EMPIRICAL MODELING OF THE STELLAR SPECTRUM OF GALAXIES

CHENG LI, TING-GUI WANG, HONG-YAN ZHOU, XIAO-BO DONG, AND FU-ZHEN CHENG

Center for Astrophysics, University of Science and Technology of China, Hefei 230026, China; leech@ustc.edu.cn

Received 2004 June 30; accepted 2004 October 20

ABSTRACT

An empirical method of modeling the stellar spectrum of galaxies is proposed, based on two successive applications of principal component analysis (PCA). PCA is first applied to the newly available stellar library STELIB, supplemented by the $J$, $H$, and $K_S$ magnitudes taken mainly from the Two Micron All Sky Survey. Next, the resulting eigenspectra are used to fit the observed spectra of a sample of 1016 galaxies selected from the Sloan Digital Sky Survey Data Release 1 (SDSS DR1). PCA is again applied to the fitted spectra to construct the eigenspectra of galaxies with zero velocity dispersion. The first nine galactic eigenspectra so obtained are then used to model the stellar spectrum of the galaxies in SDSS DR1 and synchronously to estimate the stellar velocity dispersion, the spectral type, the near-infrared spectral energy distribution, and the average reddening. Extensive tests show that the spectra of different types of galaxies can be modeled quite accurately using these eigenspectra. The method can yield stellar velocity dispersion with accuracies better than 10% for the spectra of typical signal-to-noise ratios in SDSS DR1.

Key words: galaxies: fundamental parameters — galaxies: general — methods: data analysis — methods: statistical

1. INTRODUCTION

The observed spectrum of a galaxy is a combination of three components: a continuum, absorption lines, and emission lines. For nonactive galaxies, the continuum and the absorption-line components in the optical and near-infrared are usually dominated by starlight, whereas in the mid- and far-infrared they are dominated by dust. For active galaxies, the continuum is diluted by the featureless continuum of the nucleus. For active and nonactive galaxies, the continuum is modified to varying degrees by reddening. The emission-line component is produced in H II regions around hot stars or in the emission-line regions of the active nucleus. Proper decomposition of the emission-line component on one hand and the continuum plus absorption-line component (the so-called stellar spectrum) on the other is the first step toward a physical interpretation of the spectrum. For instance, accurate measurement of emission lines is fundamental to the identification and classification of active galactic nuclei (AGNs). However, the optical spectra of many nuclei are heavily contaminated, even dominated, by stellar absorption lines coming from the host galaxy. Therefore, it is imperative that the underlying starlight be properly removed before the AGN identification. As for studying nebular emission lines and the abundances of gaseous material, proper starlight removal is also necessary. Moreover, a well-modeled stellar spectrum can provide direct information on the stellar population, accurate measurement of the stellar velocity dispersion, and hence some fundamental properties of the host galaxy, such as spectral type, kinematics, star formation history, etc.

Although various schemes have been proposed to subtract the underlying starlight, the basic procedure is quite similar: a library of absorption-line templates is built in the first place, either from spectra of stars or from that of pure absorption-line galaxies; second, the templates are used to model the non–emission-line region of the spectrum; and finally the modeled spectrum, which is taken as the integrated spectrum of the stellar component, is removed.

When spectra of stars are employed to construct the templates, the most common approach is stellar population synthesis. There are two main types of population synthesis studies: evolutionary population synthesis (Tinsley 1967; Bruzual & Charlot 2003) and empirical population synthesis (Faber 1972; Cid Fernandes et al. 2001; Kong et al. 2003). For the former, the age- and metallicity-dependent models of stellar spectra are developed by assuming the time evolution of a few main parameters and then used to reconstruct the integrated spectrum of stellar systems. However, modern evolutionary population synthesis models still suffer from serious uncertainties, which appear to originate from the underlying stellar evolution theory, the color-temperature scale of giant stars, and, for non-solar abundances in particular, the flux libraries (Charlot et al. 1996). The empirical population synthesis approach reproduces the stellar spectrum with a linear combination of the spectra of a library of stars or star clusters. Compared with the evolutionary population synthesis, the merit of this approach is that the result does not depend on any assumed parameters, whereas the drawback is its strong dependence on the coverage range of metallicity, spectral type, and luminosity class of the stellar library. Another drawback is that this approach cannot predict the past and future stellar properties of galaxies.

When spectra of galaxies are used, the templates are usually derived either from the spectrum of another galaxy with no or weak emission lines or from that of an off-nuclear position in the same galaxy (Ho et al. 1997 and references therein). However, the spectrum of another galaxy may not accurately reflect the exact stellar content of the galaxy being studied, whereas the off-nuclear spectrum may not be completely free of line emission. Instead of using a single galactic spectrum as the template, which is often chosen subjectively, Ho et al. (1997) used an objective algorithm to find the best combination of galactic spectra to create an “effective” template. The advantage of this modification is that the use of a large basis of input spectra ensures a closer match to the true underlying stellar population. Recently, Hao & Strauss (2002) applied principal component analysis (PCA; Deeming 1964) to several hundred pure absorption-line galaxies and used the first few eigenspectra as the templates. This makes the size of the template library much smaller because the most prominent features from the sample
concentrate into the first few eigenspectra. The other merit is that the best-fitting model is unique because the eigenspectra are orthogonal. However, both of the above methods have their unavoidable shortcomings. Because the templates are derived from the observed spectra of galaxies, in which the stellar velocity dispersion is nonzero and varies from object to object, they usually do not match the velocity dispersion of the galaxies in question. Furthermore, the template library consists of only pure absorption-line galaxies, which usually contain very little or no young stellar component. As a result, the modeled spectra could not reflect correct information on the stellar population of emission-line galaxies, giving rise to inaccurate representative spectra.

We present here a robust and efficient method of starlight removal in the optical and near-infrared band. Briefly, PCA is first applied to the optical spectra of STELIB (Le Borgne et al. 2003), a newly available stellar library in the wavelength coverage of 3500–9500 Å, and to the near-infrared photometric data at $J$, $H$, and $K_s$ bands collected from the Two Micron All Sky Survey (2MASS; Skrutskie et al. 1997), supplemented by the stellar library presented by Pickles (1998) (§3). The stellar eigenspectra in the visible range are then used to model a homogeneous library of 1016 galactic spectra picked up from the Sloan Digital Sky Survey Data Release 1 (SDSS DR1; Abazajian et al. 2003). PCA is again applied to the modeled galactic spectra with zero stellar velocity dispersion. The first nine eigenspectra are selected as the final absorption-line templates (§4). Using these templates, the stellar spectrum of the galaxies in SDSS DR1 can be well modeled, and the stellar velocity dispersion, the near-infrared spectral energy distribution (SED), the spectral type, and the average reddening of galaxies can be obtained simultaneously (§5). Extensive tests show that the present method is self-consistent and robust.

2. PRINCIPAL COMPONENT ANALYSIS

The purpose of this section is to summarize briefly the PCA method and provide the definition of terms used in this paper. PCA is a method used to reveal interrelations among different variables and objects contained in a large, multivariate data set. Its aim is to reduce the number of dimensions in the data space so that the most important information can be extracted. Let the sample being studied be a collection of $n$ objects, for each of which there are $m$ observational variables. Thus, one has a matrix $X = \{ x_{ij} \}_{n \times m} (i = 1, \ldots, n$ and $j = 1, \ldots, m)$, each row vector of which corresponds to the different variables of a given object, whereas each column vector corresponds to the same variable of the various objects. PCA proceeds from the given matrix $X$ and yields $m$ new variables, the principal components (PCs), which are mutually independent, and generally the first $m'$ ($m' \ll m$) PCs contain a majority of the information of the data. Each PC is a linear combination of the original $m$ variables; the corresponding coefficient vector is called an eigenvector. Furthermore, via the eigenvectors, the original $m$ variables of each object are projected onto the PCs to yield the new variables of this object. For example, the $j$th PC of the $i$th object is given by

$$pc_{ij} = e_j \cdot x_i = e_{j1}x_{i1} + \ldots + e_{jk}x_{ik} + \ldots + e_{jm}x_{im},$$

(1)

where $x_i$ is the $i$th row vector of $X$ and $e_j$ is the eigenvector of the $j$th PC.

The fundamental principles of PCA could be understood as follows. The contribution of PCs to the variance of the original data set is a measure of the amount of original information contained in the PCs. Thus, the purpose of the PCA method is to seek the set of eigenvectors that give rise to the PCs with the maximum variance. According to the theory of statistics, the expected eigenvectors are essentially the orthogonal eigenvectors of the covariance matrix $C = \{ c_{jk} \}_{m \times m}$, where

$$c_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k); \quad 1 \leq j, k \leq m.$$

(2)

Accordingly, the covariance matrix $C$ is first constructed. Then the determinant equation $|C - \lambda I| = 0$ is solved to find the $m$ eigenvalues $\{ \lambda_i \} (l_1 \geq l_2 \geq l_3 \geq \ldots \geq l_m \geq 0)$, where $I$ is the unit $m \times m$ matrix. After this, by solving the equation $(C - \lambda_j I)e_j = 0$, the eigenvectors $e_j$ are obtained. Since there are $m$ eigenvalues, there will be at most $m$ eigenvectors. The PCs are ordered by decreasing the eigenvalues, which are generally used to characterize the contribution of the corresponding PCs to the original information. The quantity $l_j / \sum_{j=1}^{m} l_j$ is called the fractional or relative contribution of the $j$th PC, and $\sum_{j=1}^{k} l_j / \sum_{j=1}^{m} l_j$ is the cumulative contribution of the first $k$ PCs. In general, it is not necessary to find all the $m$ PCs; most of the information is contained in the first few, with the first one having the lion’s share.

Much has been written about the use of PCA in studies of the multivariate distribution of astronomical data (Connolly et al. 1995 and references therein). In most of the earlier studies, the variables of the original data set are observed (generally normalized) fluxes at $m$ wavelength channels of $n$ celestial bodies (objects), giving a matrix of $X = \{ x_{ij} \}_{n \times m}$; the resulting $m$ eigenvectors with $m$ wavelength channels for each eigenvector are so-called eigenspectra and are used for further applications: spectral classification, modeling of stellar spectra, etc. In this paper, we perform PCA in a slightly different way: the input data matrix is transposed to be $X = \{ x_{ij} \}_{m \times n}$ before being analyzed, i.e., the variables now become the fluxes of various celestial bodies at a given wavelength channel, and the objects are thus all the wavelength channels. In this case, PCA carries out an $n \times n$ matrix of eigenvectors, corresponding to at most $n$ PCs. Thus, the projection of the $n$ celestial bodies onto the PCs gives rise to $n$ new spectra, which denote $n$ “new” celestial bodies, called “eigenspectra” or “eigengalaxies.” We then use these spectra as our templates for spectral modeling.

The main difference between the above methods is easy to understand. In the former case, the eigenvectors (eigenspectra) correlate strongly with the prominent features present in the spectra of celestial bodies; thus, the projection of a celestial spectrum onto a PC gives a measure of the relative contribution of this celestial spectrum to the corresponding eigenspectrum. However, in the latter case, it is the eigenvectors that represent the contributions of the celestial bodies to the eigenspectra. Eigenspectra, whereas the spectra of the eigenspectra or eigen-galaxies contain various spectral features presented in original celestial spectra. We argue that, in some sense, the “matrix-transposed” method is much more effective and convenient. It is well known that a significant drawback of implementing PCA on large or very high dimensional data sets is the required computation time (Madgwick et al. 2003). For $n$ spectra, each with $m$ wavelengths, this requires $O(mn^2)$ operations. Therefore, if the data matrix is transposed, the operation will become $O(mn^2)$, which is much smaller in the case of $m \gg n$. Given that each

---

1 Hao & Strauss (2002) also noticed this problem, and a spectrum of an A star is added during their fitting.
spectrum contains $O(10^3)$ wavelength channels and the number of spectra is $O(10^2)$, the latter method would reduce the computing expense by a magnitude. Data matrix transposition goes without saying for the analysis of spectra of higher resolution.

For convenience, the spectra of eigenstars or eigengalaxies, i.e., the projection of the spectra of stars or galaxies onto PCs, are still called eigenspectra in the following text, although its meaning is different from that in the general case.

3. DERIVATION OF STELLAR EIGENSPECTRA

The stellar library is the foundation of applying the present method. It is conceivable that the quality of fits to observed galaxy spectra seriously hinges on the resolution of the stellar spectra, as well as the coverage of spectral type, luminosity class, chemical abundance, and wavelength range.

Theoretical stellar spectra are often preferred for spectral modeling because of their uniformity and generally more extensive coverage (e.g., Kurucz 1992; Lejeune et al. 1997, 1998; Westera et al. 2002). However, at the time of writing, no such library at the spectral resolution similar to or higher than that of SDSS spectra is available. Furthermore, these libraries may be biased in color and line strength, since many minor contributors to stellar opacity and to the emergent spectrum usually cannot all be included for the full spectral range because of computational constraints (Pickles 1998).

Several observed stellar libraries have been published, covering the ultraviolet (Heck et al. 1984; Robert et al. 1993; Walborn et al. 1995), optical (Gunn & Stryker 1983; Jacoby et al. 1984; Pickles 1985; Kiehling 1987), and near-infrared (Danks & Dennefeld 1994; Serote Roos et al. 1996) wavelength ranges. They are usually obtained with different instruments, at different resolution and spectral sampling and for different purposes (Pickles 1998). The most commonly used stellar library is HILIB, presented by Pickles (1998), which consists of 131 flux-calibrated spectra with complete wavelength range from 1150 to 10620 Å in steps of 5 Å. However, the metallicity coverage of this library is limited to be near solar abundance, and the spectral resolution is relatively lower for accurately spectral modeling.

High-resolution libraries spanning a wide range in metallicity, spectral type, and luminosity class have only recently become available. The STELIB library (Le Borgne et al. 2003), a new spectroscopic stellar library, consists of a homogeneous library of over 250 stellar spectra covering the wavelength range from 3200 to 9500 Å, with a resolution of $\leq 3$ Å (1 Å sampling) and a signal-to-noise ratio of $\sim 50$. The library includes stars of a range of metallicity from 0.05 to 2.5 times solar, a range of spectral type from O5 to M6 and luminosity class from I to V. Because of its wide coverage of spectral resolution and spectral type, this library represents a substantial improvement over previous stellar libraries (Le Borgne et al. 2003).

Although the visible region is always favored, the near-infrared is equally important because it reveals stars that remain hidden in the visible by interstellar dust (Combes et al. 2002). To extend the library into the near-infrared band, we extract the infrared photometric data from 2MASS, which has mapped the full sky at three near-infrared wavelengths with 10σ sensitivity limits of $J = 15.8$, $H = 15.1$, and $K_s = 14.3$ mag.

We therefore use STELIB as our star library. The optical spectra in STELIB and the corresponding near-infrared photometric data in 2MASS supplemented with HILIB are incorporated to form a star library of the highest quality to date.

3.1. Optical Data

The current public version of the STELIB library contains 255 optical spectra, which have been corrected for interstellar extinction and radial velocity. However, the data are missing in some limited wavelength range for nearly half of the stars. We use similar stellar spectra either from STELIB itself or from SDSS DR1 to fill the gaps of these spectra. Some examples of the result of this procedure are shown in Figure 1, and at last 204 spectra are yielded.

Although about 50 spectra that could not be satisfactorily filled were ignored, the resultant stellar library is still homogeneous enough. The distributions of spectral types of the initial 255 stars and that of the final 204 stars are shown in Figure 2. It is apparent that the coverage of spectral type of the resultant library is almost as good as that of the original.

There is no denying that, over previous stellar libraries, substantially improved as STELIB is, it still leaves much to be desired. As can be seen from Figure 2, there are obviously two blank regions: early O and late K, M, and L types. The absence of these types of stars may have some influence on modeling the stellar spectrum of galaxies, especially for those mainly consisting of late-type stars.

3.2. Near-Infrared Data

We search for counterparts of the 204 stars in the 2MASS Point Source Catalog. Out of them, 182 have been detected in 2MASS.

2 See http://webast.ast.obs-mip.fr/stelib/.
The magnitudes of these stars at the $J (1.235 \pm 0.006 \mu m)$, $H (1.662 \pm 0.009 \mu m)$, and $K_s (2.159 \pm 0.011 \mu m)$ bands are then transferred to fluxes $f_J$, $f_H$, and $f_{K_s}$ in units of ergs s$^{-1}$ cm$^{-2}$ Å$^{-1}$, according to the following formulas (see Cohen et al. 2003):

\[
\begin{align*}
  f_J &= 3.129 \times 10^{-0.4(J+25.0)}, \\
  f_H &= 1.133 \times 10^{-0.4(H+25.0)}, \\
  f_{K_s} &= 4.283 \times 10^{-0.4(K+27.5)}. 
\end{align*}
\]

As for the remaining 22 stars, the three near-infrared fluxes are derived from the similar spectra in HILIB. For each of the 22 optical spectra in STELIB, a cross-correlation method is performed to find the best-matched spectrum among the 131 spectra in HILIB, and the corresponding fluxes at the $J$, $H$, and $K_s$ bands are accepted to be the real values of the near-infrared data of this star.

3.3. PCA of Stellar Library

Before performing PCA, the 204 spectra in the visible range are trimmed to the common wavelength range of 3500–9500 Å and each spectrum, including the fluxes at the $J$, $H$, and $K_s$ bands, is normalized to unit strength ($\sum f_i^2 = 1$) (Connolly et al. 1995), where $f_i$ is the flux at wavelength $\lambda$. It is universally acknowledged that the set of orthogonal eigenvectors resulting from PCA is affected by the scaling of the data because the scaling may affect lines and continuum differently (Sodrê & Cuevas 1997). Since the data here are spectra of stars or modeled stellar spectra of galaxies (see § 4) that are non–emission–line, the method of normalization does not affect the results of PCA. To verify this argument, we have repeated the analysis with other methods of normalization, e.g., $f_{5500} = 1$ (Kennicutt 1992) and $\sum f_i = 1$ (Sodrê & Cuevas 1997), and found that the results are almost independent of the adopted normalization.

The 204 scaled spectra are then analyzed by PCA after the data matrix is transposed (see § 2). The fractional contributions to variance by the first three eigenspectra (Fig. 3) are 61.7%, 34.4%, and 1.7%. In Figure 4 (top), the 204 stars are plotted in the plane of $p_1$ versus $p_2$, where $p_1$ and $p_2$ are the relative contributions by each star to the first two eigenspectra. It is remarkable that most of the stars are distributed on the circle with $r = (p_1^2 + p_2^2)^{1/2} \approx 1$. Only few very late and early-type stars deviate from this unit circle, scattering well within $r = 1$.

![Fig. 2.—Distribution of the spectral types of the STELIB library (crosses) and the 204 stars used in this paper (squares). Note that, out of the total 255 stars in STELIB, Feige 110 (dwarf A), LTT 4364 (dwarf carbon), and six WC- or WN-type stars are not plotted in this figure.](image)

![Fig. 3.—First three eigenspectra of the stellar library. The solid lines represent the spectra in the visible range, and the crosses are the $J$, $H$, and $K_s$ fluxes.](image)

![Fig. 4.—Top: Relative contributions of the 204 stars to the first two eigenspectra. Bottom: Position angles $\theta$ of the 204 stars in the top panel, as a function of their spectral types.](image)
This result is understandable because the cumulative contribution to variance by the first two eigensstars is 96.1%, indicating that they contain nearly all the information of the original 204 stars. For those stars scattering far from the unit circle, the contribution of other eigenspectra is important. In the bottom panel of Figure 4, the position angles of the 204 stars in the upper panel are plotted against their spectral types. As can be seen, the 204 stars are well separated on the unit circle and are ordered by spectral types from early to late. Therefore, the PCA method may provide a new way of spectroscopic classification of stars.

The eigenspectra so obtained will be used as the absorption-line templates to model the spectra of galaxies in the next step. As has been pointed out in the last section, one of the advantages of the PCA method is that only a few of the first eigenspectra are enough for spectral modeling. In order to determine the number of significant eigensstars, we estimate the expected level of the variance caused by the noise in the spectra as \( \sum_{i=1}^{204} \sum_{p=1}^{204} \sigma_{ij}^2 / \sum_{i=1}^{204} \sum_{p=1}^{204} m_p \), where \( m_p \) is the number of wavelength channels, \( f_{ij} \) and \( \sigma_{ij} \) are the flux and its error of the \( j \)th wavelength channel of the \( i \)th star, and \( \bar{m}_p \) is the average flux of the \( p \)th waveband. The errors \( \sigma_{ij} \) are determined by assuming a signal-to-noise ratio of \( S/N = 50 \), which is the typical value for STELIB. This estimation gives rise to a significance of 0.2% and subsequently the number of significant eigensstars of 24, indicating that the first 24 eigensstars together contribute nearly all the useful information of the stellar library.

4. DERIVATION OF GALACTIC EIGENSPECTRA

4.1. Selection of Galactic Templates

A full spectral type coverage is crucial to a library of galaxies, as it is to that of stars. SDSS is to date the most ambitious imaging and spectroscopic survey and will eventually cover a quarter of the sky (York et al. 2000). The large coverage of area and moderately deep survey limit of the SDSS make it very propitious for constructing a library of template galaxies with full coverage of spectral type.

In the first step, all the low-redshift (\( z < 0.2 \)) and high–signal-to-noise ratio [(S/N)\( _p > 30 \) or (S/N)\( _c > 40 \)] objects in the SDSS DR1 spectroscopically classified as galaxies by the SDSS pipeline are selected as candidates of the template galaxies. The resulting 7098 spectra are transformed to the rest frame using the redshift provided by the SDSS spectroscopic pipeline.

The 7098 spectra are then fitted with the first 24 eigenspectra of the star library obtained in the last section. The eigenspectra are broadened to velocity dispersions from 0 to 600 km s\(^{-1}\) using a Gaussian kernel. To avoid the effect of emission lines, the central \( \sim 5 \) Å of emission lines (e.g., Balmer system, forbidden lines) are excluded from the fit. The best-fitting model of each spectrum is therefore derived through the \( \chi^2 \) minimization by taking into account flux uncertainties provided by the SDSS pipeline. Given the 24 star eigenspectra as the input templates, our program solves for the systemic velocity, the line-broadening function, and the relative contributions of the various templates. The best-fitting model, in general, is a good solution for the absorption-line spectrum of the stellar component.

The large size of the galaxy sample provides us with an extensive collection of modeled spectra spanning a full range of spectral type. However, the distribution of the spectral types in the galaxy library is not uniform. To get a uniform library, we reselect candidates on a set of color-color diagrams. Each modeled spectrum is cut into 18 passbands with widths of \( \sim 200 \) Å, and a synthesized magnitude of each band is obtained via

\[
m_i = -2.5 \log \int f_i^j d\lambda, \quad i = 1, 2, \ldots, 18,
\]

where \( f_i^j \) is the spectrum in the \( j \)th waveband. The color-color diagrams should make use of all 18 magnitudes, whereas it does not mean that we need to make use of all the possible color-color diagrams. We assign the 18 magnitudes into six groups, each of which have three magnitudes. For each group, the three magnitudes give rise to two colors, \( c_1 \) and \( c_2 \):

\[
c_{i1} = m_i - m_{i+6}, \quad c_{i2} = m_{i+6} - m_{i+12}, \quad i = 1, 2, \ldots, 6,
\]

where \( c_{i1} \) and \( c_{i2} \) are the two colors in the \( i \)th group. In this way, all 18 magnitudes but only six color-color diagrams are used.

The candidates are then selected on the six color-color diagrams. To this end, each diagram is partitioned into square meshes with a size of 0.02 mag. The mesh size is so chosen as to have nearly a full coverage of spectral type for the galaxy sample. The galaxies located within each mesh are ordered by decreasing the signal-to-noise ratios of their spectra, and the first 10% are picked up. The galaxy with the highest spectral signal-to-noise ratio is selected if there are less than 10 galaxies in a mesh. As an example, Figure 5 shows the selection procedure on the \( m_6 - m_{12} \) versus \( m_{12} - m_{18} \) diagram. In this stage, 1126 unique galaxies are obtained. Each spectrum is then examined by eye, and 110 objects were rejected because of either the presence of bad wavelength channels in the spectrum or the contamination of nuclear activity. Finally, 1016 galaxies were chosen as our galaxy templates.

4.2. Iterative Spectral Modeling Using Stellar Templates

The procedure of spectral modeling for the 7098 template candidates is rather rough, although it is sufficient for sample
selection. In fact, the following issues that might affect the fit should be carefully addressed.

First, masking precisely the emission-line region is vital to the measurement of the profile of absorption lines, which are sometimes filled partially with one or more emission lines. A narrower masked range may include wings of the emission line, whereas a wider one will also exclude useful information of the absorption-line profile. Besides, the width and equivalent width of emission lines differ from line to line and from object to object. A specific masked region for each line in each object is required for such subtle treatment. In this paper, we use the measured emission parameters to create the masked region for each emission line of each object.

Second, galaxies often suffer from intrinsic reddening to certain extent. In this paper, we estimate the intrinsic extinction of galaxies in a way similar to that commonly used by population synthesis, i.e., a single extinction for the whole galaxy. Certainly this is only a zero-order solution and by no means a satisfactory one. During the modeling, the program searches a range of color excess $E(B-V)$ to find the most plausible value by assuming an extinction curve of Calzetti et al. (2000).

Finally, bad pixels in the SDSS spectrum, flagged by the SDSS pipeline, or in the stellar templates are also masked from fitting.

We use an iterative procedure to remodel the 1016 galaxies by taking care of the above issues. Each spectrum is modeled at least three times. Initially, the same procedure as described in § 4.1 is performed. The modeled spectrum is then subtracted from the observed one, and the emission lines are fitted with Gaussian functions (see Dong et al. 2005 for details). Because emission lines and absorption lines are coupled, this procedure must be performed iteratively. An average reddening is added to the model. Detailed parameters are listed below.

1. The range of $E(B-V)$ is set to be from 0 to 2 with a step size of 0.01.
2. Pixels with emission-line flux above $3 \sigma$, the flux uncertainty of SDSS at that pixel, are masked. See Figure 6 for an example of this procedure.
3. Pixels in the wavelength ranges from 6800 to 7100 and from 7500 to 7700 Å in the source rest frame are masked because of the atmosphere absorption in the original stellar library.

4. Pixels within 100 Å of the left end and 200 Å of the right end of each spectrum are masked to avoid possible calibration problems.

4.3. PCA of Galactic Library

Galaxy template spectra with effectively zero velocity dispersion are obtained according to the procedure described in § 4.2. Given a set of expansion coefficients, the modeled spectrum with zero velocity dispersion can be obtained via

$$f_{\ell} = \sum_{i=1}^{24} a_ie_{\ell,i},$$

where $e_{\ell,i}$ is the $i$th eigenstar and $a_i$ is the best-fitting coefficient.

Next, the 1016 zero velocity dispersion spectra are analyzed by PCA, in the same way as in § 3.3. Because the modeled spectra are constructed using the first 24 star eigenspectra, only the first 24 eigengalaxies are nonzero. The fractional contributions of the first three eigengalaxies to variance are 81.4%, 17.0%, and 0.4%. Figure 7 shows the spectra of these three eigengalaxies, including their fluxes at the $J$, $H$, and $K_s$ bands. As in the case for the stellar library, the variance is so dominated by the contributions of the first two eigengalaxies that most of the galaxies are distributed around a circle of radius $r = (p_1^2 + p_2^2)^{1/2} = 1$ on the $p_1$ versus $p_2$ plane (see Fig. 8).

The number of eigengalaxies that are required to fit the observed galaxies is determined as follows. Initially, the observed spectra are modeled using the first three eigengalaxies. By adding successively the next eigengalaxy to the model, we calculate the significance of the improvement to the fit using the $F$-test:

$$\alpha_F = \int_F^\infty dF p(F|\Delta P, N-P_1)$$

$$= I_{N-P_1-\Delta P} \left( \frac{N-P_1}{2} \right)^{-\frac{1}{2}} \frac{\Delta P}{2},$$

where the $F$-statistics $F = (\Delta \chi^2/\Delta P)/[\chi_1^2/(N-P_1)]$, $P_1 = 1$, and $\Delta P = 1$ are the numbers of thawed parameters of the previous model and of the additional freely varying parameters in the current model and $I$ is the incomplete beta function. Adopting a critical significance of $\alpha_F = 0.05$, we find that more than 97% of the galaxies can be well modeled using the first...
nine eigenspectra (see Fig. 9), which are then chosen as our final absorption-line templates.

5. APPLICATIONS AND TESTS

5.1. Modeling the Spectra in SDSS DR1

The spectra of all galaxies in SDSS DR1 (~1.4 × 10⁵ spectra) are fitted with the nine galactic eigenspectra with iterative rejection of emission lines and bad pixels (see § 4.2). On a personal computer with a CPU of main frequency 2.8 GHz, this procedure takes ~20 hr. The reduced \( \chi^2 \), namely, \( \chi^2_p = \chi^2 \) per degree of freedom), for 98.8% of the spectra are less than 1.5, with the peak of the \( \chi^2_p \) distribution around 0.96 and the mean \( \chi^2_p \sim 1.04 \), indicating that the fits are quite good (see Fig. 10). Figure 11 shows several examples of the fits. Overall residuals in the non–emission-line region are consistent with flux fluctuations. The figure also illustrates the importance of the spectral modeling to the emission-line measurements. In some of the original spectra, even H\( \beta \) is hardly visible, whereas it can be easily measured after the underlying starlight is subtracted. Furthermore, the intensities of both H\( \beta \) and H\( \alpha \) are modified substantially, and their ratio changes to be close to the theoretical value.

We would like to point out that although most of the galactic spectra could be well modeled using this method, the fit is not satisfactory for a small number of spectra (\( \sim 0.8\% \)), of which most show peculiar molecular absorption bands and a small portion show an obvious Wolf-Rayet feature around 4600–4750 Å. The result is expected because of the lack of very late type stars and of Wolf-Rayet stars in the stellar library we used (see Fig. 2). Modeling these objects requires the improvement of the stellar library.

For each spectrum, the best-fitting model is then subtracted from the original spectrum, yielding a pure emission-line spectrum from which emission-line parameters are measured. It is found that the measured equivalent width (EW) and flux of H\( \alpha \) and the modeling coefficients to solve a set of linear equations of

\[ \text{EW}_{\text{H}\alpha}^j = \sum_{i=1}^{9} a_i \text{EW}_i \tilde{c}_{ij}, \]  

\[ f_j^\text{H}\alpha = \sum_{i=1}^{9} a_i^f c_{ij}, \]  

where \( \text{EW}_{\text{H}\alpha}^j \) and \( f_j^\text{H}\alpha \) are the measured H\( \alpha \) EW and flux of the jth spectrum, \( c_{ij} \) is the modeling coefficient of the ith eigenspectrum for the jth spectrum, and \( \tilde{c}_{ij} = c_{ij}/(\sum_{i=1}^{9} c_{ij}^2)^{1/2} \) is the corresponding normalized coefficient. Using the resulting constants \( \{a_i^{\text{EW}}\} \) and \( \{a_i^f\} \) (i = 1, . . ., 9), the H\( \alpha \) EW and flux of each galaxy can be synthesized as a linear combination of the modeling coefficients of the galactic eigenspectra. The result

![Fig. 8.](image1.png)  

![Fig. 9.](image2.png)  

![Fig. 10.](image3.png)
shows that the correlation between the measured and synthesized values is obvious, although it somewhat deviates from the linear relation for galaxies with very strong emission lines (e.g., $EW_{H\alpha} > 100$ Å, $f_{H\alpha} > 0.6 \times 10^{-13}$ erg s$^{-1}$ cm$^{-2}$): the synthesized value is smaller than the measured. This result suggests that the modeling coefficients, which directly connect with the stellar component, also reflect the information on the star formation history, including the current star formation rate (SFR). To verify this argument, we further perform the same analysis as described above on the total SFR inside the fiber of these

Fig. 11.—Examples of starlight removal. In each panel, the three lines (top to bottom) are the observed, the modeled, and the starlight-subtracted spectra. For clarity, arbitrary constants are added to the observed and the modeled spectra.
galaxies. Since SFRs are particularly meaningful if expressed with respect to galaxy masses, the measured and synthesized SFRs are normalized by stellar masses $M$ to give the corresponding correlation of specific SFRs (see Fig. 12). The measured SFRs and stellar masses are extracted from the SFR catalog (Brinchmann et al. 2004) in the data catalogs from SDSS studies at MPA/JHU.\footnote{Available at http://www.mpa-garching.mpg.de/SDSS/.}

As expected, the correlation of specific SFRs, compared with those of EWs and fluxes, is much tighter, indicating that the modeling coefficients actually contain information on the current SFRs of galaxies. The deviation of the synthesized H$_\alpha$ EW and flux from the measured values could be mainly due to the following causes. First, the relation between the stellar population and emission lines is nonlinear, and thus the emission lines cannot provide exact measurement of the current SFR. In fact, in dwarf galaxies and in star formation regions in larger galaxies, there exists a delay of the H$_\alpha$-luminosity maximum with respect to the maximum of the SFR, and subsequently the rise of the SFR is not immediately reflected in a corresponding increase in H$_\alpha$-luminosity (Alvensleben 2004). Therefore, it is not a surprise that, for the galaxies with extremely strong emission lines, the synthesized value is smaller than the measured one, if the former really matches the true value of the current SFR. In this case, the H$_\alpha$ line is expected to be at or close to its maximum stage, whereas the SFR has undergone this stage and is at a relatively lower level at the current time. Next, the absence of very early type stars in the stellar library makes the eigenspectra unable to distinguish the very young stellar populations, especially in the galaxies with high SFR, and therefore makes the modeling coefficients underestimate the SFR for these galaxies. Finally, as discussed in \S 5.4, using a single average $E(B-V)$ for very late type galaxies (e.g., Sc type) will actually underestimate the extinction of dust. In principle, a young stellar population is more reddened than an old population. The underestimation of reddening therefore leads to an underestimated fraction of the young stellar population in the modeled stellar component, thus giving rise to an underestimated SFR.

5.2. Stellar Velocity Dispersion

5.2.1. Tests of the Measurement Routine

One of the merits of the method presented in this paper is that the stellar velocity dispersion of galaxies can be determined as a by-product. In order to estimate the accuracy of the measurement of stellar velocity dispersion, we created a set of testing spectra as follows. First, all the 204 stars in STELIB are classified into seven groups according to their spectral types, i.e., O, B, A, F, G, K, and M. The average spectra of these groups are then combined by specifying various fractions to create a set of 462 spectra. The spectra are then broadened by convolution with Gaussians of $\sigma$ in the range of 40–300 km s$^{-1}$. Finally, Gaussian noise is added to the broadened spectra to give $S/N = 10–120$ pixel$^{-1}$.

The resultant 59,136 spectra with known velocity dispersion, stellar population, and signal-to-noise ratio are modeled using the method described above. Figure 13 shows the relative contributions of the first two eigenspectra to the best-fitting spectrum of 59,136 created spectra with given stellar velocity dispersion, stellar population, and signal-to-noise ratio. The three thick lines are the contours of number density ($n$) of all DR1 galaxies with levels of log ($n$) = 1, 2, and 3. Bottom: Plot of the rms disagreement between the measured and input velocity dispersion of the 12,820 created spectra, which are selected on the top panel.
Figure 13 shows the rms disagreement between the measured and input velocity dispersion. As a whole, the measured dispersion is within 35 km s$^{-1}$ of the input velocity dispersion for S/N $> 10$. For synthesized spectra with S/N $\sim 20$, the typical value of the SDSS galaxies, the uncertainties of the velocity dispersion measurements is between $\sim 4\%$ and $\sim 10\%$ for higher velocity dispersion ($\sigma_v > 75$ km s$^{-1}$), whereas estimates of low-velocity dispersions are less accurate ($\sim 20\%$). For higher signal-to-noise ratios, the measurement errors decrease rapidly. We would like to point out that the accuracy of the stellar velocity dispersion determined using the method presented in this paper is comparable to previous works at high signal-to-noise ratios. For the spectra with input dispersion range of 50–300 km s$^{-1}$ and S/N $= 120$, Barth et al. (2002) yielded an accuracy within 6 km s$^{-1}$, whereas this value is less than 5 km s$^{-1}$ using our method (see Fig. 13). However, our approach should overperform these methods using only a couple of absorption lines at low signal-to-noise ratio.

5.2.2. Comparison with the SDSS Pipeline

Stellar velocity dispersion is also measured by the SDSS spectroscopic pipeline using two different methods: “Fourier fitting” and “direct fitting.”\(^4\) The latter is similar to the method presented in this paper, except that the SDSS pipeline uses only 32 K and G giant stars in M67 as templates. The SDSS velocity dispersion estimates are obtained by fitting the rest-frame wavelength range 4000–7000 Å and then averaging the estimates provided by the Fourier-fitting and direct-fitting methods. The error on the final value of the velocity dispersion is determined by adding in quadrature the errors on the two estimates. The typical error is between $\Delta(\log \sigma_v) \sim 0.02$ and 0.06 dex, depending on the signal-to-noise ratio of the spectra.

However, stellar velocity dispersions $\sigma_v$ are only measured for spheroidal systems by the SDSS spectroscopic pipeline. The main selection criteria is the PCA classification “eClass” being less than $-0.02$, typical of spectra of early-type galaxies. Adopting these criteria, $\sim 4.6 \times 10^4$ spectra of galaxies in SDSS DR1 were chosen to measure $\sigma_v$ by the SDSS pipeline. The values determined by the SDSS pipeline are compared with our measurements in Figure 14 (top). The two estimates are almost consistent with each other, except that our measurements are systematically a bit smaller than those determined by the SDSS pipeline. This is understandable because our model is expected to fit the spectrum better, reducing the template mismatch problem. The iterative rejection of emission lines in our procedure can also improve the accuracy of the velocity measurement.

In Figure 14 (bottom), the distribution of the velocity dispersion of this paper and that provided by the SDSS pipeline are presented. In addition to the $\sim 4.6 \times 10^4$ galaxies, we also measure the other $\sim 8.8 \times 10^4$ galaxies whose $\sigma_v$ are not provided by the SDSS pipeline. It can be seen that the velocity dispersions of these objects are systematically smaller than those of the $\sim 4.6 \times 10^4$ early-type galaxies. This is consistent with the general impression that the stellar velocity dispersion of spheroidal systems is generally larger than that of late-type ones. Using the method presented in this paper, the data set of $\sigma_v$ is enlarged $\sim 3$ times compared with that provided by the SDSS pipeline. As it is illustrated in Figure 6, by carefully masking the emission-line wavelength range, $\sigma_v$ can be reliably determined using our method for most emission-line galaxies.

\(^4\) See http://www.sdss.org/dr1/algorithms/veldisp.html.

Galaxy classification plays an important role in the study of galaxy formation and evolution. Three methods are often used for this purpose: morphological segregation, rest-frame colors, and direct spectrum-based classifications, and each method has its own unique drawbacks and advantages (Madgwick et al. 2003).

As shown in Figure 4, the position angle $\theta$ of the stars on the diagram of their relative contributions to the first two eigenstars is well correlated with the spectral type, suggesting that $\theta$ is a good indicator of stellar types and thus useful for stellar classification. Similarly, most of the galaxies on the diagram of their relative contributions to the first two eigengalaxies also are distributed on the unit circle (see Fig. 8). The corresponding position angle of galaxies is therefore expected to be a good single-parameter classifier of galaxies. However, such a PCA-based classification of galactic spectra is only applicable to small samples, such as the galaxy sample presented in § 4.1, because of the required computational expense of implementing PCA on large or very high dimensional data sets. Nevertheless, we note that there is another similar parameter, namely, the position angle $\theta'$ of galaxies on the diagram of the contributions of the first two eigengalaxies to the modeled spectra (the $p_1'$-$p_2'$ plane; see Fig. 13), which is also a good classifier.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure14.png}
\caption{Top: Comparison of stellar velocity dispersion as provided by the SDSS pipeline ($\sigma_v^{\text{SDSS}}$) and as obtained during the spectral modeling ($\sigma_v^{\text{Model}}$) for 46,229 spectra in SDSS DR1. The solid (dashed) line is the fitted (identical) relation. Bottom: Distribution of stellar velocity dispersion estimated using the present method. The dotted line shows the $\sim 4.6 \times 10^4$ galaxies that have SDSS velocity dispersions (the corresponding SDSS velocity dispersions are plotted with a dot-dashed line), the dashed line represents the other $\sim 8.8 \times 10^4$ galaxies without SDSS velocity dispersion, and the solid line shows all the DR1 galaxies.}
\end{figure}
In fact, the SDSS DR1 pipeline also provides a spectral classification of galaxies by cross-correlating with eigentemplates constructed from early SDSS spectroscopic data using the PCA method. Five eigencoefficients and a classification number are stored in parameters "eCoeff1"–"eCoeff5" and "eClass," respectively (Stoughton et al. 2002). The parameter eClass, based on the expansion coefficients eCoeff1–eCoeff5, ranges from about $\frac{\sigma_0}{0.35}$ to 0.5 for early- to late-type galaxies. Since the sign of the second eigenspectrum has been reversed with respect to that of the Early Data Release (see Stoughton et al. 2002), the expression $\tan^{-1}\left(-\frac{eCoeff2}{eCoeff1}\right)$ is recommended rather than eClass as the single-parameter classifier.5

To compare our classifier $\theta'$ with that provided by the SDSS pipeline, we randomly select 1000 galaxies from SDSS DR1 and fit their spectra to obtain the modeling coefficients of templates and subsequently the position angles $\theta'$. As can be seen clearly in Figure 15 (top), the position angle $\theta'$ is actually a good single-parameter classifier, which is well correlated with the classifier $\tan^{-1}\left(-\frac{eCoeff2}{eCoeff1}\right)$ provided by the SDSS pipeline. This is understandable because both of the classifiers are obtained on the basis of PCA. Moreover, our classifier $\theta'$ is also well correlated with the mean velocity dispersion $\langle \sigma_0 \rangle$ of the 1000 galaxies (see Fig. 15, bottom), with smaller $\theta'$ corresponding to larger velocity dispersion and thus to earlier type galaxies.

Using the rest-frame colors, Strateva et al. (2001) studied the color distribution of a large uniform sample of galaxies detected in SDSS commissioning data and showed that the $g^* - r^*$ versus $u^* - g^*$ color-color diagram is strongly bimodal, with an optimal color separation of $u^* - r^* = 2.22$ and the two peaks corresponding roughly to early- (E, S0, and Sa) and late-type (Sb, Sc, and Irr) galaxies. As can be seen clearly in Figure 16, the early- and late-type galaxies separated by $u^* - r^* = 2.22$ could be also well classified using our single-parameter classifier $\theta'$, indicating that $\theta'$ might be also compatible for galactic classification.

5.4. Near-Infrared SED

Once a spectrum has been modeled using the nine galactic eigenspectra, the contributions of the 24 eigenstars to the modeled spectrum can be conveniently obtained. Since the 24 stellar eigenspectra contain the infrared information of the 204 original stars (§ 3), the infrared fluxes at $J$, $H$, and $K_s$ bands could be reconstructed from a set of expansion coefficients. We test the reliability of the stellar SED in the near-infrared using this method on a set of "average" spectra of normal galaxies along the Hubble diagram between E and Sc between 0.1 and 2.4 $\mu$m, presented by Mannucci et al. (2001).6 The composite spectra were derived from 28 nearby galaxies, and the overall precision of the calibrated spectra is about 2%.  

---

5 See http://www.sdss.org/dr1/algorithms/redshift_type.html.

6 Available at http://www.arcetri.astro.it/~filippo/spectra.
The five spectra are modeled using the nine templates, and the results are presented in Figure 17. As can be seen, the modeled infrared fluxes are systematically lower than the observed ones by ~10% for E to Sb types and by 30% for Sc galaxies. The discrepancy is partly, but slightly, due to the contribution of the additional emission from hot dust to the observed infrared flux, presumably caused by stochastic heating of small grains by UV photons, which is expected to increase with both the dust content and the young stellar population along the Hubble sequence. The most probable explanation to the discrepancy may be the fact that, using a single average E(B - V), we actually underestimate the true extinction in the optical, and also the stellar mass, etc., by that amount. A more reasonable model for the absorption of starlight by dust (e.g., Charlot & Fall 2000) is desired to be added in the modeling. Nevertheless, we argue that the result of our method could give relatively useful information of starlight at infrared bands for a galaxy (especially for early types) by only using its optical spectrum.

5.5. Average Reddening

Our solution also yields an average stellar reddening E(B - V) for each galaxy (see § 4.2). The extinction is quite significant for 85.5% of the galaxies, with the E(B - V) in the range of 0.01–0.5. On the other hand, the internal reddening can be also estimated using Balmer decrements of the emission lines from H II regions. To check the consistency of the two estimates, we select a sample of ~10^4 H II galaxies from SDSS DR1, adopting the criteria of Kewley et al. (2001):

$$\log \left( \frac{[\text{O} \text{III}] \lambda5007}{\text{H} \beta} \right) < \frac{0.61}{\log (\text{N} \beta)/\text{H} \alpha} + 1.19. \quad (11)$$

Using the effective extinction curve $\tau_V = \tau_H(\lambda/5500 \text{ Å})^{0.7}$, which was introduced by Charlot & Fall (2000), the color excess arising from attenuation by dust in the selected galaxy, $E(B - V)_{\text{Balmer}}$, can be written

$$E(B - V)_{\text{Balmer}} = A_V/R_V = 1.086 \tau_V/R_V, \quad (12)$$

$$\tau_V = -\frac{\log \left( \frac{F(\text{H} \beta)/F(H\beta)}{F(\text{H} \alpha)/F(H\alpha)} \right) - \log \left( \frac{I(\text{H} \beta)/I(H\beta)}{I(\text{H} \alpha)/I(H\alpha)} \right)}{(\lambda_{\text{H} \alpha} / \lambda_{5500})^{0.7} - (\lambda_{\text{H} \beta} / \lambda_{5500})^{0.7}}, \quad (13)$$

where $I(\text{H} \beta)/I(H\beta) = 2.87$ is the intrinsic Balmer flux ratio, $F(\text{H} \alpha)/F(H\beta)$ is the observed Balmer flux ratio, $\tau_V$ is the effective V-band optical depth, $\lambda_{\text{H} \alpha} = 6563$ Å, $\lambda_{\text{H} \beta} = 4861$ Å, and $R_V = 3.1$. For objects with the observed flux ratio $F(\text{H} \alpha)/F(H\beta) = 2.87$, $\tau_V$ is set to zero. The resulting $E(B - V)_{\text{Balmer}}$ versus $E(B - V)'$ is plotted in Figure 18. Although the dispersion of $E(B - V)'$ is large (~0.2), the tendency is apparent: the $E(B - V)'$ increases with increasing $E(B - V)_{\text{Balmer}}$, and the former is systematically smaller than the latter, which is also consistent with previous works (e.g., Calzetti et al. 1994; Charlot & Fall 2000).

6. SUMMARY

We have developed an empirical method for modeling the stellar spectrum of galaxies. The absorption-line templates with zero velocity dispersion are constructed on the basis of two successive applications of Principal Component Analysis (PCA), first to 204 stars in the stellar library STELIB, then to a uniform sample of galaxies selected from SDSS DR1. With nine templates, we can fit quite well the stellar spectra of galaxies in the SDSS DR1. As by-products, the stellar velocity dispersion, the near-infrared SED, the spectral type, and the average reddening
are determined simultaneously. For a spectrum of S/N = 20, typical for SDSS galaxies, the velocity dispersion can be determined to an accuracy of ~4%–10% at larger values (>75 km s\(^{-1}\)) and ~20% at low dispersion (<75 km s\(^{-1}\)). The average reddening of stellar light is correlated and is systematically smaller than that derived from Balmer decrements of H\(\beta\) regions. After stellar light has been subtracted, emission-line parameters are measured for all galaxies. An interesting result is that the measured SFR is well correlated with the modeling coefficients of the templates, suggesting that these coefficients also contain the information on the current SFR of galaxies.

The success of this approach can be understood from the following aspects. First, two applications of PCA highly concentrate the most prominent features from both the stellar library and the galaxy sample into the first few galactic eigenspectra and hence significantly reduce the number of the final templates. This makes the procedure of spectral modeling much quicker and the results much more stable and reliable, in particular when dealing with large data sets, in comparison with using direct stellar populations. Second, the templates derived from the large galaxy sample with full coverage of spectral type ensure a close match to the true underlying stellar population of different types of galaxies, including emission-line galaxies that contain much young stellar components. As a result, the modeled spectra reflect the true information on the stellar population in galaxies and provide useful clues for their spectral classifications. Third, the templates are not obtained directly from the observed galactic spectra, but from the modeled one with zero velocity dispersion, and therefore could match better the absorption-line profiles of galaxies. Finally, the extended wavelength coverage (from optical to near-infrared) in the templates enables us to study the stellar component at the near-infrared band using only the optical spectrum of galaxies.

It should be pointed out that there is still room for improvement of this method. First, our lack of very late type stars limits our application of templates in modeling a small fraction of galaxies that show prominent, peculiar molecular absorption features. This can be improved by adding very late type stars to the stellar library. Second, the implication of each PC should be tackled by setting a relation between modeling coefficients of templates and stellar population can be addressed by combining the stellar population synthesis with the PCA method, which is still under investigation. Finally, a more reasonable model for the extinction of dust should be added in the modeling to replace the single average reddening used in the present method.

We thank Dr. T. Kiang for help with our English. We are grateful to the anonymous referee, whose comments were useful for improving the paper. This work is supported by the Chinese National Science foundation through NSF 10233030, the BaiRen project of Chinese Academy of Science, and a key program of the Chinese Ministry of Science and Technology. This work is also partly supported by the Excellent Young Teachers Program of MOE, P.R.C. This paper has made use of the stellar spectra libraries STELIB and HILIB and data from 2MASS and SDSS. Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are the University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, Los Alamos National Laboratory, the Max Planck Institute for Astronomy, the Max Planck Institute for Astrophysics, New Mexico State University, Princeton University, the US Naval Observatory, and the University of Washington.