSC companions are significantly closer to the AC star than the brightest GSC star. Magnitude difference can be rather large, more than one magnitude. Some scenarios can be proposed. The more possible is low quality of the GSC magnitude estimates. It would be possible to change the GSC star chosen criteria, i.e. to introduce some limit for magnitude differences. In such a manner we can eliminate erroneous identifications, but at the same time we would miss a lot of correct identifications. This is a subject for further research.
   - 2.5.3 and 2.5.4 algorithms have same problems as the 2.5.2 algorithm.

4. It is possible of course to improve proposed concepts and especially the algorithms, but in general they can be regarded as successful.

Acknowledgements. The author wishes to thank the Deutsche Forschungsgemeinschaft for financial support and Dr. S. Roeser and Dr. U. Bastian for fruitful discussions.

https://doi.org/10.1017/S0252921100047035 Published online by Cambridge University Press