Journal of The Korean Astronomical Society
37: 79 ∼ 85, 2004

EFFICIENT PERIOD SEARCH FOR TIME SERIES PHOTOMETRY

Min-Su Shin and Yong-Ik Byun
Yonsei University Observatory and Department of Astronomy, Yonsei University, Seoul, 120-749, Korea
E-mail: astroshin@yonsei.ac.kr and byun@obs.yonsei.ac.kr
(Received June 25, 2003; Accepted June 15, 2004)

ABSTRACT

We developed an algorithm to identify and determine periods of variable sources. With its robustness and high speed, it is expected to become an useful tool for surveys with large volume of data. This new scheme consists of an initial coarse process of finding several candidate periods followed by a secondary process of much finer period search. With this multi-step approach, best candidates among statistically possible periods are produced without human supervision and also without any prior assumption on the nature of the variable star in question. We tested our algorithm with 381 stars taken from the ASAS survey and the result is encouraging. In about 76% cases, our results are nearly identical as their published periods. Our algorithm failed to provide convincing periods for only about 10% cases. For the remaining 14%, our results significantly differ from their periods. We show that, in many of these cases, our periods are superior and much closer to the true periods. However, the existence of failures, and also periods sometimes worse than manually controlled results, indicates that this algorithm needs further improvement. Nevertheless, the present experiment shows that this is a positive step toward a fully automated period analysis for future variability surveys.

Key words : methods: data analysis — stars: variables — surveys

I. INTRODUCTION

The project of Yonsei Survey Telescopes for Astronomical Research (YSTAR) aims to carry out time series observations for the entire sky (Byun et al. 2001). One of our science goal is to identify and characterize variable sources. There are other survey observations of similar nature such as MACHO, OGLE, ROTSE, and ASAS. While the primary goals of these projects are different from each other and range from Gamma-Ray Bursts to gravitational microlensing events, they invariably contribute to the discovery of a large number of previously unknown variable stars.

Variable stars are most readily studied by the way their brightness changes. It is therefore natural that their characterization and classification are primarily based on the periodic properties of their phased light curves (e.g. Sterken and Jaschek 1996). The period of brightness variation is not only an important physical parameter for a variable star, but also an absolute prerequisite for the construction of phased light curves.

There exist several programs written for the estimate of variability periods. However, all the programs we investigated require very attentive manual control throughout the process. Depending on the cases, it involves an effort of repeated trials which can take several minutes or more of human time to process a single variable star. In the age of terra byte survey data, it is therefore essential to automate this process of period determination.

Corresponding Author: Y.-I. Byun

We present here an algorithm to derive precise periods in an efficient and very robust way. In the sections 2 and 3, we discuss various period search methods and describe how we combine them into a new process. During the development phase, we experimented with several time series data from OGLE (Udalski et al. 1994) and ASAS (Pojmanski 2000) databases. Some examples are presented here to illustrate the workings of our algorithm. In section 4, we show the results of our experiment with a large number of variable stars from the ASAS survey. Detailed comparisons are made between our periods and published estimates.

II. PERIOD SEARCH METHODS

Several methods have been previously conceived for period search, each focusing on different aspects of time series data. These methods can be classified into two major categories. One is to fit a periodic function to the data and evaluate the fitting statistic. For example, Scargle (1982) method is a modification of Fourier transformation, which uses orthogonal trigonometric functions. However, this approach is not only contaminated by aliasing effects but also relies on a poorly known statistic. In the other category, time series data are folded with a trial period and checked whether it has an appropriate period by calculating a relevant statistic. There are a few different methods based on folded data, but the phase dispersion minimization method (Stellingwerf 1978) is most widely used.

A critical measure of successful period search is how distinctively the significance of estimated period can be evaluated from a given set of data. This is especially important for an automated period search, where the most plausible period is selected by the statistical inference only without the aid of human judgement.

In a previous study, Carbonell et al.(1992) pointed out that, according to the type and size of gaps present in time series data, we need to choose a period search method which avoids spurious periods. It was also demonstrated that methods which adopt smooth periodic model functions are more sensitive than those which use phase binning. In addition, methods using the same model function appear to have an equal sensitivity in spite of different statistics (Schwarzenberg-Czerny 1998, 1999).

There has been no serious attempt to evaluate sensitivity and efficiency of every available period search algorithms. However, Schwarzenberg-Czerny (1996, 1997) argued that those methods using the analysis of variance (hereafter AoV) statistic are of superior performance and can be calculated with relatively high efficiency.

For the project YSTAR, we tested several methods including those mentioned above, for their efficiency and accuracy. Most of them turned out to be quite useful and gave apparently well assembled light curves. In order to reach a satisfactory solution, however, we invariably had to repeat several trials each with parameter adjustments. This would not be acceptable if we need to deal with a large number of variable star candidates. Besides the need for human supervision, we sometimes find that good looking phased light curves do not always correspond to accurate periods. In addition, we also find that each method had different strength and weakness, and none of them can treat the large variety of periodic variability all by itself. For our variability survey project, it is clear that we need a better tool.

III. MULTI-STEP PERIOD SEARCH

(a) First Step : Phase Dispersion Minimization and Polynomial Function Fitting

In our algorithm, we first fit periodic orthogonal polynomials to find candidate periods. The candidates are selected based on the AoV statistic (Schwarzenberg-Czerny 1997). This step is to use a smooth function for fitting. Independent from this, we also apply the phase dispersion minimization (hereafter PDM) scheme to the data. PDM provides additional candidate periods. Another AoV statistic (Schwarzenberg-Czerny 1996) is employed in this process. The application of two very different methods acts as a test of reliability for each other and also give us complimentary set of candidate periods. In order to use these methods, we wrote computer codes using the algorithms given in the aforementioned papers of Schwarzengerg-Czerny (1996, 1997).

Both of these methods require a preset range of

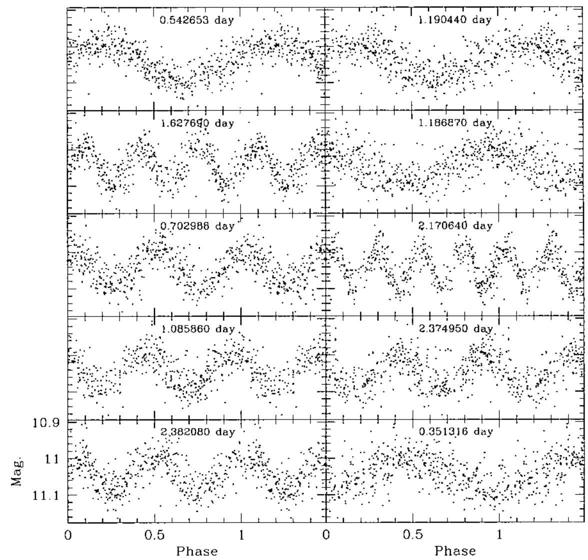

Fig. 1.— Candidate periods chosen by function fitting (*left*) and by PDM using AoV (*right*). Light curves are arranged according to the size of AoV statistic; the top corresponds to the largest value. This is the variable star 015647-0021.2 in ASAS Catalog 2 where its period was given as 0.351447 day, similar to the bottom panel in the right.

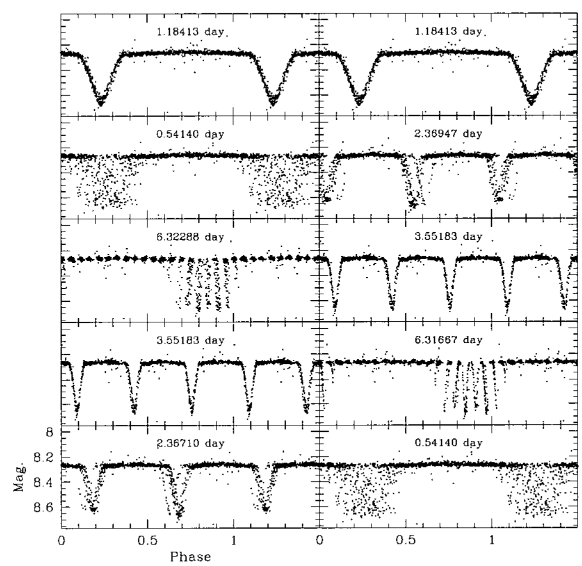

Fig. 2.— Periods selected for an eclipsing binary star, 204045+0056.4 of ASAS Catalog 2; the light curve arrangement is the same as in Figure 1. The period estimated by ASAS team is 2.368131 day.

trial periods, i.e. frequencies. Determination of the frequency range should depend on data sampling parameters. Unfortunately, optimal sampling frequency is usually not obvious from data in most cases. Therefore we set the maximum trial frequency to be half of the highest frequency in the time series data. On the other hand, the minimum frequency limit is naturally set to the frequency equal to the entire time span of time series observations. The range of allowed frequencies, and therefore the number of trial frequencies, have a great impact on the processing time of frequency analysis. However, frequency range too small or too sparse will result in misleading and wrong period estimates.

In order to verify the usefulness of these two different period search methods, we experimented with several data sets from the survey database released by OGLE and ASAS (Pojmanski et al. 1998) projects. As described below, we find several interesting cases which demonstrate the limitations of this first step as well as the need for further step of refined period search.

According to our experiments, there are cases where both function fitting method and PDM method work similarly well and provide plausible phased light curves. However the apparent look can be quite deceiving. Fig. 1 is an example of such cases. Five candidates are chosen from each method ordered by their AoV statistic. If these were to be examined by human, some of light curves with too many peaks can be readily disregarded as not physically possible. The choice of best solution however is not so easy.

The top panels, with largest AoV statistic, are apparently similar to each other and both may be regarded as good solutions. However they have very different periods of 0.542653 and 1.190440 days. Which one is the correct period? The second and fifth solutions from PDM method are also of similar quality, but again with quite different values of period, 1.186870 and 0.351316 days. Maybe one of them is the real period? It turns out that all of these are subject to the serious alias effect of 1 cycle/day present in the ASAS data sets. This shows that even if we are successful in making smooth and attractive looking phased light curves, it can be still difficult to choose among them, and the period cannot be conclusively determined.

In some other cases, both method give best candidate periods of exactly same value, but fail to be physically meaningful. For an eclipsing binary shown in Fig 2, they recommend 1.18413 day as the best candidate period. However, a close inspection of phased light curves indicates that the second best choice by PDM, 2.36947 day, or the fifth choice of function fitting method, 2.36710 day, is more representative of the light curve of an eclipsing binary. There exists two minima of different depth in the light curve, but the largest AoV solutions fail to separate the primary dim data points from the secondary. We should favor the physically meaningful solution, although the light curve indicates that exact phasing has not been achieved. Does this mean that statistical inference do not work in this case? It will be shown below that more refined period search is still capable of successfully finding physically meaningful solution.

There are also cases where function fitting and PDM methods present completely different phased light curves and periods. An example is shown in Fig 3(*left*). While PDM favors the solution of about 56 days, which seem quite appropriate in this case, the function fitting method gives much larger solution. The cause of this very different period estimation can be found in their AoV statistic distributions, compared in Fig 3(*right*). As shown in the figure, the distribution and ordering of AoV peaks differ greatly from each other. Unlike common belief, this diagram clearly shows that the highest peak of any given method does not always correspond to the most accurate period solution.

(b) Second Step : Cubic Spline Function

Our experiment with several different sets of published data shows that the most widely used methods such as function fitting and PDM often provides contradicting results. In addition to this, their best candidate period is not always the best solution. In order for the period search process to be fully automated, the application of function fitting and PDM method is not enough. As we have shown in the previous section, there are cases where both methods fail or give very different suggestions.

Our algorithm, therefore, starts with the function fitting and PDM process, but subjects the data to go through another period search method based on cubic spline (Akerlof et al. 1994). The basic idea behind this algorithm is to use a sum of cubic splines for the least square fitting, and then to find the most appropriate period which minimizes χ^2 values. This is very robust period estimation by itself and any form of light curve can be accounted for. However, its efficiency is far behind the two methods introduced above and sometimes the solution gets trapped inside a local minimum instead traveling toward the global minimum.

This problem is overcome by running function fitting and PDM routines first, and then to use their short lists of candidate periods as the search region for refined cubic spline method. This narrows the search range, boosts up the efficiency of cubic spline analysis, and gives us another statistic value χ^2, which helps to confirm more appropriate period. For this process, we constructed a computer program which uses the subroutine library provided by Akerlof et al. (1994).

In this second step, the candidate periods are more precisely estimated. For each candidate period, a new search is made around the given candidate looking for another nearby χ^2 minimum. The benefit of this is very clearly demonstrated in Fig. 4, where we run the case of Fig. 2 by supplying the same ten candidate periods into cubic spline method. More precise period of 2.367882 days is found almost immediately. The upper panel of

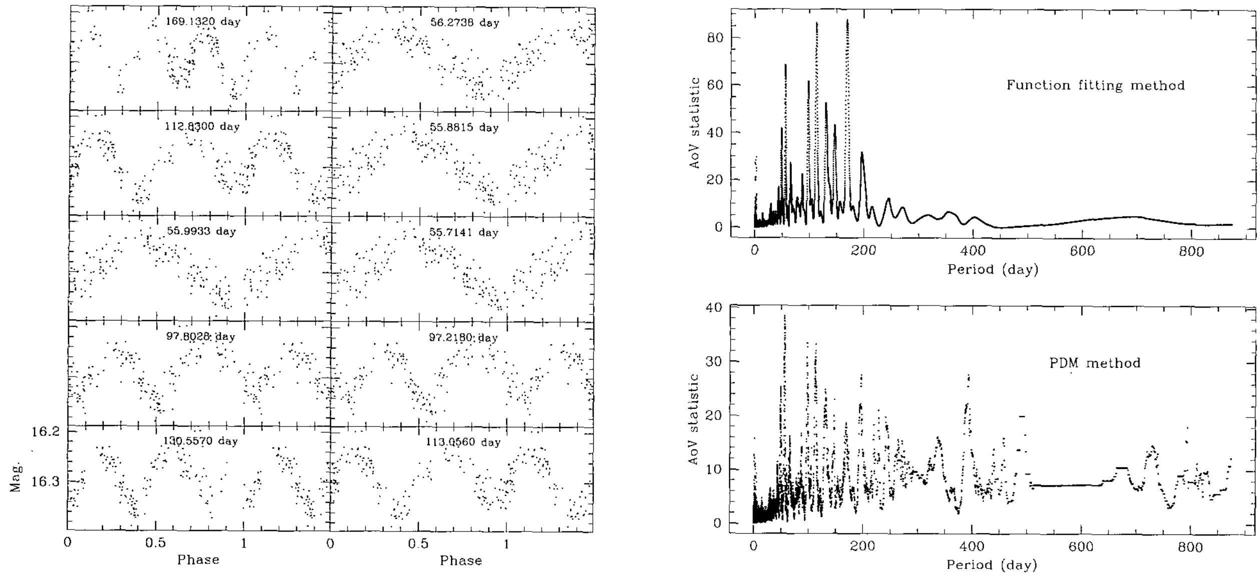

Fig. 3.— (*left*) An example of period search where function fitting and PDM methods give wildly different solutions. Light curve arrangement is the same as in Figure 1. This variable star is Star no. V70 in OGLE field BW1. OGLE team did not specify its period. (*right*) Comparison of AoV statistic from function fitting and PDM method for the variable star shown in Fig 3(a). Note the difference in peak locations. The highest peak of the PDM AoV is much lower than that of the function fitting AoV.

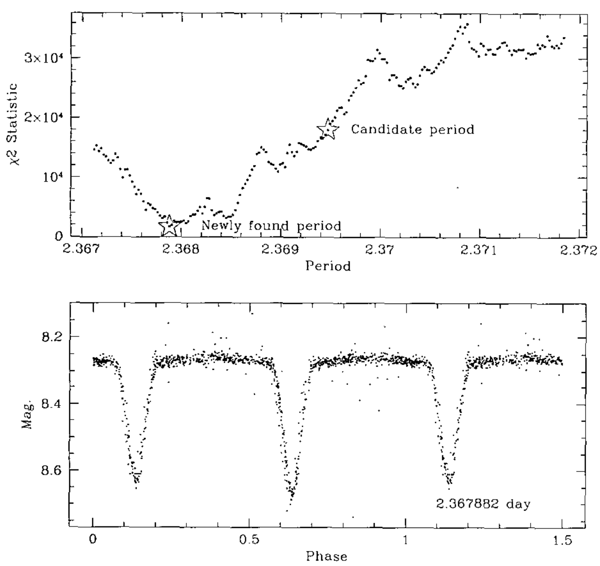

Fig. 4.— Result of cubic spline method applied to the case of Fig. 2. Upper panel shows the locations of candidate periods in χ^2 space, while the bottom panel shows the light curved phased with newly found period.

Fig. 4 shows the location of input candidate and newly found period in χ^2 space.

IV. EXPERIMENT WITH ASAS VARIABLES

We experiment here the accuracy and efficiency of our multi-step approach by applying it to many periodic variable stars. For this test, we chose the ASAS2 project database (*http://www.astrouw.edu.pl/~gp/asas/asas.html*), which contains 381 variable stars with a large number of photometric measurements ranging from a few hundreds to well over a thousand.

The entire period finding process, based on our multi-step algorithm, took only about 6 hours on a Pentium 4 1GHz Linux machine. The routine ran successfully for about 90% of sample stars, but failed for the rest. Fig. 5 compares our periods with ASAS periods. From short periods of a few tens of a day to large periods of over hundreds of days, the match is quite good. The agreement with ASAS periods turns out to be excellent for about 76% cases. However, we note that there are stars for which our solution differs significantly from ASAS periods. Including these, there are about 14% cases that called for closer examination. Our inspection of phased light curves indicates that about half of these corresponds to cases where it is difficult to tell whether one solution is superior than the other. A large fraction of them are variables whose amplitudes are very weak compared to the photometric

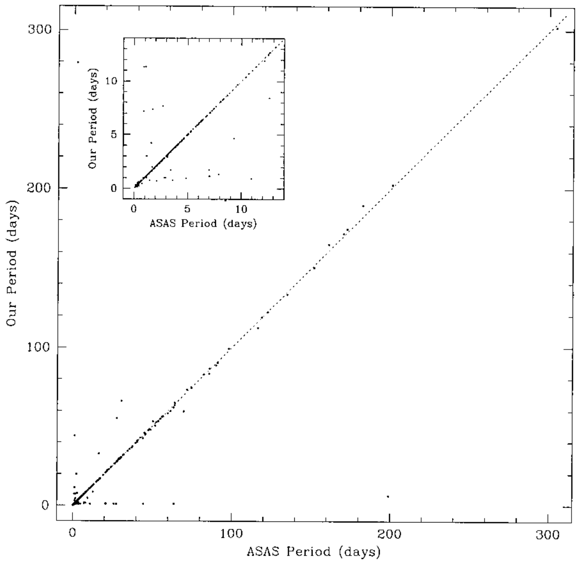

Fig. 5.— Our periods compared with those estimated by ASAS2. The dotted line represents the line of unity. Most of our results are in good agreement with theirs, however there exist cases of large discrepancies.

noise.

The rest of discrepancies can be divided into three groups. The first group, which is the majority, is where our solutions are considered better than ASAS periods. Some examples are illustrated in Fig. 6. The individual ASAS phased light curves may appear reasonable, but when compared with our light curves, it can be seen that they suffer serious problem of alias effect. This results in great differences in estimated periods. For stars 060013+0046.9, 103753-5140.1, 174611+0043.8, and 184435-0049.4, ASAS predicts relatively long periods of a few tens of days. Their light variation however can be quite well explained by much shorter periods of around 1 day or less. The eclipsing binary nature of stars 052832-6836.2, 065128-0122.3, and 113916-6026.1, and also the pulsation nature of the star 075021-0114.6 are very clearly seen in our phased light curves. This is all due to the refined search of cubic spline fitting. Even without considering the nature of variability, the superiority of our solutions are easily noticed by the apparent reduction in the dispersion of the light curve and also the uniformity of phased data distribution.

There are however second group of stars for which our solutions are not as good as ASAS. While this is a minor group compared to the opposite case, it suggests that our methods still need to be improved. Examples are given in Fig. 7. All of them have period less than 0.5 days and the discrepancy is relatively small. Nevertheless the resulting phased light curves indicate our solution got trapped in some local minimum and failed to proceed toward global minimum some short distance away. This is most likely caused by incomplete initial candidates given by our first search step, i.e. function fitting and PDM methods. Adjustment for the frequency range and intervals during this step is likely to be the correct way to solve this problem.

There exists a third group of stars. Both ASAS and our solutions produce reasonable looking phased light curves, but very different values for the period. For these stars, we need to know or make an assumption about the nature of their variability before we can choose either ASAS period or our period. Examples are shown in Fig. 8. In the cases of stars 135340-3036.0 and 220248-1218.7, ASAS considers eclipsing natures while our solutions are for pulsations. Without multicolor observations, it would be difficult to tell which nature, and therefore which period is more appropriate. Single band light curve alone is simply not sufficient to distinguish EW binary system from sinusoidal pulsation. This limitation affects both automatic and manual period determination. On the other hand, star 185821-4040.8 is obviously an eclipsing binary. With the extended flat regions in its light curve, it is likely to be a detached or semi-detached eclipsing binary. This star may indeed have primary and secondary minima of identical depth (as indicated by ASAS solution), or may be rather much weaker secondary minimum not readily shown in the present data (as preferred by our solution). Considering that ASAS is an I-band survey, the latter explanation is unlikely; nevertheless color data will provide a more definite answer. Finally, the star 104528-5321.9 is regarded as a long period variable by both ASAS and our solutions. Raw data indicate that there is a hundred day seasonal gap during which the observations were not made for this star. While ASAS solution tries to form a well connected light curve with period of 95 days, our solution simply leaves the gap as it is and suggests a period more than twice longer. Whether this is a long period variable of ASAS period or our period, observations covering much longer term will need to be made.

The experiment shows that our routine is robust enough to produce accurate periods for variety of variable stars. Our method use both AoV statistics and χ^2 to choose the best candidate period. This approach is certainly much faster than any other methods involving human judgement. Yet, we cannot claim that its result is always accurate. In a few cases, our results are not convincing and clearly worse than other published periods. Also, there are about 10% cases where our method failed to operate. We suspect that the present problem lies with the first step of finding candidate periods. By simply increasing the number of allowed initial candidates from 10 to 20, we find that the failure cases decrease to 8%. However we believe that more fundamental improvement is needed to handle all the problems found in this experiment. One obvious way is to introduce more fine and dynamically changing frequency intervals during our first step procedures. We should

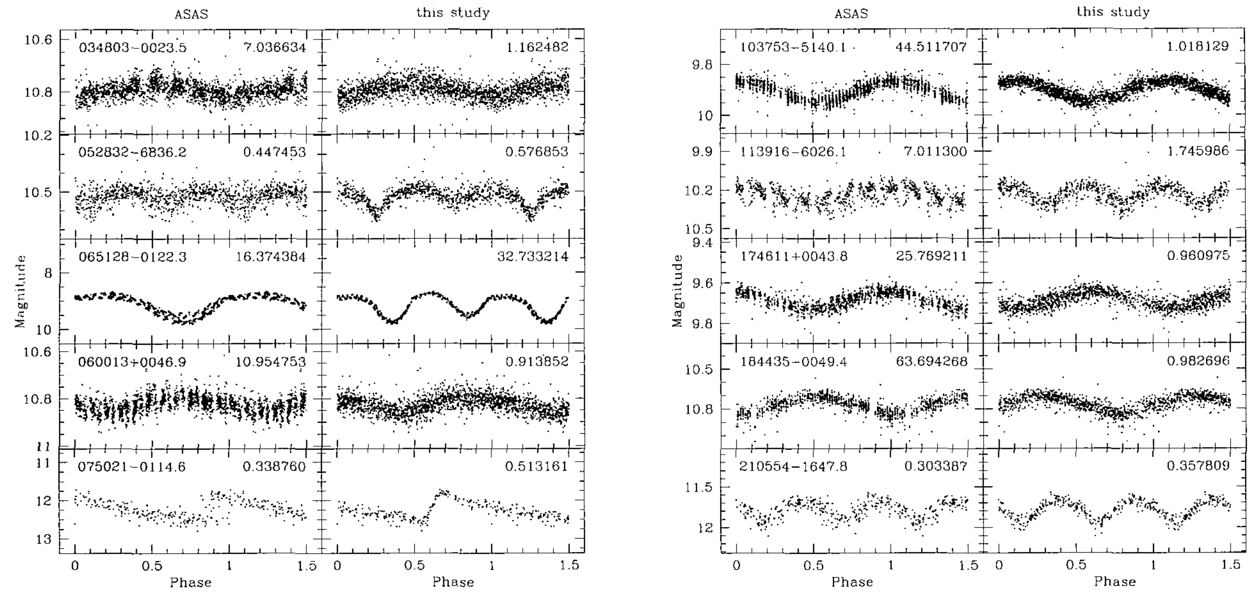

Fig. 6.— Stars for which our solutions are considered superior than ASAS. In each panel, ASAS star ID and periods are given. In many cases, ASAS solutions are subjected to the alias effect of 1 cycle/day.

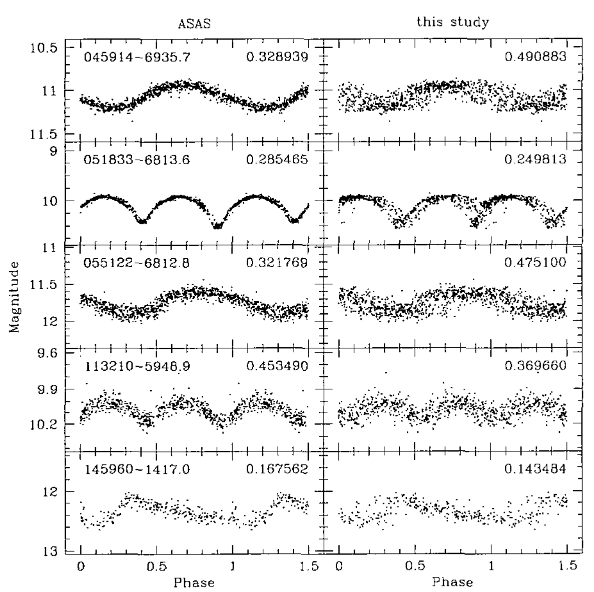

Fig. 7.— Stars for which our solutions are inferior to ASAS.

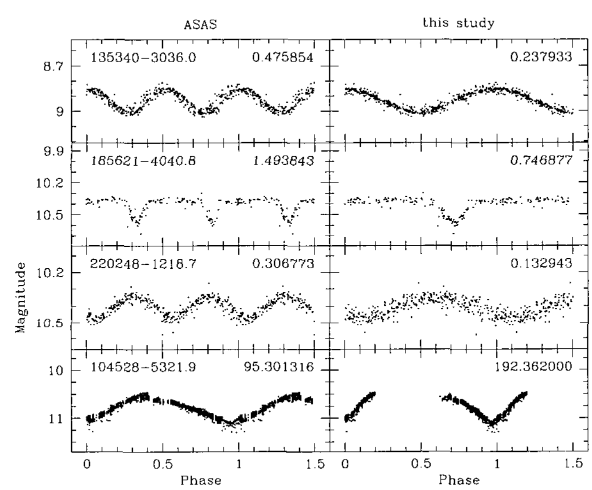

Fig. 8.— Stars for which we need to know its variability nature before we choose between our solution and ASAS period.

also consider frequencies higher than shortest interval between observations, because when large number of data is gathered from long term survey observations, the shortest possible period one can detect does not relate to the sampling intervals anymore.

V. SUMMARY

We developed an efficient algorithm which combines function fitting, PDM, and cubic spline methods. The final step of cubic spline analysis acts not only as a period refiner but also as an additional check for the candidate solutions suggested by function fitting and PDM methods. By applying our algorithm to a large data set of ASAS, we find that in most cases our algorithm provide periods as good as or even better than the original catalogued periods. However, there still are cases where our routine fails or produces periods worse than published values. We are presently modifying our algorithm and have begun another experiment with about 1700 variable stars in the ROTSE database (Akerlof et al. 2000).

This is a part of our effort toward a fully automated determination of periods. There will be a limit in the automation. Statistical inference is not everything, and we will need to include some empirical knowledge into the algorithm to distinguish a certain type of eclipsing variables from a certain type of pulsating ones. There are cases, as also shown in this paper, for which the single band data alone cannot make a conclusion. Except these particular cases, however, it should be possible to calculate the most plausible period and also to classify the nature of each variable star in the process of period analysis.

The need for automation is clear. There are too much data that a digital survey can produce. YSTAR project monitors many millions of stars, and several hundreds of thousands new variables will need to be analyzed. It should be noted that an automated period search routine such as the one developed here will be useful not only for the study of variable sources, but also for the initial identification of periodic variables from the vast amount of photometric data.

ACKNOWLEDGEMENTS

We are grateful for the constructive comments made by the referee, Dr. S. -L. Kim. This work was supported by Korea Research Foundation Grant (KRF-2002-070-C00045).

REFERENCES

Akerlof, C. et al. 1994, Application of cubic splines to the spectral analysis of unequally spaced data, ApJ, 436, 787

Akerlof, C. et al. 2000, ROTSE All-Sky Surveys for Variable Stars. I. Test Fields, AJ, 119, 1901

Byun, Y. I., Han, W.-Y., & Kang, Y.-W. et al. 2001, in IAU Colloq. 183, ed. B. Paczynski, W.-P. Chen, & C. Lemme, Small Telescopes Astronomy on Global Scales (ASP Conf. Ser. 246; San Francisco: ASP), 83

Carbonell, M. & Ballester, J. L., 1992, Power spectra of gapped time series - A comparison of several methods, A&A, 264, 350

Pojmanski, G., 1998, The All Sky Automated Survey. The Catalog of the Periodic Variable Stars in the Selected Fields, AcA, 48, 35

Pojmanski, G., 2000, The All Sky Automated Survey. Catalog of about 3800 Variable Stars, AcA, 50, 177

Scargle, J. D. 1982, Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data, ApJ, 263, 835

Schwarzenberg-Czerny, A. 1996, Fast and Statistically Optimal Period Search in Uneven Sampled Observations, ApJ, 460, L107

Schwarzenberg-Czerny, A. 1997, The Correct Probability Distribution for the Phase Dispersion Minimization Periodogram, ApJ, 489,941

Schwarzenberg-Czerny, A. 1998, Period Search in Large Datasets, Baltic Astronomy, 7, 43

Schwarzenberg-Czerny, A. 1999, Optimum Period Search: Quantitative Analysis, ApJ, 516, 315

Stellingwerf, R. F. 1978, Period determination using phase dispersion minimization, ApJ, 224, 953

Sterken, C. & Jaschek, C. 1996, Light Curves of Variable Stars (Cambridge: Cambridge Univ. Press)

Udalski, A., Kubiak, M., Szymanski, M., Kaluzny, J., Mateo, M., & Krzeminski, W., 1994, The Optical Gravitational Lensing Experiment. The Catalog of Periodic Variable Stars in the Galactic Bulge. I. Periodic Variables in the Center of the Baade's Window, AcA, 44, 317