

AUTOMATIC CLASSIFICATION OF OBJECTIVE PRISM SPECTRA

R. M. WEST

*European Southern Observatory, Hamburg, Germany**

Abstract. Methodological problems and hardware/software requirements of efficient automatic classification of objective prism stellar spectra are described on the basis of experience from 'microphotometer-computer' classification. Significant advances in hardware (two-dimensional scanners, on-line computers) as well as in software (image processing techniques) have now brought high-speed, exhaustive, automatic classification within reach. Some astronomical implications of automatic spectral surveys are discussed.

1. Introduction

There is hardly any field of modern astronomy in which digital equipment has not yet been successfully applied. Many laborious tasks which are unavoidable in astronomical work and which formerly occupied significant amounts of an astronomer's time, are now performed by various kinds of machines. It is quite natural that first of all those straightforward problems have been automated that only to a lesser degree require the astronomer's skill; those problems that demand a greater degree of versatility and personal judgment are still, to a large extent, solved by the astronomer's manual efforts.

The classification of stellar spectra is a fundamental astronomical function that has not yet been automated. Anyone who has tried to classify spectra visually, or who has had the opportunity to acquaint himself with the working methods of the great visual classifiers, must admit that spectral classification requires nothing less than an artist's approach. The whole method of spectral classification, and especially the visual application of spectral criteria to objective prism spectra, is a very delicate operation for which a successful result can only be reached by great skill and patience.

In view of the many inherent difficulties one might therefore ask if it justified to investigate the possibilities of automating spectral classification. However, attempts to increase the objectivity of the classification by means of microphotometer registrations have definitely demonstrated the advantage of *partial automation* in this field.

Automation is always applied most efficiently where large masses of data are treated in a not necessarily simple, but at least monotonous way. Within the field of classification of stellar spectra, objective prism plates with hundreds or even thousands of spectra constitute an obvious object for automation efforts. The usefulness of quantitative, two-dimensional interpretation of objective prism spectra has been demonstrated repeatedly. Moreover, investigations by McCarthy *et al.* (1966), Samson (1969) and West (1970, 1971) have shown that quantitative three-dimensional classification of late-type objective prism spectra is possible, even at rather low dispersion. The magnificent spectra of the Curtis Schmidt telescope on Cerro Tololo prove that objective prism plates that are taken under excellent atmospheric conditions also yield

* Now at ESO Sky Atlas Laboratory, Geneva.

excellent results; thus further adding to the increasing importance of objective prism work in modern astronomy.

With this in mind, we shall therefore now consider the possibility of *complete automation* of spectral classification, applied to objective prism spectral plates. The word 'complete' here signifies the automatic assignment of parameter values to stellar spectra on a photographic plate by means of an apparatus (an 'automatic classifier') into which the plate is inserted.

Let us for a moment forget about the vast difficulties which must be overcome before an automatic classifier can be put into operation and consider the major advantages that would then become available (not listed in order of relative importance):

- Position determination and automatic identification of the stars.
- Magnitude and colour determination by means of appropriate mathematical filters.
- Possibility of determination of coarse radial velocities from plate sets with prism in normal and reversed position.
- Direct estimate of *basic* stellar parameters (e.g. T_e , g , abundances).
- Possibility of classification of higher dimension.
- High speed and great endurance, i.e. many spectra can be classified within a reasonable time.
- Homogeneity, elimination of personal errors.
- Accuracy.
- Detection of spectral variability.
- Possibility of rapid and easy re-classification, e.g. if the calibration of the stellar parameters changes or when new spectral criteria become available.

In what follows, an outline is given of the methods (Section 2) and the means (Section 3) that are now available for high-speed, exhaustive, automatic classification of objective prism plates. The present state of automatic classification and some of the problems that still have to be solved are summarized in Section 4, and possible astronomical applications of future, large-scale spectral surveys are discussed in Section 5.

2. The Methods

Automatic classification of objective prism spectra can be accomplished in several ways within the following general scheme:

2.1. REGISTRATION

The spectral plate is scanned and the transmission is measured in a network of points.

2.2. RECOGNITION

A spectrum is identified as a plate area with reduced transmission. The measured transmission values within the spectrum are arranged in a predetermined way, e.g. according to wavelength.

2.3. TRANSFORMATION

The spectrum is transformed in some way or another in order to have it in a conve-

nient form for further reduction, e.g. to intensity units by means of the photographic calibration curve; Fourier- or Hadamard-transformation; convolution with suitable function, etc. In what follows the transformed spectrum will be referred to as the *image* of the spectrum.

Two different procedures are possible at this step:

2.4a. CRITERIA EVALUATION

Certain numerical or logical values (criteria) are computed which characterize the image of the spectrum in an unambiguous way, e.g. line depth ratios, equivalent widths, amplitudes in power spectrum, etc.

2.4b. CROSS CORRELATION

The image of the spectrum is compared with the images of stored standard spectra by cross correlation techniques. By iteration those standard spectra are found that surround the spectrum in the n -dimensional classification space.

2.5. CLASSIFICATION

If step 2.4a was performed, then the values of the stellar parameters are found from the computed values of the criteria and a certain class is assigned to the spectrum. Following step 2.4b, however, an interpolation in the classification space by means of the appropriate correlation measures leads to the spatial position of the image of the spectrum and thus the class. Peculiarities, e.g. unusual features in the spectrum, spectroscopic binaries, etc., are detected here or possibly at an earlier stage.

It follows from this scheme that a large number of standard spectra (or synthetic spectra of the relevant dispersion, if these should become available) for which the stellar parameters are known, must be secured. In case the cross correlation method is used (step 2.4b), the images of these spectra will be stored in the computer, possibly after some preliminary processing to ensure homogeneity.

When the criteria evaluation method (step 2.4a) is used, then the automatic classification must be preceded by the establishment of optimized spectral criteria that are evaluated from standard stars.

The accuracy of the automatic classification is readily estimated from automatic re-classification of standard spectra.

3. The Means

There has been a tremendous progress in image processing techniques during the past few years and the image types that can now be automatically interpreted range from cells in biological tissue to elementary particle tracks and all kinds of objects on aerial survey photographs. An extensive literature exists on this subject, and a series of excellent review papers is found in the Proceedings of a NATO Summer School (Graselli, 1969) and a Conference on Methodologies of Pattern Recognition (Watanabe, 1969). For a very recent, brief survey of the INTER/MICRO Exhibition in London

in September 1971 that almost exclusively dealt with the analysis of images by computer (see Fisher, 1971).

The automatic classification of objective prism spectra has much in common with other types of automatic image interpretation. Similar to features on aerial and satellite photographs the spectra may be *easily* identified when they stand out clearly from the plate background, but only *with difficulty* if they are 'camouflaged' on a variable background or by overlapping. The complexity of the software required for the classification is rather *high*, since the spectra are classified according to distribution of surface density and not to the size and shape alone as most other images that are now being automatically interpreted.

We shall now take a look on the hardware and software that will be necessary for the realization of automatic classification. In general, only qualitative facts will be given, since the ultimate speed and accuracy of an automatic classifier will depend critically on the software that controls the scanning device and performs the classification. Although hardware already exists that is very similar to what will be needed for automatic classification, no detailed study of the necessary amount and substructure of the software has so far been completed.

The hardware consists of two closely interacting parts: the scanning device and the on-line computer. If, for practical reasons such as speed and storage limitation, the entire classification procedure can not be carried out in real-time by the on-line computer, then a storage medium (tape- or disc-unit) must be provided for. From this medium the data are later transferred to another (larger) computer.

3.1. HARDWARE: THE SCANNING DEVICE

It is convenient to distinguish between *one-dimensional* and *two-dimensional* scanning on the basis of the freedom of the spot/slit during computer controlled movement. A one-dimensional device can, of course, simulate two-dimensional scanning by consecutive offsettings perpendicular to the scanning direction, but only at a vast expense in time. Considering the general advantages of two-dimensional scanning (that will be further discussed below), there is no doubt that this technique must be preferred in automatic classification of objective prism spectra.

The most important characteristics of the scanning device for the present purpose may be summarized as follows:

- Spot/slit dimensions.
- Spot/slit speed; plateholder speed.
- Positional accuracy and stability.
- Light source stability.
- Ability to keep the plate continuously and optimally focussed.

3.1.1. *One-Dimensional Scanning*

Digitized slit microphotometers are now in use at many observatories. So far they have mainly been used for providing the input material in computer analysis of high-dispersion spectra, but recently studies of automatic classification of objective prism

spectra have also been undertaken (cf. Section 4). The conventional microphotometer is a very stable device and the slit can be set to almost any size. A major drawback in this context is, however, the necessity of individual setting on each spectrum, including manual focussing. Moreover, most of the digitized microphotometers that have been put into operation have been somewhat limited in speed, partly because of slow output devices.

3.1.2. *Two-Dimensional Scanning*

Besides a few machines that have been specifically developed for astronomical purposes, a large number of automatic two-dimensional scanning devices for the measurement of particle tracks are now in operation. Among these, reference will here only be given to those that were developed at CERN (the European Organization for Nuclear Research) for scanning of bubble chamber tracks. These machines are, however, very representative for the whole class, as it can also be inferred from the Proceedings of the International Conference on Data Handling Systems in High Energy Physics (Lord and Powell, 1970). A brief survey of two-dimensional scanning devices has been given by Hambleton (1971).

One type of two-dimensional, 'flying spot' scanning device is based on a light spot that is moved across the plate by means of *digitally controlled, mechanically moved mirrors*. The light source may be *conventional* (US Naval Observatory Parallax Measuring Machine (cf. e.g. Strand, 1971); the HPD Machine at CERN (Benot *et al.*, 1968)) or a *laser* (Luyten Proper Motion Machine (Newcomb *et al.*, 1970); the Cambridge Machine (Kibblewhite, 1971)). This scanner type is fairly stable and the obtained speeds of the light spot are high; however, the complete digital control of the light spot is, where present, somewhat slowed by the inertia of the mirror mechanics. The laser devices offer the advantages of high spot intensity and small spot size.

In recent years, *Cathode Ray Tubes (CRT)* have been used to generate flying spots. Typical examples hereof are the GALAXY Machine (cf. e.g. Walker, 1971) and the LUCY Machine at CERN (Anders *et al.*, 1970). The major problems here are the stability of the light source, due to variations in the CRT phosphor screen, and unavoidable image aberrations. Furthermore, multiple reflections in the CRT screen may diffuse the spot and create problems in the astronomical application. On the other hand, the CRT technique allows a very high speed and certainly a great versatility. Comparison of the CERN machines HPD and LUCY seems to indicate that the CRT devices can now be made as stable as the 'moving mirror' devices, if geometric and photometric calibrations are performed carefully and regularly by means of standard grids and sensitometric scales.

The latest technical innovation which might prove to be of great use for scanning apparatus is the *Diode Array*. It consists of several hundred mini-diodes (until now manufactured with areas of 25×25 or $100 \times 100 \mu\text{m}^2$) in a one-dimensional array on the surface of an integrated block that also contains the circuitry for reading a digitized signal from the diodes. A scanning apparatus that moves a photographic plate per-

pendicularly to a Diode Array at high speed and thus scans the plate in a *band* has very recently been drafted by Høg (1971).

One of the major problems in fast two-dimensional scanning is to obtain a high light intensity level so that the statistical noise that is due to the corpuscular nature of light is reduced.

Further comments on the possibility of using modified versions of the machines that were mentioned above for the purpose of automatic spectral classification will be given in Section 4. Here it is sufficient to mention that the spot sizes that have so far been applied range from 8 μm (GALAXY) to 20–30 μm (older machines). However, if the need arises, the spot size may be demagnified further at the expense of the scanning speed.

The correct focus can be automatically maintained by continuously ascertaining maximum grain noise through a computer feed-back loop to a focus screw during the operation.

3.2. HARDWARE: THE ON-LINE COMPUTER

During the last years many small- to medium-size computers have become commercially available. The cost per word has gone considerably down, and the introduction of fourth-generation computers, based on large-scale integration, will undoubtedly mean new improvements in terms of speed, amount of storage space and versatility, at the same or even lower cost.

The computer that is connected on-line to the scanning device has a double function. It controls the scanning process (spot and/or plateholder position) and it performs at least a part of the classification. Whether the whole classification can be carried out in real-time is largely a question of software structure, computer storage space, cycling time and scanning speed. Before a detailed study has been made of the software it is very difficult to judge the total classification time, but something of the order of 1–2 s per spectrum will probably have to be expected if the cycling time is 1–2 μs .

If real-time, complete classification is not feasible, then an intermediate storage medium, such as a magnetic tape station or a disc file must be added. Paper tape or cards must be disregarded here because of excessive punching time. It seems very natural to let the on-line computer carry out at least the recognition of the spectrum on the plate in order to avoid the storage of large amounts of transmission data. At any step from then on the characteristic data of the spectrum can be stored for later evaluation with another (larger) computer, according to the greatest convenience in the data handling. The storage of all transmission readings (possibly transformed in some way) of a spectrum is advantageous from the re-classification point of view, but complete classification in real-time will save money and trouble in handling of the magnetic tapes or discs and probably also classification time.

3.3. SOFTWARE

The software consists of a program of algorithms that is stored in the on-line com

puter. It has the double function of controlling the scanning device and performing (part of) the classification. It is a great advantage to have these functions closely connected in order to perform the registration of the plate in the most efficient way. Especially useful is a software with time-sharing property, so that the registered transmission values of one spectrum are analyzed simultaneously with the registration of the next spectrum, etc.

In designing the software great care must be taken to ensure that it is fast, efficient and by all means complete at the same time. To reach this goal we can be guided to a large extent by the experience with circular images that has been gained during the use of the above mentioned two-dimensional machines and also by the recent developments in 'microphotometercomputer' classification (cf. Section 4).

Some further comments will now be given on the software functions that were outlined in Section 2.

3.3.1. *Registration – Recognition*

This is a crucial point in the classification procedure and the accomplishment of an efficient and unambiguous recognition of the spectra is of fundamental importance for the subsequent classification. Any spectral feature that is lost in this phase will increase the risk of faulty classification or even make the assignment of class to a spectrum impossible.

A spectrum may be provisionally identified as a region of lower transmission in the elsewhere almost transparent plate. This holds true in most cases, but there are a number of exceptions. First the plate may be highly exposed or the spectrum weak, so that the contrast is low. The plate background may also be variable, e.g. due to extended emission objects. Secondly, the plate may contain plate errors, such as bright or dark lines, images of dust specks, grain clusters or holes in the emulsion. And thirdly there is the very important question of overlapping of adjacent spectra.

The computer program guides the spot and/or the plateholder so that the idle time that is spent between the spectra is at minimum. However, registration of the plate background is necessary in the neighbourhood of each spectrum and serves also for focal adjustment by continuous monitoring and maximizing of the grain noise through feed-back. When the spot encounters a spectrum, it appears to be most economic to let it scan the spectrum repeatedly along the direction of dispersion, in small steps across the width of the spectrum. Overlapping spectra can be identified by the increased width and/or length of the combined image. The program separates the overlapping spectra as far as this is possible and the overlapped area may be utilized through decomposition, especially in spectra where the non-overlapped width is small. The separation is more difficult when the overlapping spectra only can be discerned because of increased (combined) length, but use must of course be made as much as possible of the extreme wavelength regions where the contamination is smallest. Very close stellar pairs must be treated as spectroscopic binaries and sought decomposed at a later stage in the analysis.

One way of overcoming some of the recognition problems is to use predetermined stellar positions that are obtained by scanning a direct plate in advance and storing the positions of the stars for which the spectra are later to be classified. This method is advantageous in several respects, especially for the safe recognition of stellar pairs that might otherwise simulate spectroscopic binaries during the spectral classification. It presupposes, on the other hand, that a direct plate is available and that the measured plate positions can be stored in the computer or in a peripheral unit and later made available at the right moment. It can probably best be judged after practical tests whether or not a direct plate should be registered before the spectral plate.

After registration and possible disentangling from neighbouring spectra, a spectrum is now stored in the computer as a two-dimensional array of transmission registrations. At this point it is possible to correct for some types of plate errors, e.g. lines that extend across the spectrum and thus can be recognized above the surrounding plate background. Smaller errors entirely within the spectrum or just at the border must still await detection.

In making the recognition of a spectrum complete, a rigid wavelength calibration is needed. Since the mean dispersion curve of the plate may be considered as known, it is first of all necessary to establish at least one fixpoint of known wavelength in the spectrum. This fixpoint can be the cut-off of the photographic emulsion or the telescope optics. It may also be some of the major lines in the spectrum, although this implies some kind of rough classification already at this point, which might be somewhat impractical. The aforementioned use of predetermined plate positions from a direct plate may solve this problem if the corrections between the direct and the spectral plate – caused by the addition of the objective prism to the optics – are sufficient accurately determined.

Following the preliminary wavelength calibration by means of one fixpoint and a mean dispersion curve, it now becomes possible to identify major spectral features so that an individual, rigid wavelength calibration can be established for each spectrum. It must be emphasized that the problem of a precise wavelength calibration is a very fundamental one; without a correct knowledge of the wavelengths, the classification criteria may be incorrectly evaluated due to faulty feature identification. Furthermore, the existence of peculiarities at certain wavelengths, e.g. Barium and Silicon lines, can not be convincingly ascertained.

It is in principle possible to perform the wavelength calibration by means of optimal fitting through cross correlation with a standard spectrum, similar to the Griffin (1967) method for radial velocity determination. However, this method implies that the spectral type is fairly well known and therefore necessitates a preliminary classification of the spectrum already at this stage.

By integrating over the spectrum, a measure of the total density is obtained from which the apparent magnitude can be computed if the plate has been photographically calibrated or contains a standard magnitude sequence. With the application of one or more mathematical filters, the magnitudes and colours can be computed in any photometric system within the sensitivity region of the plate emulsion. When the

photographic calibration curve is wavelength dependent, as is practically always the case, this must be taken into account during the integration.

The final output of the recognition part of the program consists of a two-dimensional array of plate transmissions that has been wavelength calibrated.

3.3.2. Transformation

In almost all cases where images are handled by a computer, it has been found very useful to let an image transformation follow after the recognition phase in order to facilitate the access to the important image features. Several monographs and articles are available on the subject of image transformation; e.g. two excellent introductions have been written by Rosenfeld (1969) and Andrews (1970). There are evidently many levels of complexity in this part of image processing, ranging from very simple to extremely sophisticated statistical-mathematical treatment of the image array. It is difficult to foresee exactly what kind of technique that will be the most advantageous to apply in our case, but in view of the limited time available for each spectrum, the highly advanced and thus very time consuming methods appear to be out of question. But we should not *a priori* discard the possibility of making at least partly use of some of the powerful tools of modern numerical analysis.

It is possible to reduce immediately the two-dimensional transmission array to one dimension by averaging or integrating the transmission readings at the same wavelength. This is what is implicitly done by the slit scanning microphotometer. However, notwithstanding the increased amount of computation involved, there are advantages in a further two-dimensional evaluation. This concerns especially the plate errors which do not extend across the whole width of the spectrum and the noise. The recorded noise in the spectrum depends on the intrinsic graininess of the plate, the spot size and, depending on the scanning principle, on internal noise in the scanning device. In the past, noise has been eliminated either by smoothing (averaging) or cutting of the high frequencies in the one-dimensional Fourier transform.

After high-speed, two-dimensional Fourier transformation has become available (cf. Special Issues of IEEE Trans. Audio Electroacoustics AU-15 (June 1967) and AU-17 (June 1969) on Fast Fourier Transforms), this specific kind of transformation has repeatedly been successfully applied in image processing. Some one-dimensional astronomical applications of the Fast Fourier Transform have been given by Brault and White (1971).

Ideally, the density at a certain wavelength should be the same across the spectrum, but it is seldom so because of noise and plate errors, and sometimes because of incorrect guiding of the telescope. It would therefore be very interesting to see two-dimensional Fourier transformation applied to objective prism spectra for noise smoothing and error detection in the inverse image of the spectrum. Other transformations, e.g. the Hadamard-transformation, are presently in much use in image processing, mostly in their two-dimensional forms. Their usefulness in the present context should also be investigated.

There are other ways of transforming the spectrum and simultaneously make the

elimination of noise possible, e.g. convolution or cross correlation with suitable functions (filtering).

A transformation from plate transmission to intensity can take place if the plate has been photographically calibrated. It is advantageous to use individual calibration curves for each wavelength, obtained by interpolation between the curves for each 50 or 100 Å.

After the spectrum has been transformed and possibly cleaned for noise and plate errors, the image of the spectrum can now be compressed to a one-dimensional image. Hopefully the plate errors for which no correction could be made are now known so that the affected spectral regions can be disregarded in the subsequent analysis.

3.3.3. *Criteria Evaluation*

This part of the program must be distinguished from the succeeding part, the classification, because it establishes inherent numerical or logical values of the spectrum. The classification part thereafter interpretes these values in terms of stellar parameters, but this interpretation, i.e. the calibration of the criteria, may change, for instance if the basis for the calibration becomes better.

Well defined spectral criteria exist at present for visual classification at various dispersions and in several wavelength regions. Criteria that have been established in the MK system and which play a role in the classification of objective prism spectra have been published in many places, but are best illustrated in the spectral atlases of Morgan, *et al.* (1943), Abt *et al.* (1968), Seitter (1970), and Landi Dessy *et al.* (1972). However, since automatic classification may not be optimally performed with the already existing criteria, and since new criteria are needed if the criteria values are to be derived from transformed spectra, it is necessary to establish (new) criteria independently. Moreover, the same set of criteria can only be applied to stellar spectra taken through the same optics and on the same photographic emulsion. New, optimal criteria must be determined whenever one of these factors is changed.

Before any automatic classification can be performed, it is therefore necessary that a large number of standard spectra with known values of the stellar parameters are processed until the stage of criteria evaluation. At this point the images of the spectra are thoroughly searched for optimal criteria. This may seem an enormous task in view of the many spectral features, but it can be very efficiently conducted by means of a computer. All possible criteria, such as line depths, line-depth ratios, peak heights, feature areas (e.g. equivalent widths) and absence/presence of specific features in the image of the spectrum can be tested for correlation with the stellar parameters, and those criteria that appear the most promising are selected for refinement. In this process the *accuracy* as well as the *general availability* of the criteria must be considered. There should preferably be two or more criteria for each spectral parameter so that any reasonable number of plate errors does not exclude the evaluation of at least one criterion for each parameter and thus the complete classification of any spectrum.

The possible automatic use of the present spectral criteria in visual classification should be investigated.

When the optimal criteria have been decided on, they are calibrated directly in terms of the stellar parameters.

It is possible only to use major criteria in the beginning of the evaluation of the criteria in order to save time. Depending on a preliminary 'coarse' assignment of spectral class, other, more refined criteria are thereafter activated for the 'fine' classification.

3.3.4. *Cross Correlation with Standard Spectra*

An alternative to the criteria evaluation is the cross correlation with standard spectra. This technique implies that a sufficient number of standard spectra is stored in the computer. The spectrum for which values of the stellar parameters shall be found is brought into the same format as the standard spectra; this involves a very accurate wavelength calibration and some standardizing transformation. As earlier mentioned the wavelength calibration may be part of the cross correlation procedure which in that case begins with a cross correlation with a standard spectrum of similar spectral type to obtain the best fit in wavelength. Thereafter the spectrum is correlated with a number of standard spectra in order to place it in the correct box in the n -dimensional classification space, n being the number of independent classification parameters.

It will be very interesting to see the cross correlation method applied to automatic classification of objective prism spectra. However, notwithstanding its many advantages, it might be feared that a rather large number of correlations must be performed, even in two-dimensional classification, which will be of disadvantage in computing time required. It is quite possible that a combination of the criteria evaluation and the cross correlation technique will be the most useful approach. By alternate use of spectral criteria and cross correlation in narrow spectral regions, the correct assignment of a stellar parameter may be achieved in a minimum of time.

3.3.5. *Classification*

As the last step the classification is now performed by assigning to each spectrum those values of the stellar parameters that correspond to the criteria values or, if the cross correlation method is used, by interpolation in the classification space. Peculiar features are looked for, and it is noted if there are significant discrepancies between the values of any stellar parameter when determined by means of different criteria.

It is possible that the classification procedure will have the form of an iteration, e.g. first the luminosity class is estimated from certain criteria and, depending on this class, other criteria give the temperature class, the knowledge of which in turn makes possible a more accurate determination of the luminosity class, etc.

The endproduct of all these computations is a list, giving the plate positions of the spectra, the magnitudes and colours and the stellar parameters with their internal errors. External errors may be computed if some of the spectra have been classified by other investigators. The plate positions can be utilized for automatic identification

of the stars by number or designation if an astrometric catalogue of the field is available on tape. When other spectral plates of the same field have been processed, mean values of the stellar parameters can be computed and stars with spectral variability detected. The intrinsic accuracy of the automatic classification method is determined through the comparison of the results obtained from plates of the same field.

By analysis of spectral plates with the prism in normal and reversed (180° rotation) positions, coarse radial velocities may be determined from the positions of the lines (cf. Stock and Osborn, 1973).

4. The Present State of Automatic Classification

Apart from a few, isolated experiments with flying spot scanners, all experience with automatic classification of stellar spectra has so far been gained in one-dimensional 'microphotometer-computer' classification. This means that the recognition problem hardly has been touched upon, except for some experiments with automated wavelength calibration (West, 1971). Various aspects of the computer reduction of digitally registered slit spectra have been described by Bonsack (1970), Furenlid (1971), Gratton *et al.* (1971), Hutchings (1971), Hutchison (1971), Latham (1971), Leushin *et al.* (1970), Peat and Pemberton (1970), Thompson (1967, 1971a, 1971b), and Yoss and Lutz (1968). Automatic reduction of objective prism spectra is described by Cassatella *et al.* (1973), Clausen (1973), Gratton *et al.* (1971) and Yoss (1973). Mathematical discussions of the computer analysis of stellar spectra are found in the papers of Rusconi and Sedmark (1971) and Thompson (1971a). Radial velocities were determined by a computer from microphotometer scannings of late-type, slit spectra (Lutz and Yoss, 1969; Yoss, 1966). A search for optimal criteria in the ultraviolet part of objective prism spectra of late-type stars has been described by West (1970). Partially automated classification of early-type objective prism spectra with a digitized microphotometer has been performed by Rossen (1969) and of late-type spectra by West (1971). Unpublished work on automatic reduction of objective prism spectra has been reported by Th. Schmidt-Kaler and collaborators in Bochum, by B. Virdefors in Lund and F. Simien in Marseille.

Obviously quite an amount of work is presently being invested in *partially* automated reduction of stellar spectra by means of digitized microphotometers. With the advantages of two-dimensional scanning in mind, let us now discuss some of the major problems that still have to be solved before *complete* automation be realized.

Some of the scanning devices that are available for automatic processing of circular images have been referred to above. A few of these regularly measure direct astronomical plates and much of the experience can undoubtedly be directly applied to the problem of automatic stellar classification. However, there are still important modifications that must be made.

Among the necessary changes in the *hardware*, one of the most significant is the digitization of the plate transmission, i.e. of the recorded spot intensities. Until now, most devices only discriminate between 'clear' and 'dark' plate areas (1 bit digitization)

since this is fully adequate in the registration of circular images for which only the position and size are important. An exception herefrom is the GALAXY Machine in the measuring mode.

The digitization will slow down the scanning speed for two reasons. First, the illumination time of each image element must be increased to reduce correspondingly the statistical noise due to the corpuscular nature of light, and secondly, more computing time is needed for the discrimination of the light levels. To retain the photographic accuracy, a transmission digitization with at least 8 bits (0–255) or preferably 10 bits (0–1023) is needed in spectral work. The intrinsic digitization of the Diode Array makes this solution a very attractive one, and it should certainly be investigated in detail.

The digitization also sets higher demands for the photometric stability of the device and some convenient means for accurate and fast calibration must be available. When used for identification only, the positional accuracy over the whole plate may not be so critical, but a determination of radial velocities calls for a very high accuracy that may only be accomplished over small plate areas.

The memory of the *on-line computer* should preferably be larger than the presently normal size (8–16 Kwords) for minicomputers because of the storage demand for the transmission readings and the program, but this is merely a cost problem.

The *software* for automatic classification is much more complicated than that for circular image processing. This means that, even when the necessary hardware becomes available, it will still take some time and certainly much experimentation before the automatic classification program will run satisfactorily.

The speeds that have already been attained range from about 7.5 h (GALAXY in search phase, 16 μm spot) to 6.5 min (LUCY, spot size 13 μm , spot speed $5\mu\text{m } \mu\text{s}^{-1}$, line distance 5 μm) for the registration of 100 cm^2 . Whereas the accuracy of the GALAXY measurements is high enough for astronomical purposes, this is not the case for LUCY.

TABLE I

Some numerical values for the automatic classification of objective prism spectra

Size of spectra	200 × 5000 μm^2
Distance between parallel scans	20 μm , i.e. 10 scans along a spectrum
Distance between transmission readings	2 μm
Computer cycling time	1 μs
Spot speed	0.2 $\mu\text{m } \mu\text{s}^{-1}$
Background registration, focussing, search for and registration of the spectrum	~ 0.1–0.2 s
Scanning time per spectrum	0.25 s
Classification time (criteria evaluation)	~ 0.6 s
Total time per spectrum	~ 1 s

Without prior knowledge of the amount and detailed substructure of the software and the scanning speed that will ultimately be adopted for the automatic classification of objective prism spectra, the numerical example in Table I gives only an indication of the obtainable speed.

The estimate of the classification time is based on a extrapolation of the figures in West (1971).

5. Some Implications of Automatic Spectral Classification

On the preceding pages we have discussed the possibility of complete automation of the classification of objective prism spectra. With its many advantages it must be foreseen as a future, very powerful tool in astronomical research. It is appropriate to add some words about the possible applications of an automatic classifier.

Clearly, the most obvious astronomical program that can be carried out with an automatic classifier is a *large-scale investigation of the galactic structure*. The automatic classifier would be an excellent mean for producing a large, very homogeneous material of spectral classes which would in turn be invaluable in the study of the spatial distribution of various stellar types. In this connection numerous other important problems could also be investigated; among these one might mention the determination of the luminosity functions of different population types, the peculiarity frequencies, spectral variability, discovery of spectroscopic binaries, etc. Of special importance is the possibility of direct calibration to basic stellar parameters that is offered by the automatic classification techniques.

It is here relevant to emphasize that the performance of an automatic classifier completely depends on the accumulated knowledge in the field of stellar classification and that an astronomically useful automatic classification is only possible in continued, close contact with the astronomers that are actively engaged in conventional spectral classification. In other words, the automatic classifier can best be regarded as an extremely efficient means of accomplishing what could otherwise only have been carried out in a tremendous effort of astronomers with normal equipment.

Let us briefly illustrate with an example the amount of information that could be processed with an automatic classifier. The 1 m ESO Schmidt telescope that will be in operation at Cerro La Silla in Chile from the beginning of 1972 will be equipped with an ultraviolet transparent prism, giving a dispersion of about 580 \AA mm^{-1} at H γ .

With a plate limiting magnitude of 14^m (a conservative estimate on IIa-O plates, 200 μm widening), about 11 000 stars may be expected on a $29 \times 29 \text{ cm}^2$ ($5.5 \times 5.5 \text{ deg}^2$) plate at galactic latitude 0° (1300 stars at 90°), cf. Seares *et al.* (1925). Neglecting overexposure and overlapping, the classification time would probably amount to several months (a few weeks) for a visual classifier, who classifies at a rate of 40 spectra h^{-1} and with time allowed for checking purposes. The corresponding time would be a few hours or less for an automatic classifier at 60 spectra min^{-1} , and there would be no need for checking. As a matter of fact, several plates of the same field could be processed on the same day in order to eliminate gross errors and to improve the accuracy.

A final problem that arises, concerns the storage of all the information that is obtained from the automatic classifier. Should lists be printed with the spectral classification and how shall the stars be identified? How can such a large material best be made available to the astronomical community? One might doubt whether there is any reason to publish extensive lists of spectral classes that are mainly of statistical value; it would most certainly be more economic and practical to make the original plates and the computer print-out available on request.

The Schmidt telescope has a very low ratio between observing and evaluation time. The spectral plates that were mentioned above will be obtained in 20–30 min. If such plates could be processed and prepared for astronomical interpretation of the spectral classes found during some hours, then a much more reasonable ratio would be obtained.

Experience in other fields of astronomy and in other sciences as well has since long proved that the introduction of automatic methods not only facilitates the current research; it also opens up entirely new possibilities. The classification of objective prism spectra will be no exception from this principle.

References

- Abt, H. A., Meinel, A. B., Morgan, W. W., and Tabscott, J. W.: 1968, *An Atlas of Low-Dispersion Grating Stellar Spectra*, Kitt Peak National Observatory and Steward Observatory, Tucson.
- Anders, H., Jacobs, D., Krishner, W., Lewis, A., Lingjaerde, T., Lord, D., Oropesa, J., Van Prag, A., Stumpe, B., Turek, L., Wiscott, D., and Zurbrechen, R.: 1970, in D. H. Lord and B. W. Powell (eds.), *Proc. of the International Conference on Data Handling Systems in High Energy Physics*, Cambridge, March 23–25, 1970, CERN Document 70–21, Geneva, p. 427.
- Andrews, H. C.: 1970, *Computer Techniques in Image Processing*, Academic Press, New York and London.
- Benot, M., Evershed, B. W., Messerli, R., and Powell, B. W.: 1968, *The HPD Mark 2 Flying-Spot Digitizer at CERN*, CERN Document 68–4, Geneva.
- Bonsack, W. K.: 1970, *Bull. Am. Astron. Soc.* **2**, 298.
- Brault, J. W. and White, O. R.: 1971, *Astron. Astrophys.* **13**, 169.
- Cassatella, A., Maffei, P., and Viotti, R.: 1973, this volume, p. 127.
- Clausen, J. V.: 1973, this volume, p. 134.
- Fisher, C.: 1971, in *New Sci. and Sci. J.*, September 23, 1971, p. 676.
- Furenlid, I.: 1971, *Astron. Astrophys.* **10**, 321.
- Graselli, A. (ed.): 1969, in *Automatic Interpretation and Classification of Images*, Proceedings of the NATO Summer School, Pisa-Terrenia, Italy, August 26–September 6, 1968, Academic Press, New York and London.
- Gratton, L., Martini, A., Martino, F., Natali, G., and Viotti, R.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 142.
- Griffin, R. F.: 1967, *Astrophys. J.* **148**, 465.
- Hambleton, J.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 157.
- Høg, E.: 1971, personal communication.
- Hutchings, J. B.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 185.
- Hutchison, R. B.: 1971, *Astron. J.* **76**, 711.
- Kibblewhite, E. J.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 122.
- Landi Dessy, J., Jaschek, C., and Jaschek, M.: 1972, *A Spectral Atlas at 42 Å mm⁻¹*, Cordoba, Argentine, to be published.
- Latham, D. W.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 183.
- Leushin, W. W., Jacomo, A. A., and Rusak, N. P.: 1970, *Izv. Krym. Astrofiz. Observ.* **41–42**, 384.
- Lord, D. H. and Powell, B. W. (eds.): 1970, *Proc. of the International Conference on Data Handling Systems in High Energy Physics*, Cambridge, March 23–25, 1970, CERN Document 70–21, Geneva.

- Lutz, T. E. and Yoss, K. M.: 1969, *Astron. J.* **74**, 91.
- McCarthy, M. F., Treanor, P. J., Bertiau, S. J., and Bertiau, F. C.: 1966, in K. Lodén, L. O. Lodén, and U. Sinnerstad (eds.), 'Spectral Classification and Multicolour Photometry', *IAU Symp.* **24**, 59.
- Morgan, W. W., Keenan, P. C., and Kellman, E.: 1943, *An Atlas of Stellar Spectra with an Outline of Spectral Classification*, Univ. Chicago Press, Chicago.
- Newcomb, J., La Bonte, A., and Luyten, W. J.: 1970, *The Automated Plate Scanner and Measuring Machine*, University of Minnesota, Minneapolis, Minnesota.
- Peat, D. W. and Pemberton, A. C.: 1970, *Observatory* **90**, 141.
- Rosenfeld, A.: 1969, *Picture Processing by Computer*, Academic Press, New York and London.
- Rossen, J.: 1969, Thesis, Copenhagen University Observatory.
- Rusconi, L. and Sedmark, G.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 188.
- Samson, W. B.: 1969, *Publ. Roy. Obs. Edinburgh* **6**, No. 10.
- Seares, F. H., Van Rhijn, P. J., Joyner, M. C., and Richmond, M. L.: 1925, Contributions Mt. Wilson Obs. No. 301.
- Seitter, W. C.: 1970, *Atlas for Objective Prism Spectra – Bonner Spektral-Atlas I*, Ferd. Dümmlers Verlag Bonn.
- Stock, J. and Osborn, W.: 1973, this volume, p. 290.
- Strand, K. Aa.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 119.
- Thompson, G. I.: 1967, *Publ. Roy. Obs. Edinburgh* **5**, No. 12.
- Thompson, G. I.: 1971a, *Publ. Roy. Obs. Edinburgh* **7**, No. 2.
- Thompson, G. I.: 1971b, *Publ. Roy. Obs. Edinburgh* **8**, 179.
- Walker, G. S.: 1971, *Publ. Roy. Obs. Edinburgh* **8**, 103.
- Watanabe, S. (Editor): 1969, in *Methodologies of Pattern Recognition*, Proceedings of the International Conference on Methodologies of Pattern Recognition, Honolulu, Hawaii, January 24–26, 1968, Academic Press, New York and London.
- West, R. M.: 1970, *Bull. Abastumani Astrophys. Observ.* **39**, 29.
- West, R. M.: 1971, *Bull. Abastumani Astrophys. Observ.* **42**, 109.
- Yoss, K. M.: 1966, in K. Lodén, L. O. Lodén, and U. Sinnerstad (eds.), 'Spectral Classification and Multicolour Photometry', *IAU Symp.* **24**, 111.
- Yoss, K. M.: 1973, this volume, p. 125.
- Yoss, K. M. and Lutz, T. E.: 1968, *Astron. J.* **73**, S41.