

## THE CUT-AND-ENHANCE METHOD: SELECTING CLUSTERS OF GALAXIES FROM THE SLOAN DIGITAL SKY SURVEY COMMISSIONING DATA

TOMOTSUGU GOTO,<sup>1,2</sup> MAKI SEKIGUCHI,<sup>1</sup> ROBERT C. NICHOL,<sup>2</sup> NETA A. BAHCALL,<sup>3</sup> RITA S. J. KIM,<sup>3</sup> JAMES ANNIS,<sup>4</sup> ŽELJKO IVEZIĆ,<sup>3</sup> J. BRINKMANN,<sup>5</sup> GREGORY S. HENNESSY,<sup>6</sup> GYULA P. SZOKOLY,<sup>7</sup> AND DOUGLAS L. TUCKER<sup>4</sup>

Received 2001 May 28; accepted 2001 December 19

### ABSTRACT

We describe an automated method, the cut-and-enhance (CE) method, for detecting clusters of galaxies in multicolor optical imaging surveys. This method uses simple color cuts, combined with a density enhancement algorithm, to up-weight pairs of galaxies that are close in both angular separation and color. The method is semiparametric, since it uses minimal assumptions about cluster properties in order to minimize possible biases. No assumptions are made about the shape of clusters, their radial profile, or their luminosity function. The method is successful in finding systems ranging from poor to rich clusters of galaxies, of both regular and irregular shape. We determine the selection function of the CE method via extensive Monte Carlo simulations that use both the real, observed background of galaxies and a randomized background of galaxies. We use position-shuffled and color-shuffled data to perform false-positive tests. We have also visually checked all the clusters detected by the CE method. We apply the CE method to the 350 deg<sup>2</sup> of the Sloan Digital Sky Survey (SDSS) commissioning data and construct an SDSS CE galaxy cluster catalog with an estimated redshift and richness for each cluster. The CE method is compared with other cluster selection methods used on SDSS data such as the matched filter, “maxBCG,” and Voronoi tessellation techniques. The CE method can be adopted for cluster selection in any multicolor imaging survey.

**Key words:** galaxies: clusters: general — methods: analytical

### 1. INTRODUCTION

Clusters of galaxies are the most massive virialized systems known and provide powerful tools in the study of cosmology and extragalactic astronomy. For example, clusters are efficient tracers of the large-scale structure in the universe and aid in determining the amount of dark matter on megaparsec scales (Bahcall 1998; Carlberg et al. 1996; Borgani & Guzzo 2001; Nichol 2002 and references therein). Furthermore, clusters provide a laboratory within which to study a large number of galaxies at the same redshift and, thus, assess the effects of dense environments on galaxy evolution, for example, the morphology-density relation (Dressler 1980, 1984; Dressler et al. 1997), the Butcher-Oemler effect (Butcher & Oemler 1978, 1984), and the density dependence of the luminosity function of galaxies (Garilli, Maccagni, & Andreon 1999). In recent years, surveys of clusters of galaxies have been used extensively in constraining cosmological parameters such as  $\Omega_m$ , the mass density parameter of the universe, and  $\sigma_8$ , the amplitude of mass fluctuations at a scale of 8 h<sup>-1</sup> Mpc (see, e.g., Oukbir & Blanchard 1992; Viana & Liddle 1996, 1999; Eke, Cole, &

Frenk 1996; Bahcall, Fan, & Cen 1997; Henry 1997, 2000; Reichart et al. 1999). Such constraints are achieved through the comparison of the evolution of the mass function of galaxy clusters, as predicted by the Press-Schechter formalism (see Jenkins et al. 2001 for the latest analytical predictions) or simulations (e.g., Evrard et al. 2001; Bode et al. 2001), with the observed abundance of clusters with redshift. Therefore, to obtain robust constraints on  $\Omega_m$  and  $\sigma_8$ , we need large samples of clusters that span a large range in redshift and mass and that possess a well-determined selection function (see Nichol 2002).

Despite their importance, existing catalogs of clusters are limited in both their size and quality. For example, the Abell catalog of rich clusters (Abell 1958) and its southern extension (Abell, Corwin, & Olowin 1989) are still some of the most commonly used catalogs in astronomical research, even though they were constructed by visual inspection of photographic plates. Another large cluster catalog by Zwicky et al. (1961–1968) was similarly constructed by visual inspection. Although the human eye can be efficient in detecting galaxy clusters, it suffers from subjectivity and incompleteness. For cosmological studies, the major disadvantage of visually constructed catalogs is the difficulty in quantifying the selection bias and, thus, the selection function. Furthermore, the response of photographic plates is not uniform. Plate-to-plate sensitivity variations can disturb the uniformity of the catalog. To overcome these problems, several cluster catalogs have been constructed using automated detection methods on CCD imaging data. They have, however, been restricted to small areas because of the lack of large-format CCDs. For example, the PDCS catalog (Postman et al. 1996) only covers 5.1 deg<sup>2</sup> with 79 galaxy clusters. The need for a uniform, large cluster catalog is strong. The Sloan Digital Sky Survey (SDSS; York et al. 2000) data offer the opportunity to produce the largest and

<sup>1</sup> Institute for Cosmic Ray Research, University of Tokyo, 5-1-5 Kashiwa, Kashiwa City, Chiba 277-8582, Japan.

<sup>2</sup> Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890.

<sup>3</sup> Princeton University Observatory, Peyton Hall, Princeton, NJ 08544-1001.

<sup>4</sup> Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510.

<sup>5</sup> Apache Point Observatory, P.O. Box 59, Sunspot, NM 88349.

<sup>6</sup> US Naval Observatory, 3450 Massachusetts Avenue, NW, Washington, DC 20392-5420.

<sup>7</sup> Astrophysikalisches Institut Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany.

most uniform galaxy cluster catalog in existence because the SDSS is the largest CCD imaging survey currently underway, scanning 10,000 deg<sup>2</sup> centered approximately on the north Galactic pole.

The quantity and quality of the SDSS data demand the use of sophisticated cluster-finding algorithms to help maximize the number of true cluster detections while suppressing the number of false positives. The history of automated cluster-finding methods goes back to Shectman's counts-in-cells method (Shectman 1985). He counted the number of galaxies in cells on the sky to estimate the galaxy density. Although this provided important progress over the visual inspection, the results depend on the size and position of the cell. Currently, the most commonly used automated cluster-finding method is the matched filter technique (Postman et al. 1996; Kawasaki et al. 1998; Kepner et al. 1999; Schuecker & Böhringer 1998; Bramel, Nichol, & Pope 2000; Lobo et al. 2000; da Costa et al. 2001; Willick et al. 2001). The method assumes a filter for the radial profile of galaxy clusters and for the luminosity function of their members. It then selects clusters from imaging data by maximizing the likelihood of matching the data to the cluster model. Although the method has been successful, galaxy clusters that do not fit the model assumption (density profile and luminosity function) may be missed. We present here a new cluster-finding method called the *cut-and-enhance* (CE) method. This new algorithm is semiparametric and is designed to be as simple as possible, using the minimum number of assumptions possible about cluster properties. In this way, it should be sensitive to all types of galaxy overdensities, even those that may have recently undergone a merger and therefore are highly nonspherical. One major difference between the CE method and previous cluster finders is that the CE method makes full use of colors of galaxies, which become available as a result of the advent of the accurate CCD photometry of the SDSS data. We apply this detection method on 350 deg<sup>2</sup> of the SDSS commissioning data and construct a large cluster catalog. The catalog ranges from rich clusters to the more numerous poor clusters of galaxies over this area. We also determine the selection function of the CE method.

In § 2, we describe the SDSS commissioning data. In § 3, we describe the detection strategy of the CE method. In § 4, we present a performance test of the CE method and the selection function using Monte Carlo simulations. In § 5, we visually check the success rate of the method. In § 6, we compare the CE method with the other detection methods applied to the SDSS data. In § 7 we summarize the results.

## 2. THE SDSS COMMISSIONING DATA

The data we use to construct the SDSS CE galaxy cluster catalog are equatorial scan data taken in 1998 September and 1999 March during the early part of the SDSS commissioning phase. Contiguous areas of 250 deg<sup>2</sup> ( $145^{\circ}1 < \text{R.A.} < 236^{\circ}0$ ,  $-1^{\circ}25 < \text{decl.} < 1^{\circ}25$ ) and 150 deg<sup>2</sup> ( $350^{\circ}5 < \text{R.A.} < 56^{\circ}51$ ,  $-1.25 < \text{decl.} < 1.25$ ) were obtained during four nights, where seeing varied from 1''.1 to 2''.5. Since we intend to use the CE method at the faint end of the imaging data, we include galaxies to  $r^* = 21.5$  (Petrosian magnitude), which is the star-galaxy separation limit. Since the SDSS photometric system is not yet finalized, we refer to the SDSS photometry presented here as  $u^*$ ,  $g^*$ ,  $r^*$ ,  $i^*$ , and  $z^*$ . The technical aspects of the SDSS camera

are described in Gunn et al. (1998). Fukugita et al. (1996) describe the color filters of the SDSS. The details of the SDSS commissioning data are described in Stoughton et al. (2002).

## 3. CUT-AND-ENHANCE CLUSTER DETECTION METHOD

### 3.1. Color Cut

The aim of the CE method is to construct a cluster catalog that has as little bias as possible by minimizing the assumptions about cluster properties. If a method assumes a luminosity function or radial profile, for example, the resulting clusters will be biased toward the detection model used. We thus exclude all such assumptions except for a generous color cut. The assumption on colors of cluster galaxies appears to be robust, as all the galaxy clusters appear to have the same general color-magnitude relation (Gladders & Yee 2000). Even a claimed "dark cluster" (Hattori et al. 1997) was found to have a normal color-magnitude relation (Benítez et al. 1999; Clowe, Trentham, & Tonry 2001; Soucail et al. 2001).

Galaxy clusters are known to have a tight color-magnitude relation; among the various galaxy populations within a cluster (i.e., spiral, elliptical, dwarf, irregular), bright red elliptical galaxies have similar colors and populate a red ridgeline in the color-magnitude diagram (called the color-magnitude relation). Bower, Lucey, & Ellis (1992) obtained high-precision  $U$  and  $V$  photometry of spheroidal galaxies in two local clusters, Virgo and Coma. They observed a very small scatter,  $\delta(U-V) < 0.035$  rms. Ellis et al. (1997) studied the  $U-V$  color-magnitude relation at high redshift ( $z \sim 0.54$ ) and found a scatter of less than 0.1 mag rms. Similarly, Stanford, Eisenhardt, & Dickinson (1998) studied optical-infrared colors ( $R-K$ ) of early-type (E+S0) galaxies in 19 galaxy clusters out to  $z = 0.9$  and found a very small dispersion in the optical-infrared colors,  $\sim 0.1$  mag rms. Figure 1 shows the  $(r^*-i^*, r^*)$  color-magnitude diagram using SDSS data for galaxy members in the cluster A168 ( $z = 0.044$ ). The member galaxies are identified by matching the positions of galaxies in the SDSS commissioning data with the spectroscopic observations of Katgert et al. (1998). The error bars show the standard errors of  $r^*-i^*$  color estimated by the SDSS reduction software (Lupton et al. 2002). The red ridgeline of the color-magnitude relation is seen at  $r^*-i^* \sim 0.4$  from  $r^* = 17.5$  to  $r^* = 20$ . The scatter is 0.08 mag from the brightest to  $r^* = 18$ . Figure 2 shows the color-magnitude diagram (in  $g^*-r^*$  vs.  $r^*$ ) for all galaxies in the SDSS fields ( $\sim 8.3 \times 10^{-2}$  deg<sup>2</sup>) that contain A1577. A1577 has a redshift of  $z \sim 0.14$  and Abell richness class  $\sim 1$ . Figures 3 and 4 show color-magnitude diagrams for the same field in  $r^*-i^*$  and  $i^*-z^*$ , respectively. All the galaxies in the region are included. Even without spectroscopic information, the red ridge of the color-magnitude relation is clearly visible as the horizontal distribution of galaxy colors. The scatter in the color-magnitude relation is the largest in  $g^*-r^*$  because the difference of the galaxy spectral energy distribution due to the age or metallicity difference is prominent around 3500–5000 Å. The color distribution is much wider at faint magnitudes, partly because fainter galaxies have larger color errors, and partly because of the increase in the number of background galaxies.

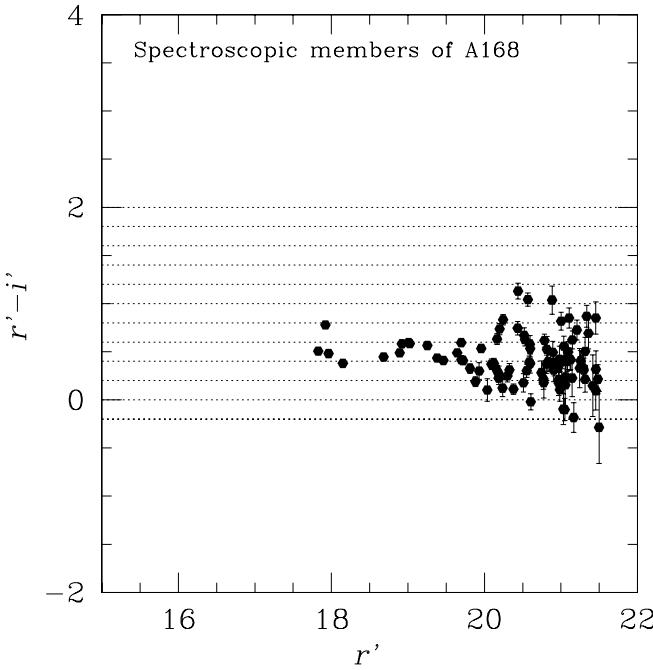


FIG. 1.—The  $r^* - i^*$  color-magnitude diagram for A168. The  $r^* - i^*$  color is plotted against  $r^*$  magnitude for confirmed member galaxies of A168. Colors and magnitudes are taken from the SDSS commissioning data by matching up the positions with the spectroscopic observations of Katgert et al. (1998). The standard errors of the colors estimated by the reduction software are shown as error bars. The  $r^* - i^*$  color-cut bins are superposed. Horizontal dotted lines are the borders of the color cuts.

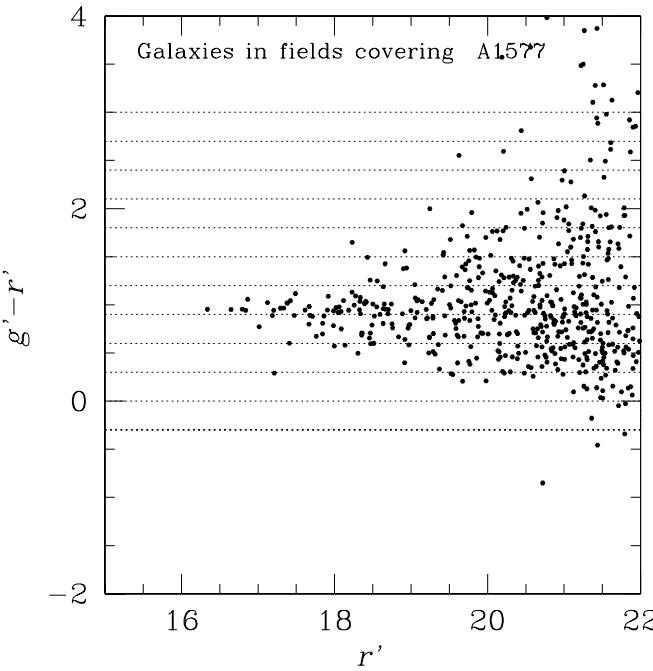


FIG. 2.—A1577  $g^* - r^*$  color-magnitude diagram. The  $g^* - r^*$  color-cut bins are superposed. The abscissa is the  $r^*$  apparent magnitude. The ordinate is  $g^* - r^*$  color. Galaxies in the SDSS fields covering A1577 ( $\sim 8.3 \times 10^{-2} \text{ deg}^2$ ) are plotted. Data are taken from the SDSS commissioning data. Horizontal dotted lines are the borders of the color cuts.

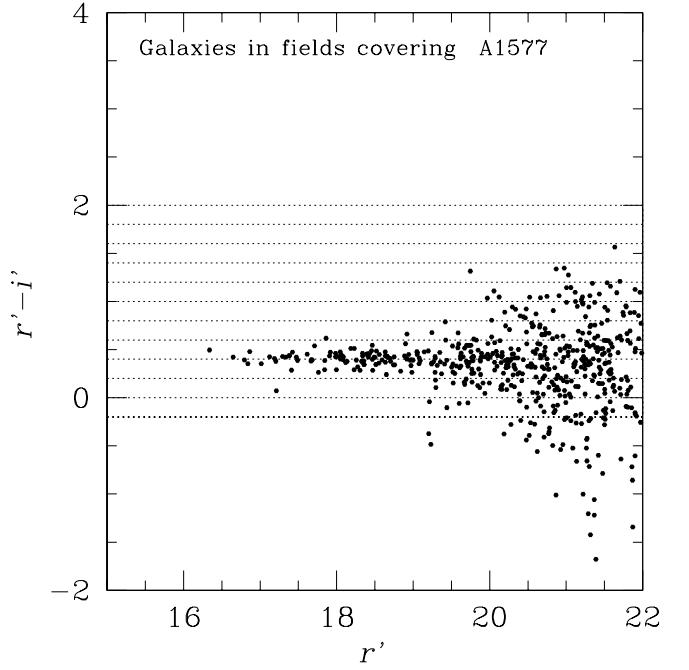


FIG. 3.—Same as Fig. 2, but for  $r^* - i^*$

The color-magnitude relation is known to have a slight tilt (Kodama et al. 1998). The tilt is small in the SDSS color bands. The tilt and its scatter in the case of A1577 (Fig. 2) are summarized in Table 1. The tilt is small in  $g^* - r^*$  and  $r^* - i^*$  ( $\sim 0.08$ ) and even smaller in  $i^* - z^*$  (0.0018). These values are much smaller than the color cuts of the CE method. The scatters are also small: 0.081, 0.040, and 0.033 in  $g^* - r^*$ ,  $r^* - i^*$ , and  $i^* - z^*$ , respectively, well smaller than the color cuts of the CE method. The small scatter of less than 0.1 mag is consistent with previous work (Bower et al.

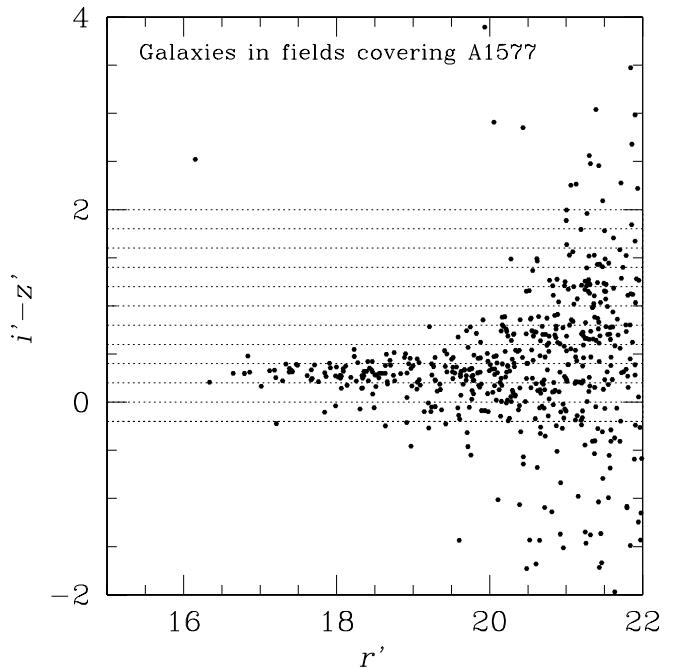


FIG. 4.—Same as Fig. 2, but for  $i^* - z^*$

TABLE 1  
TILT OF THE COLOR-MAGNITUDE RELATION OF A1577

Color	Tilt (mag)	Magnitude Range	Scatter (mag)	Magnitude Range
$g^* - r^*$ .....	0.0737	$r^* < 19$	0.081	$r^* < 17$
$r^* - i^*$ .....	0.0898	$r^* < 19$	0.040	$18 < r^* < 19$
$i^* - z^*$ .....	0.0018	$r^* < 21$	0.033	$18 < r^* < 19$

1992; Ellis et al. 1997; Stanford et al. 1998). The tilt of the color-magnitude relation is smaller than the scatter in the SDSS color bands.

We use the color-magnitude relation to enhance the detection signal of galaxy clusters. Such a use of colors of galaxies has become possible only recently as a result of the appearance of large-format CCD-based data (e.g., SDSS). Since cluster members have similar colors, we use specific but generous color cuts, to enhance the contrast of galaxy clusters. The colors of red elliptical galaxies change as a function of redshift. Figure 5 presents the  $g^* - r^*$  versus  $r^* - i^*$  color-color diagram for all galaxies brighter than  $r^* = 22$  in the SDSS fields that cover A1577, as well as the color predictions of elliptical galaxies at different redshifts (*triangles*; Fukugita, Shimasaku, & Ichikawa 1995). The  $g^* - r^*$  color becomes redder from  $z = 0$  to  $z = 0.4$ , and  $r^* - i^*$  reddens monotonically. At  $z \sim 0.4$ , the 4000 Å break of an elliptical galaxy crosses the border between  $g^*$  and  $r^*$  bands and appears as a sharp turn in the color at this redshift (Fig. 5). By using this color change, we can reject foreground and background galaxies and select galaxies likely to be in a certain redshift range in the following way. This is a big advantage of having multicolor data, since optical

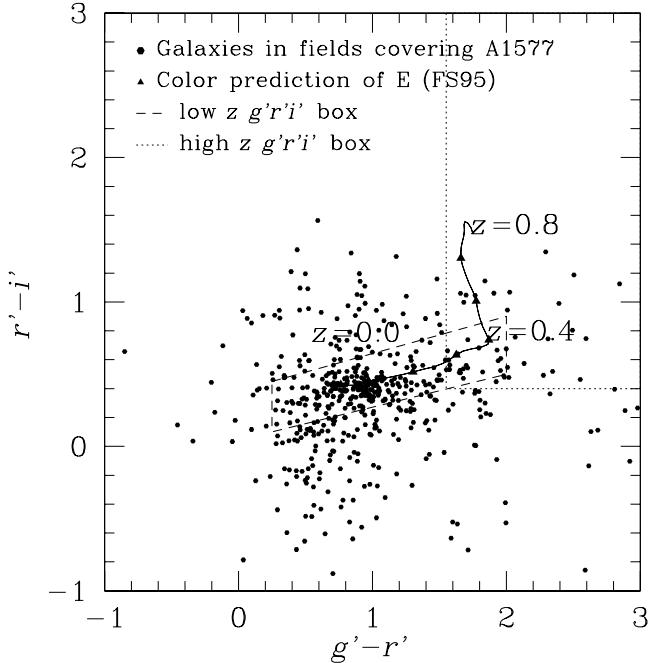


FIG. 5.—Plot of  $g^* - r^* - i^*$  color-color boxes to find galaxy clusters. The abscissa is the  $g^* - r^*$  color. The ordinate is  $r^* - i^*$  color. The low- $z$   $g^* - r^* - i^*$  box is drawn with dashed lines. The high- $z$   $g^* - r^* - i^*$  box is drawn with dotted lines. Galaxies brighter than  $r^* = 22$  in the SDSS fields ( $\sim 8.3 \times 10^{-2}$  deg $^2$ ) that cover A1577 are plotted with small dots. The triangles show the color prediction for elliptical galaxies (Fukugita et al. 1995).

cluster finders have suffered chance projections of galaxies in the sky. To select galaxies with similar colors, we divide the  $g^* - r^*$  versus  $r^*$  color-magnitude diagram into 11 bins. The bins are shown in Figure 2 as horizontal dashed lines. The bins are not tilted, because the tilt is almost negligible in the SDSS bands (see above) and because we wish to minimize the assumptions used for cluster selection. Any specific color bin reflects the redshifts of the cluster: blue color bins represent low-redshift clusters, while red bins represent higher redshift clusters. We use two bins as one color cut in order to produce overlap in the color cuts; the cut is shifted by one bin each time we step to a higher redshift (redder cut).

Similarly, we use 10 color cuts in both  $r^* - i^*$  (Fig. 3, *dotted lines*) and  $i^* - z^*$  (Fig. 4, *dotted lines*). The widths of the bins in  $g^* - r^*$ ,  $r^* - i^*$ , and  $i^* - z^*$  are 0.2, 0.1, and 0.1 mag, respectively. The width of the  $r^* - i^*$  and  $i^* - z^*$  bins is smaller than the  $g^* - r^*$  width because the colors of elliptical galaxies have less scatter in  $r^* - i^*$  and  $i^* - z^*$  than in  $g^* - r^*$ . The above color cuts, in the three colors, are applied independently. Galaxies that have color errors larger than the size of the color bin are rejected. The standard color errors estimated by the SDSS reduction software at  $r^* = 21.5$  (the limiting magnitude used in the CE method) are  $0.20 \pm 0.09$ ,  $0.16 \pm 0.06$ , and  $0.26 \pm 0.1$  in  $g^* - r^*$ ,  $r^* - i^*$ , and  $i^* - z^*$ , respectively. In  $g^* - r^*$  and  $r^* - i^*$ , the color error is smaller than the size of the color-cut box. In  $i^* - z^*$ , the color error at  $r^* = 21.5$  is slightly larger than the size of the color-cut boxes (0.2 mag); at  $r^* = 20.5$ , however, the error in  $i^* - z^*$  is  $0.11 \pm 0.05$ .

In Figures 6 and 7, we demonstrate the effect of the color cut. Dots are the galaxies within  $2.7'$  ( $1.5 h^{-1}$  Mpc at  $z = 0.37$ ) from the center of RX J0256.5+0006 (Romer

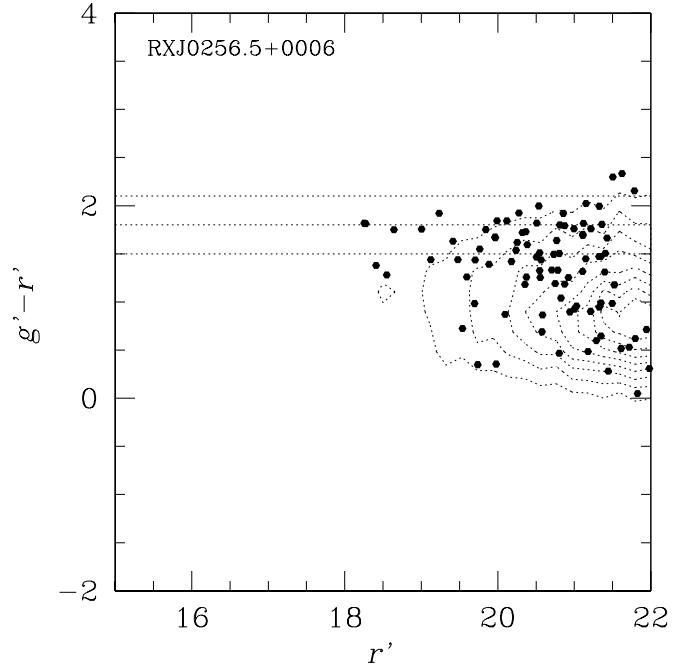
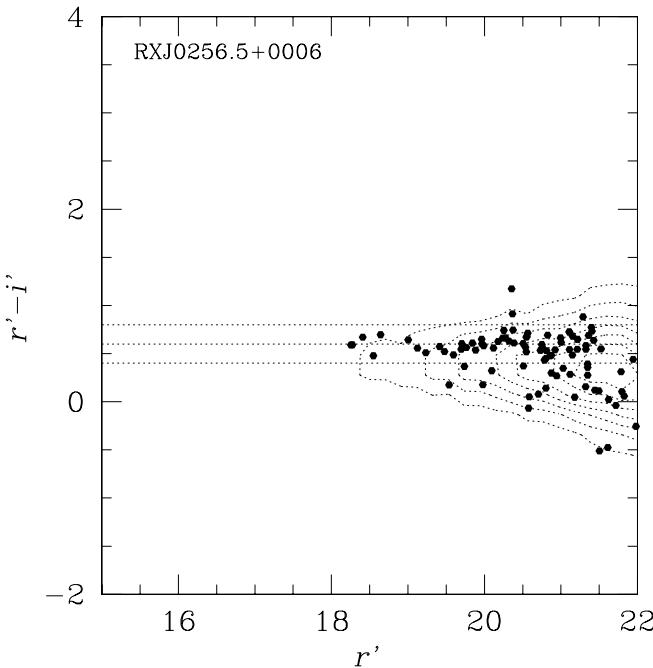


FIG. 6.—Example color cut capturing the color-magnitude relation. Galaxies within a  $1.5 h^{-1}$  Mpc aperture around RX J0256.5+0006 ( $z = 0.36$ ) are plotted as dots. The distribution of all the galaxies in the SDSS commissioning data is drawn as contours. The  $g^* - r^*$  color cut successfully captures the red sequence of RX J0256.5+0006 and removes foreground galaxies bluer than the sequence.

FIG. 7.—Same as Fig. 6, but for  $r^* - i^*$ 

et al. 2001). No background or foreground correction is applied. Contours represent the distribution of all the galaxies in the SDSS imaging data. The corresponding color cuts to the redshift of the cluster are drawn in each figure. In each case, the color cuts capture the red sequence of RX J0256.5+0006 successfully and reject foreground galaxies, as designed. In fact, we show in Table 2 the fraction of galaxies inside the color cut for both inside and outside of the cluster region. As shown in Figures 6 and 7, the fraction in the color cut indeed increases dramatically, from 13.5% to 36.9% in the  $g^* - r^*$  cut and from 42.4% to 62.1% in the  $r^* - i^*$  cut. The efficiency of the color cut increases as we see higher redshift apart from the foreground color distribution of galaxies. The top left panel in Figure 8 shows the galaxy distribution of the SDSS commissioning data around RX J0256.5+0006 before applying any cut. The top right panel shows the galaxy distribution after applying the  $g^* - r^*$  color cut at the cluster redshift, illustrating the color-cut enhancement of the cluster (see also Fig. 9).

### 3.2. Color-Color Cut

When more than two colors are available, it is more effective to select galaxies in color-color space. We thus added

TABLE 2

FRACTION OF GALAXIES INSIDE THE COLOR CUT FOR BOTH INSIDE AND OUTSIDE OF RX J0256.5+0006

Color Cut	In Cluster Region (%)	Outside Cluster Region (%)
$g^* - r^*$ .....	$36.9^{+7.0}_{-6.0}$	$13.57 \pm 0.03$
$r^* - i^*$ .....	$62.1^{+8.8}_{-7.7}$	$42.35 \pm 0.06$
$i^* - z^*$ .....	$59.2^{+8.6}_{-7.6}$	$44.55 \pm 0.06$
$g^* - r^* - i^*$ .....	$58.3^{+8.5}_{-7.6}$	$48.77 \pm 0.06$
$r^* - i^* - z^*$ .....	$76.7^{+10.7}_{-7.7}$	$65.68 \pm 0.07$
$g^* - r^* - i^*$ , high $z$ .....	$29.1^{+6.3}_{-5.3}$	$10.86 \pm 0.03$
$r^* - i^* - z^*$ , high $z$ .....	$6.8^{+3.7}_{-2.5}$	$9.94 \pm 0.02$

four additional color-color-cut boxes to enhance the contrast of galaxy clusters. The cuts are low- $z$  and high- $z$  boxes in  $g^* - r^* - i^*$  space and in  $r^* - i^* - z^*$  space, as shown in Figures 10 and 11. These color boxes are based on the fact that cluster galaxies concentrate in specific regions in color-color space (Dressler & Gunn 1992). In Figure 10, we show the  $g^* - r^*$  versus  $r^* - i^*$  color-color diagram of A168 for spectroscopically confirmed member galaxies (Katgert et al. 1998) brighter than  $r^* = 21$ . The low- $z$   $g^* - r^* - i^*$  color-color-cut box is shown by dashed lines, and the high- $z$   $g^* - r^* - i^*$  color-color-cut box is shown by the dotted lines. The triangles present the color prediction as a function of redshift for elliptical galaxies ( $\Delta z = 0.1$ ; Fukugita et al. 1995). The scatter in the plots comes from the mixture of different types of morphology. Similar results are shown in Figure 11 for the  $r^* - i^* - z^*$  color-color diagram of A168. Member galaxies of A168 ( $z = 0.044$ ; Struble & Rood 1999) are well centered in the low- $z$   $g^* - r^* - i^*$  and  $r^* - i^* - z^*$  boxes.

Figure 5 is the  $g^* - r^* - i^*$  color-color diagram of galaxies (brighter than  $r^* = 22$ ) in the SDSS fields covering A1577 ( $z = 0.14$ ). The low- $z$  and high- $z$  color-color-cut boxes are also shown. The triangles show the color prediction for elliptical galaxies. Figure 12 represents similar results in the  $r^* - i^* - z^*$  color-color space for the same field. Even though both cluster members and field galaxies are included in the plot, the concentration of cluster galaxies inside the low- $z$  boxes is clearly seen.

The color-color cuts are made based on the spectroscopic observation of Dressler & Gunn (1992) and the color prediction for elliptical galaxies (Fukugita et al. 1995). We reject galaxies that have standard color errors larger than the size of the color-color boxes. The standard color errors at  $r^* = 21.5$  (the limiting magnitude of the CE method) are  $0.20 \pm 0.09$ ,  $0.16 \pm 0.06$ , and  $0.26 \pm 0.1$  in  $g^* - r^*$ ,  $r^* - i^*$ , and  $i^* - z^*$ , respectively. The smallest size of the color-color boxes is the  $r^* - i^*$  side of the low- $z$   $g^* - r^* - i^*$  box, which is 0.34 in  $r^* - i^*$ . The standard color error is well within the color-cut boxes even at  $r^* = 21.5$ . In Figure 9, the top left panel shows the galaxy distribution in the SDSS commissioning data over  $23.75 \text{ deg}^2$  before applying any cut. The top right panel shows the galaxy distribution after applying the  $g^* - r^* - i^*$  color-color cut. Table 2 shows the fraction of galaxies inside the color cut for both inside and outside of RX J0256.5+0006 ( $z = 0.36$ ). Indeed, the fraction of galaxies in the color cut increases from 48.8% to 58.3% in the  $g^* - r^* - i^*$  cut and from 65.7% to 76.7% in the  $r^* - i^* - z^*$  cut. Since the color cuts have overlaps at  $z \sim 0.4$ , the  $g^* - r^* - i^*$  high- $z$  cut also increases somewhat.

We thus use 30 color cuts and four color-color cuts independently to search for clusters. We then merge the 34 cluster candidate lists into a final cluster catalog. Because of the star-galaxy separation limit, we do not use galaxies fainter than  $r^* = 21.5$ . The only main assumptions made in the CE detection method are the above color cuts.

In Figure 13, we plot the color prediction for galaxies with the evolving model with star formation (Pentagons) and the same model without star formation (squares) from  $z = 0$  to  $z = 0.6$  (PEGASE model; Fioc & Rocca-Volmerange 1997). Triangles show the color prediction for elliptical galaxies ( $\Delta z = 0.1$ ; Fukugita et al. 1995). The model galaxies with star formation are the extreme star-forming galaxies. We plot spectroscopic galaxies as green dots. Black dots are the galaxies around A1577, for reference. Although the evolving model goes outside of the high- $z$  color-cut box

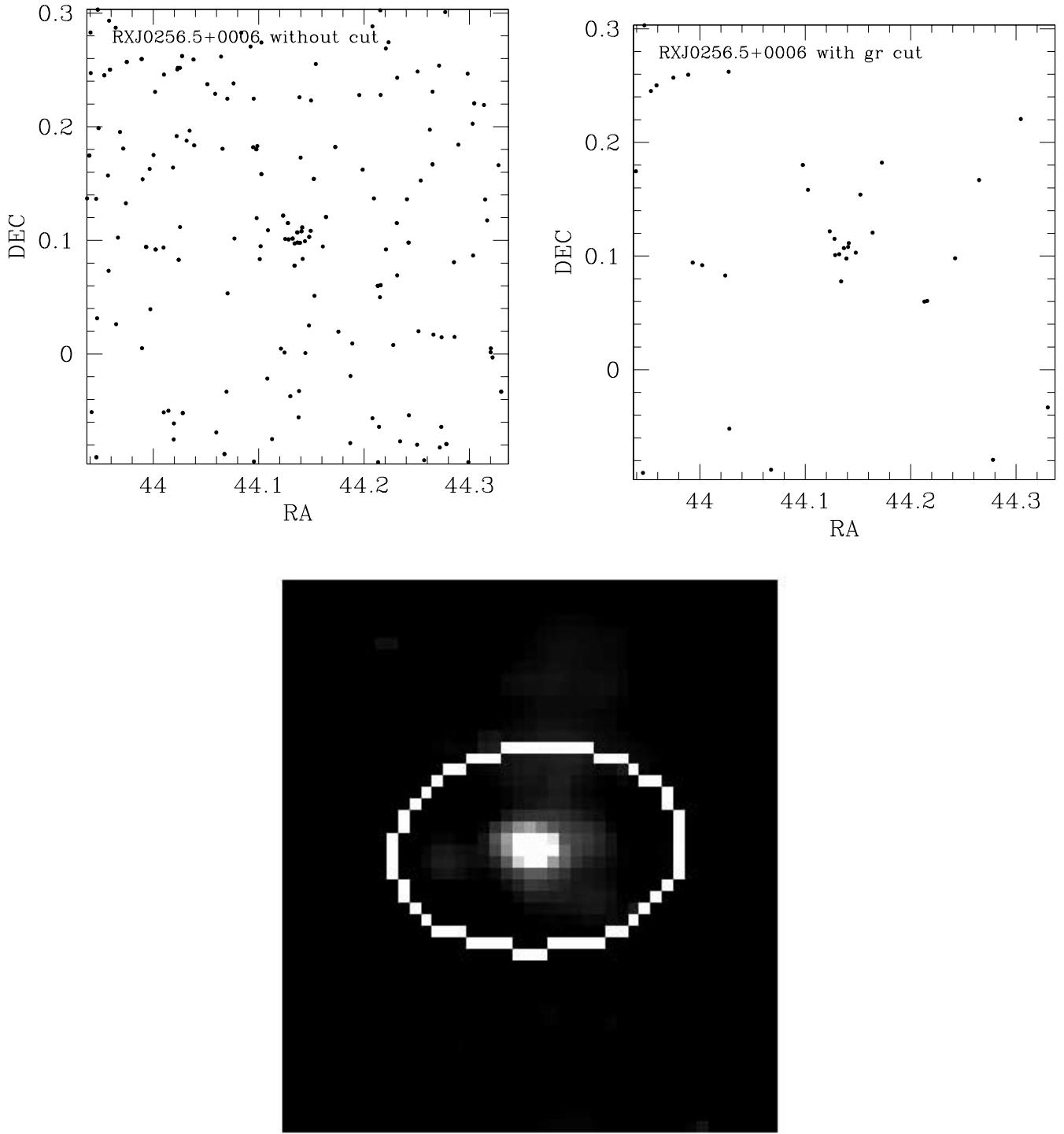


FIG. 8.—Distribution of galaxies brighter than  $r^* = 20.0$  around RX J0256.5+0006. *Top left*, the distribution before applying any cut; *top right*, the distribution after applying the  $g^* - r^*$  color cut; *bottom*, the enhanced map. The color cut removes foreground and background galaxies as designed. RX J0256.5+0006 is successfully detected (white circle).

at  $z \sim 0.6$ , the CE method is designed to detect galaxy clusters if enough red galaxies (*squares*) are in the color cut by weighting the galaxies with similar color. In fact, spectroscopic galaxies (*green dots*;  $0.4 < z \leq 0.5$ ) are well within the high- $z$  box (100 galaxies with  $z > 0.4$  were randomly taken from the SDSS spectroscopic data). As seen in the real catalog in § 3.6, because of the magnitude limit of SDSS, it is difficult to find many clusters beyond  $z \sim 0.4$ . On the other hand, if we move the color cut bluer, we increase the contamination from  $z \sim 0.3$  galaxies (which are well within

the magnitude limit of SDSS). This is how the color-cut box was optimized.

### 3.3. Enhancement Method

After applying the color cuts, we use a special method to enhance the signal-to-noise ratio of clusters further. First, we find all pairs of galaxies within  $5'$ ; this scale corresponds to the size of galaxy clusters at  $z \sim 0.3$ . Selecting larger separations blurs high- $z$  clusters, while smaller separations

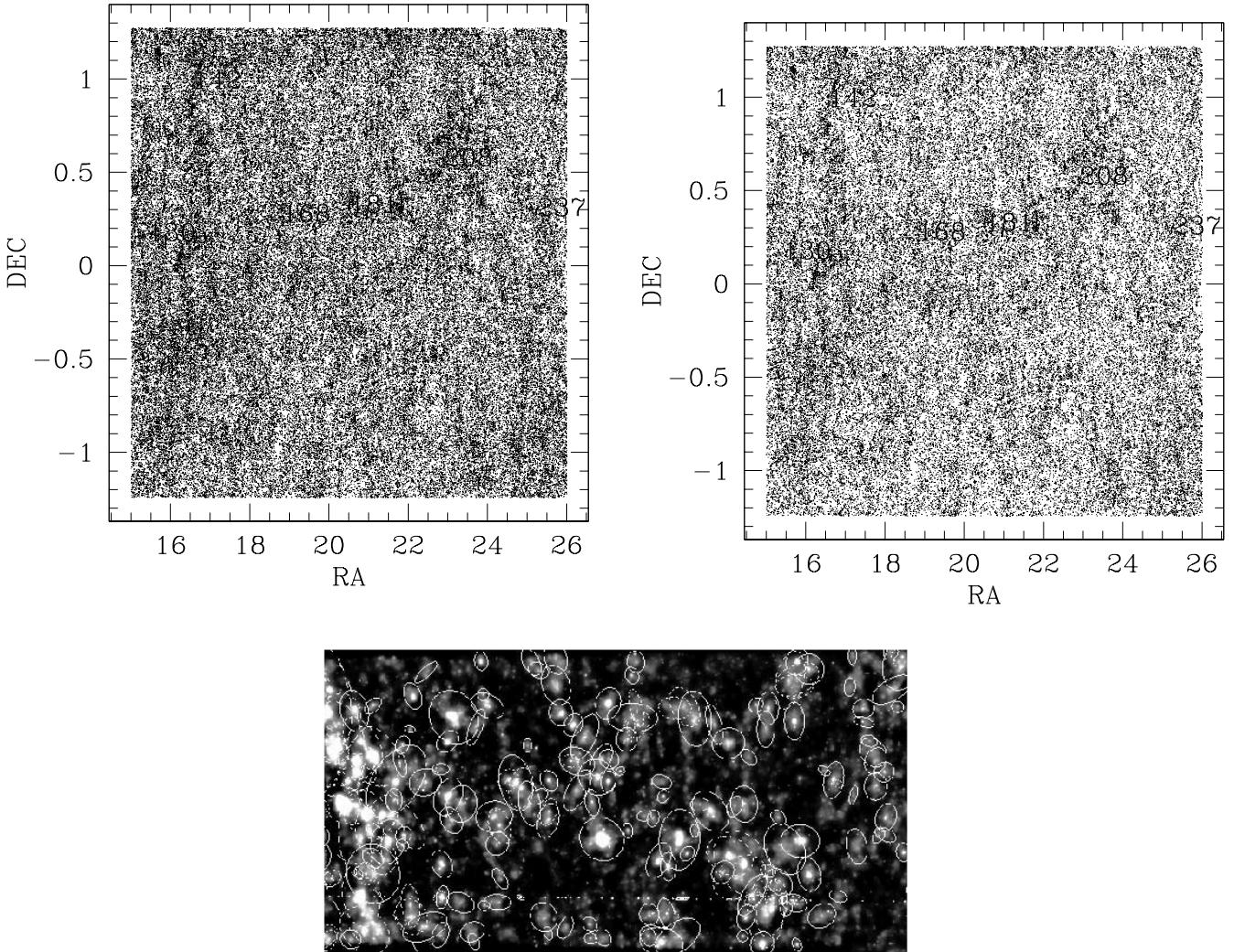


FIG. 9.—Distribution of galaxies brighter than  $r^* = 21.5$ . *Top left*: The distribution before applying any cut. *Top right*: The distribution after applying the  $g^*-r^*-i^*$  color-color cut. The numbers show the positions of Abell clusters. *Bottom*: The enhanced map in the  $g^*-r^*-i^*$  color-color cut. Detected clusters are circled.

weaken the signal of low- $z$  clusters. We empirically investigated several separations and found  $5'$  to be a good mean value. We then calculate the angular distance and color difference of each pair of galaxies. We distribute a Gaussian cloud around the center position of each pair. The width of a Gaussian cloud is the angular separation of the pair, and the volume of a Gaussian cloud is given by its weight  $W$ , which is calculated as

$$W = \frac{1}{\Delta r + 1''} \frac{1}{\Delta(g^* - r^*)^2 + 2.5 \times 10^{-3}}, \quad (1)$$

where  $\Delta r$  is the angular separation between the two galaxies and  $\Delta(g^* - r^*)$  is their color difference. Small softening parameters (empirically determined) are added in the denominator of each term to avoid values becoming infinite. This enhancement method provides stronger weights to pairs that are closer both in angular space and in color space and thus are more likely to appear in galaxy clusters. Gaussian clouds are distributed in  $30'' \times 30''$  cells on the sky. The  $30''$  cells are small compared with sizes of galaxy clusters (several arcminutes at  $z \sim 0.5$ ).

An enhanced weighted map of high-density regions is obtained by summing the Gaussian clouds. The bottom

panels of Figures 8 and 9 present such enhanced maps of the region in the top panels. RX J0256.5+0006 is successfully enhanced in Figure 8. Figure 9 illustrates how the CE method finds galaxy clusters. The advantage of this enhancement method in addition to the color cuts is that it makes full use of color concentration of cluster galaxies. The color cuts are used to reduce foreground and background galaxies and to enhance the signal of clusters. Since the color-magnitude relation of cluster galaxies is frequently tighter than the width of our color cuts, the use of the second term in equation (1)—the inverse square of the color difference—further enhances the signal of clusters, despite the larger width of the color cuts. Another notable feature is that the enhancement method is adaptive, that is, larger separation pairs have large Gaussian clouds and small-separation pairs have sharp, small Gaussian clouds. In this way, the enhancement method naturally fits to any region with any number density of galaxies in the sky. It is also easy to apply it to data from another telescope with different depth and different galaxy density. Another benefit of the enhancement method is that it includes a smoothing scheme, and thus conventional detection methods commonly used in the astronomical community can be used to

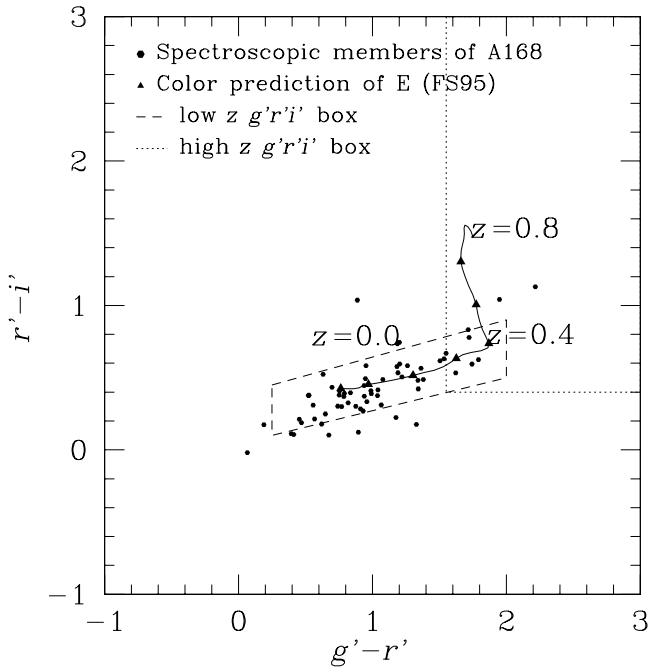


FIG. 10.—Color-color diagram of spectroscopically confirmed member galaxies of A168. The abscissa is  $g^* - r^*$  color. The ordinate is  $r^* - i^*$  color. The low- $z$   $g^* - r^* - i^*$  box is drawn with dashed lines. The high- $z$   $g^* - r^* - i^*$  box is drawn with dotted lines. Galaxies brighter than  $r^* = 21$  that matched up with the spectroscopically confirmed galaxies (Katgert et al. 1998) are plotted with dots. The triangles show the color prediction for elliptical galaxies (Fukugita et al. 1995).

detect clusters in the enhanced map. The enhancement method uses the angular separation in the  $W$ -values. This might bias our catalog against nearby clusters ( $z < 0.1$ ), which have a large angular extent (and thus are given less  $W$ ). However, these nearby clusters are already well documented in existing catalogs; these nearby clusters will also be well sampled in the SDSS spectroscopic survey with fiber

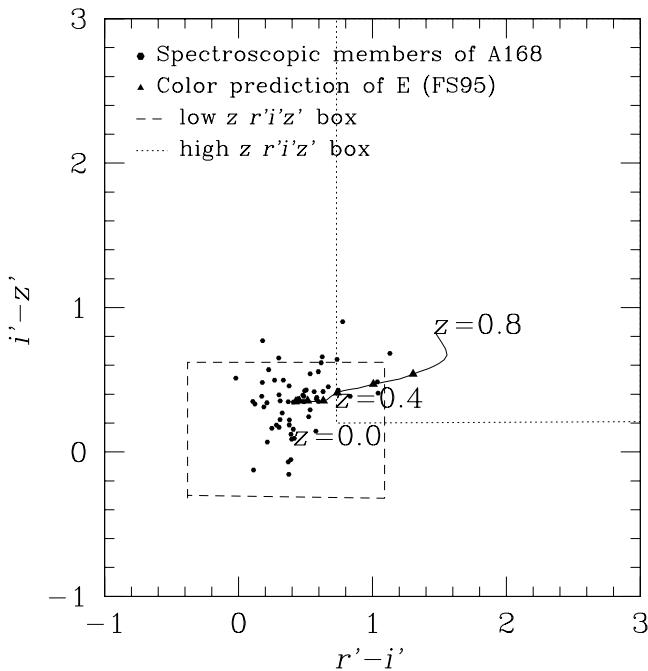


FIG. 11.—Same as Fig. 10, but for  $r^* - i^* - z^*$

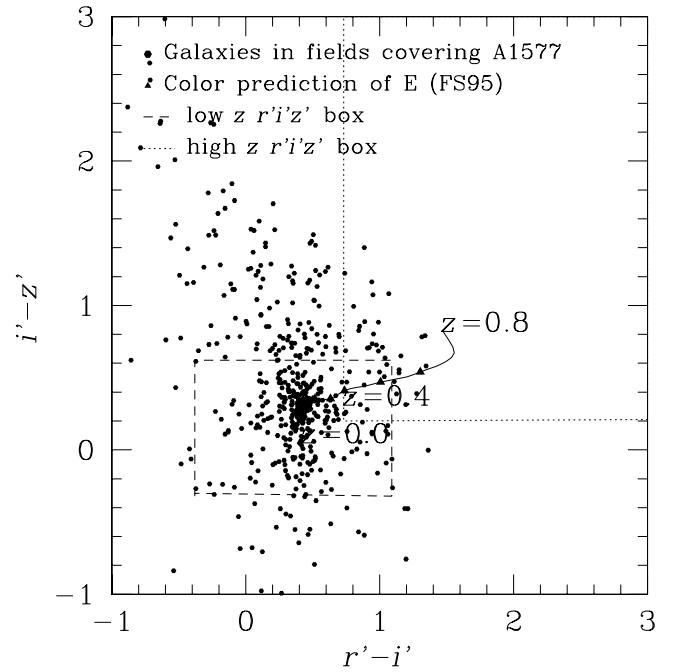


FIG. 12.—Plot of  $r^* - i^* - z^*$  color-color boxes to find galaxy clusters. The abscissa is  $r^* - i^*$  color. The ordinate is  $i^* - z^*$  color. The low- $z$   $r^* - i^* - z^*$  box is drawn with dashed lines. The high- $z$   $r^* - i^* - z^*$  box is drawn with dotted lines. Galaxies brighter than  $r^* = 22$  in the SDSS fields ( $\sim 8.3 \times 10^{-2} \text{ deg}^2$ ) that cover A1577 are plotted with small dots. The triangles show the color prediction for elliptical galaxies (Fukugita et al. 1995).

redshifts and will thus be detected in the SDSS three-dimensional cluster selection (the CE method is intended to detect clusters using imaging data only). These nearby clusters do not have a significant effect on angular or redshift-space correlations, because the number of such clusters is a small fraction of any large volume-limited sample.

### 3.4. Detection

We use SExtractor (Bertin & Arnouts 1996) to detect clusters from the enhanced map discussed in § 3.3. SExtractor identifies high-density peaks above a given threshold, measuring the background and its fluctuation locally. The threshold selection determines the number of clusters obtained. A high threshold selects only the richer clusters. We tried several thresholds, examining the colored image, color-magnitude diagram, and color-color diagram of the resulting cluster catalog. The effect of changing the threshold is summarized in Table 3. The number of clusters detected is not very sensitive to the threshold.<sup>8</sup> Based on the above analysis, we have selected the threshold to be 6 times the background fluctuation, a threshold that yields a large number of clusters while still keeping the spurious detection rate low.

Monte Carlo simulations are sometimes used to decide the optimal threshold at which the most true clusters are recovered while the spurious detection rate is still low. However, the simulations reflect an ideal situation, and they are inevitably different from true data; for example, a uniform

<sup>8</sup> The number of detections goes up and down with increasing  $\sigma$  because the following two effects cancel each other out: (1) lower threshold detects faint sources and thus increases the number of detections; (2) higher threshold deblends the peaks and increases the number of detections.

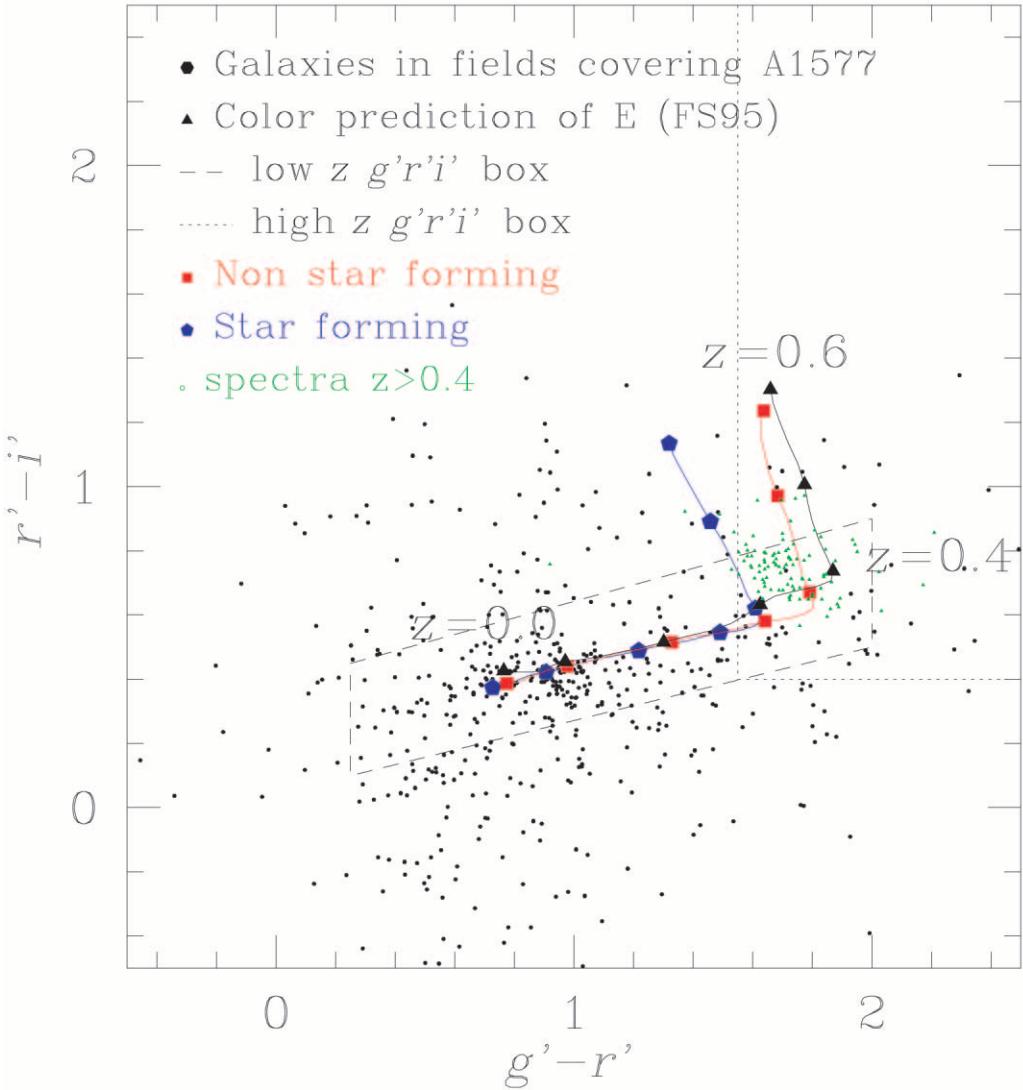


FIG. 13.—Evaluation of the high- $z$  color cut. Triangles show the color prediction for elliptical galaxies (Fukugita et al. 1995). Squares show the color prediction for non-star-forming galaxies from the PEGASE model (Fioc & Rocca-Volmerange 1997). Pentagons show the color prediction for star-forming galaxies from the PEGASE model. Black dots are the galaxies around A1577. The high- $z$  and low- $z$  color cuts are drawn with dotted and dashed lines, respectively.

background cannot represent the true galaxy distribution with its large-scale structure. There are always clusters that do not match the radial profile or luminosity function assumed in Monte Carlo simulations, and this may affect the optimization of the threshold. The optimal threshold in a Monte Carlo simulation differs from that in the real data. Therefore, we select the threshold empirically using the actual data and later derive the selection function using Monte Carlo simulation.

TABLE 3  
SIGMA-CUT TEST

$\sigma$	Detections
2.....	402
4.....	437
6.....	453
8.....	434
10.....	415

At high redshifts ( $z > 0.4$ ), the number of galaxies within the color cuts is small; therefore, the rms of the enhanced map is generally too low and the clusters detected at high redshift have unusually high signal. To avoid such spurious detections, we applied another threshold at a maximum absolute flux of 1000 in the enhanced map<sup>9</sup>. Spurious detections with high signal would generally have low values because they are not true density peaks. This maximum absolute flux threshold can thus reject spurious detections. The value is determined by investigating the image and the color-magnitude and color-color diagrams of the detected clusters and iterating the detection with different values of the maximum absolute flux threshold. The effect of changing the absolute flux threshold is summarized in Table 4.

To secure the detection further, at all redshifts we demand at least two detections in the 34 cuts. This is because the

<sup>9</sup> FLUX\_MAX + BACKGROUND = 1000, where FLUX\_MAX and BACKGROUND are the parameters of SExtractor. FLUX\_MAX + BACKGROUND is the highest value in the pixels within the cluster. It is an absolute value and not affected by rms value.

TABLE 4  
EFFECT OF CHANGING THE ABSOLUTE  
FLUX THRESHOLD

Maximum Flux	Detections
500.....	890
750.....	655
1000.....	464
1500.....	260
2000.....	10

cluster galaxies have color concentrations in all  $g^*-r^*$ ,  $r^*-i^*$ , and  $i^*-z^*$  colors; real clusters should thus be detected in at least two color cuts.

### 3.5. Merging

We apply the procedure of cut, enhance, and detect to all of the 34 color cuts (30 color cuts plus four color-color cuts) independently. After creating the 34 cluster lists, we merge them into one cluster catalog. We regard the detections within 1/2 as one cluster. To avoid two clusters with different redshifts being merged into one cluster as a result of chance alignment, we do not merge clusters that are detected in two color cuts of the same band unless the successive color cuts both detect it.

An alternative way to merge clusters would be to merge only those clusters that are detected in the consistent color cut in all  $g^*-r^*$ ,  $r^*-i^*$ , and  $i^*-z^*$  colors, using the model of the elliptical galaxy colors. However, the catalog would be biased against clusters that have colors different from the model elliptical galaxies. In order to minimize the assumptions on cluster properties, we treat the three-color space,  $g^*-r^*$ ,  $r^*-i^*$ , and  $i^*-z^*$ , independently.

### 3.6. Redshift and Richness Estimation

We estimate the redshift and richness of each cluster as follows: Instead of using the same richness estimator as Abell, we count the number of galaxies inside the detected cluster radius that lie in the 2 mag range ( $r^*$  band) from  $m_3$  (the third brightest galaxy) to  $m_3 + 2$  (CE richness). The difference from Abell's estimation is that he used a fixed radius of  $1.5 h^{-1}$  Mpc. Here we use the detection radius of the cluster detection algorithm, which can be larger or smaller than the Abell radius, typically slightly smaller than  $1.5 h^{-1}$  Mpc. The background galaxy count is subtracted using the average galaxy counts in the SDSS commissioning data.

For the redshift estimates, we use the strategy of the “maxBCG” technique (Annis et al. 2002). We count the number of galaxies within the detected radius that are brighter than  $M_{r^*}^* = -20.25$  for a given assumed redshift and are within a color range of  $\pm 1$  mag in  $g^*-r^*$  around the color prediction for elliptical galaxies (Fukugita et al. 1995). This is determined in estimated redshift steps of  $\delta z = 0.01$ . After subtracting the average background number counts from each bin, the redshift of the bin that has the largest number of galaxies is taken as the estimated cluster redshift. The estimated redshifts are calibrated using the spectroscopic redshifts from the SDSS spectroscopic survey. Our redshift estimation depends on the model of Fukugita et al. (1995), but the differences from other models are not so significant, as seen in the difference between the squares (PEGASE model) and triangles (Fukugita et al. 1995) in

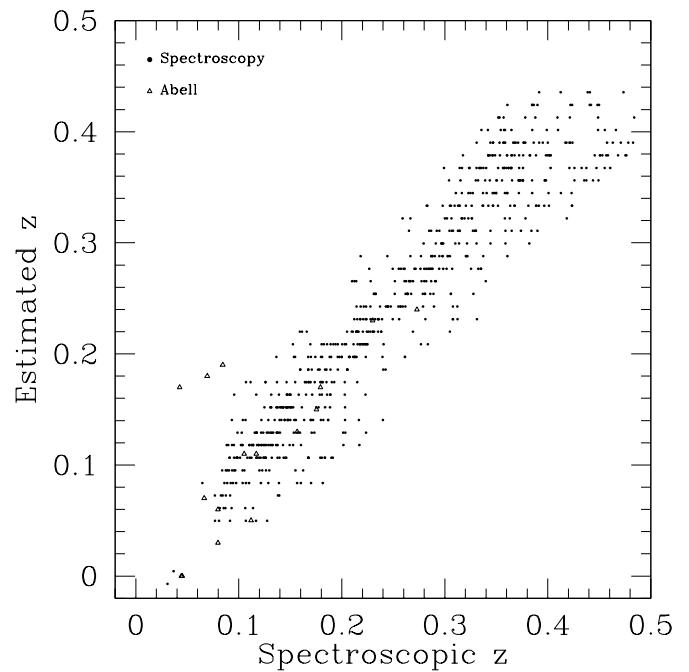


FIG. 14.—Redshift estimation accuracy. The estimated redshifts are plotted against spectroscopic redshifts. Abell clusters are plotted with triangles. Dots are the redshifts from SDSS spectroscopic galaxies. Extensive outliers,  $\delta z > 0.1$ , are removed. The dispersion is 0.0147 for  $z < 0.3$  and 0.0209 for  $z > 0.3$ .

Figure 13. If a cluster has enough elliptical galaxies, the redshift of the cluster is expected to be well measured. If a cluster is however dominated only by spiral galaxies, as seen in the difference between triangles and pentagons, the redshift of the cluster will be underestimated.

Figure 14 shows the redshift accuracy of the method. The estimated redshifts are plotted against observed redshifts from the spectroscopic observations. The redshift of the SDSS spectroscopic galaxy within the detected radius and with the nearest spectroscopic redshift to the estimation is adopted as the real redshift. In the fall equatorial region, 699 clusters have spectroscopic redshifts. The correlation between true and estimated redshifts is very good: the rms scatter is  $\delta z = \pm 0.0147$  for  $z < 0.3$  clusters and  $\delta z = \pm 0.0209$  for  $z > 0.3$  clusters. Triangles show 15 Abell clusters measured with available spectroscopic redshifts; there are three outliers at low spectroscopic redshifts. CE counterparts for these three clusters all have very small radii of several arcminutes. Since these Abell clusters are at  $z < 0.1$ , the discrepancy is probably not in the redshift estimation but rather in the detection radius. We thus construct the SDSS CE galaxy cluster catalog, containing 4638 galaxy clusters.<sup>10</sup>

## 4. MONTE CARLO SIMULATION

In this section, we examine the performance of the CE method and determine the selection function using Monte Carlo simulations. We also perform false-positive tests.

<sup>10</sup> The catalog is available at the following Web site:  
<http://astrophysics.phys.cmu.edu/~tomo>.

#### 4.1. Method

We perform Monte Carlo simulations both with a real background using the SDSS commissioning data and with a shuffled background (explained below). For the real background, we randomly choose a  $1 \text{ deg}^2$  region of the SDSS data with seeing better than  $1''.7$ .<sup>11</sup> For the shuffled background, we redistribute all the galaxies in the above  $1 \text{ deg}^2$  of SDSS data randomly in position but keep their colors and magnitudes unchanged.

Then we place artificial galaxy clusters on these backgrounds. We distribute cluster galaxies randomly using a King profile (King 1966; Ichikawa 1986) for the radial density, with a concentration index of 1.5 and cutoff radius of  $2.1 h^{-1} \text{ Mpc}$ , which is the size of A1577 (Struble & Rood 1987). For colors of the artificial cluster galaxies, we use the color and magnitude distribution of A1577 (at  $z \sim 0.14$ , richness  $\sim 1$ ) as a model. We choose the SDSS fields that cover the entire A1577 area and count the number of galaxies in each color bin. The size of the bins is 0.2 mag in both colors and magnitude. The color and magnitude distribution spans a four-dimensional space,  $r^*, g^* - r^*, r^* - i^*$ , and  $i^* - z^*$ . We count the number of field galaxies using the same size fields near A1577 and subtract the distribution of field galaxies from the distribution of galaxies in the A1577 fields. The resulting color distribution is used as a model for the artificial galaxy clusters. Galaxy colors are selected randomly so that they reproduce the overall color distribution of A1577. The distribution is linearly interpolated when allocating colors and magnitudes to the galaxies.

For the high-redshift artificial clusters, we apply a  $k$ -correction and the color prediction for elliptical galaxies from Fukugita et al. (1995). For the color prediction, only the color difference, not the absolute value, is used. Galaxies that become fainter than  $r^* = 21.5$  are not used in the CE method.

#### 4.2. Monte Carlo Results

First we run a Monte Carlo simulation with only the background, without any artificial clusters, in order to measure the detection rate of the simulation itself. The bias detection rate is defined as the percentile in which any detection is found within  $1/2$  from the position where we later place an artificial galaxy cluster. The main reason for the false detection is that a real cluster sometimes comes into the detection position where an artificial cluster is later placed. This is not a false detection by the CE method but rather the noise in the simulation itself. The bias detection rate with the real SDSS background is 4.3%. This is small relative to the actual cluster detection rate discussed below. The bias detection rate using the shuffled background is lower (as expected), 2.4%. We run Monte Carlo simulations with a set of artificial clusters with redshifts ranging from  $z = 0.2$  to  $z = 0.6$ , and with richesses of  $N_{\text{gal}} = 40, 60, 80$ , and 100 at each redshift. ( $N_{\text{gal}}$  is the number of galaxies within  $2.1 h^{-1} \text{ Mpc}$  input to a cluster, whose magnitudes are  $r^* < 21.5$  at the redshift of A1577. If a galaxy becomes fainter than  $r^* = 21.5$ , it is not counted in the CE detection

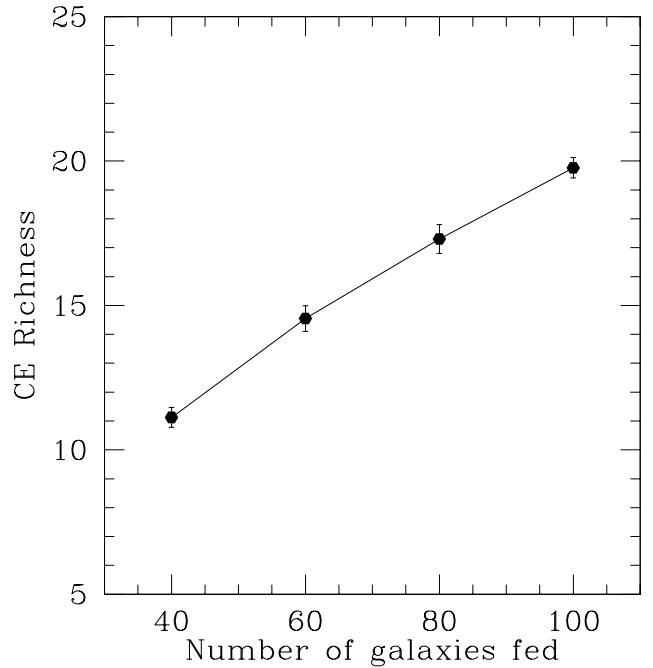


FIG. 15.— $N_{\text{gal}}$  vs. richness. The number of galaxies put in the artificial cluster is compared with richness (the number of galaxies within the detected radius whose magnitude is between the magnitude of the third brightest galaxy and the magnitude fainter by 2). The bars show  $1\sigma$  standard errors.

method even if it is included in  $N_{\text{gal}}$ .) For each set of parameters, the simulation is iterated 1000 times.

In Figure 15, we compare  $N_{\text{gal}}$  with cluster richness, where richness is defined (§ 3) as the number of galaxies within the 2 mag range below the third-brightest galaxy, located within the cluster radius that the CE method returns. The error bars are  $1\sigma$  standard errors.  $N_{\text{gal}} = 50$  corresponds to Abell richness class  $\sim 1$ .

Figure 16 shows the recovery rate in the Monte Carlo simulations on the real background. The percentage recovery rate is shown as a function of redshift. Each line represents different-richness input clusters,  $N_{\text{gal}} = 100, 80, 60$ , and 40 (top to bottom). Because the false detection rate in the simulation with real background is 4.3%, all the lines converge to 4.3% at high redshift. The detection rate drops suddenly at  $z = 0.4$  because, at this point, a large fraction of the cluster member galaxies are lost as a result of the magnitude limit of  $r^* = 21.5$ . Roughly speaking, it determines the depth of an SDSS cluster catalog. The  $N_{\text{gal}} = 80$  clusters are recovered  $\sim 80\%$  of the time to  $z < 0.3$ , dropping to  $\sim 40\%$  beyond  $z \sim 0.4$ . Clusters of the lowest richesses,  $N_{\text{gal}} = 40$ , are more difficult to detect, as expected. The detection rates of  $N_{\text{gal}} = 40$  clusters are less than 40% even at  $z = 0.3$ . The recovery rate for  $N_{\text{gal}} = 100$  at  $z = 0.2$  is not 100%. If we widen the detection radius from  $1/2$  to  $5/4$ , the recovery rate increases to 100%. Note that a radius of  $5/4$  is still small in comparison with the size of A1577:  $11'$  at  $z = 0.2$  (Struble & Rood 1987). The reason may be that a real cluster (in the real background) is located close to the artificial cluster, and the detected position may then be shifted by more than the detection radius ( $1/2$ ) away from the cluster.

Figure 17 shows the positional accuracy of the detected clusters in the Monte Carlo simulation with the real SDSS background, as a function of redshift and richness. The  $1\sigma$

<sup>11</sup> Though the SDSS survey criterion for seeing is better than  $1''.5$ , some parts of the SDSS commissioning data have seeing worse than  $2''.0$ . It is expected that the seeing is better than  $1''.5$  for all the data after the survey begins.

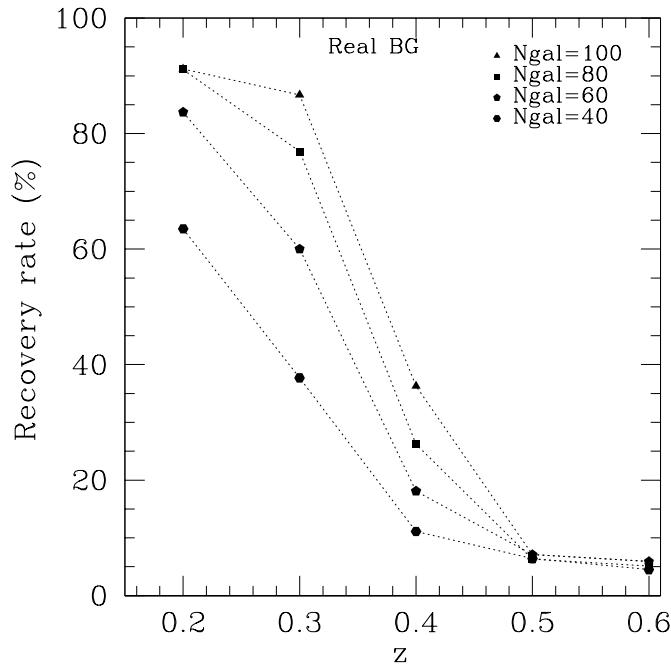


FIG. 16.—Recovery rate in Monte Carlo simulation with the real SDSS background plotted against redshift. The artificial clusters are added on the real SDSS background randomly chosen from the SDSS commissioning data. The detection is iterated 1000 times for each data point. Even at  $z = 0.5$ , an  $N_{\text{gal}} = 50$  cluster is detected with better than 82.5% probability.

positional errors of the detected clusters are shown. Note that since the CE method does not detect many clusters beyond  $z = 0.4$ , there is not much meaning to discussing the positional accuracy beyond  $z = 0.4$ . The positional accu-

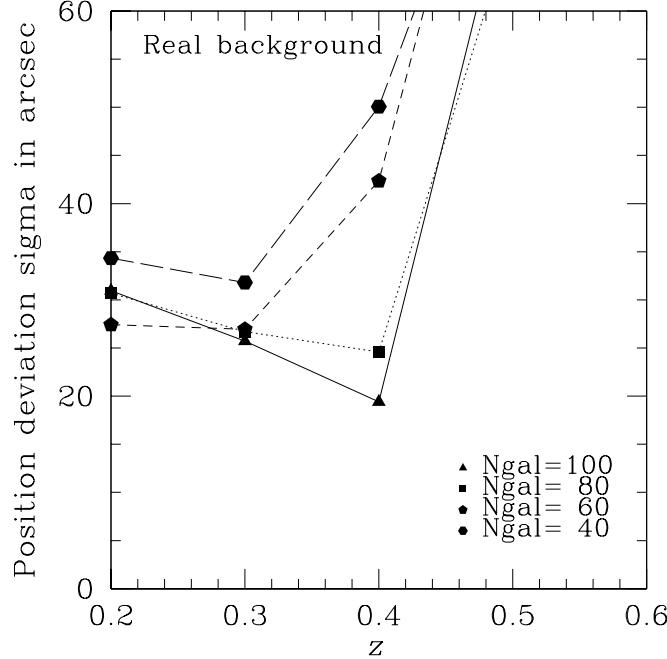


FIG. 17.—Positional accuracy with the real SDSS background. The positional accuracy is almost constant because a more distant cluster is more compact in angular space. Positional accuracy of  $\sim 0^{\circ}01$  is reasonable, considering that the mesh size of the enhancement method is  $30''$  ( $= 0^{\circ}0083$ ). The lack of some points at low richness and high redshift is due to the failure to fit using poor statistics.

racy is better than  $1'$  until  $z = 0.4$  in all richness ranges used. The deviation is nearly independent of the redshift because the high-redshift clusters are more compact than the low-redshift ones. This partially cancels the effect of losing more galaxies at high redshift as a result of the flux limit of the sample. The positional accuracy roughly corresponds to the mesh size of the enhancement method,  $30''$ . As expected, the positional accuracy is worse for high-redshift poor clusters ( $z = 0.4$  and  $N_{\text{gal}} \leq 60$ ). The statistics for these objects are also less good; the detection rates of  $N_{\text{gal}} = 60$  and 40 clusters are less than 20% at  $z = 0.4$ .

Figure 18 presents the recovery rate of artificial clusters in Monte Carlo simulations with the shuffled background. The recovery rates are slightly better than with the real background. Again, the recovery rates drop sharply at  $z = 0.4$ . The  $N_{\text{gal}} = 100$  clusters are recovered with  $\sim 90\%$  probability to  $z \sim 0.3$  and  $\sim 40\%$  to  $z \sim 0.4$ . The detection rate is slightly higher than with the real background. At  $z \leq 0.3$ ,  $N_{\text{gal}} > 40$  clusters are recovered at better than 40%. Figure 19 shows the positional accuracy of the detected clusters in the simulations (with shuffled background). The results are similar to these with the real data background. The positional accuracy is better than  $40''$  until  $z = 0.3$  for all richnesses.

#### 4.3. False-Positive Test

In order to test the false-positive rate, we prepared four sets of the data: (1) real SDSS data of  $25 \text{ deg}^2$ , (2) with positions of galaxies in the same  $25 \text{ deg}^2$  randomized (galaxy colors untouched), (3) with colors of the galaxies shuffled (galaxy positions untouched), and (4) with color shuffled and position smeared ( $5'$ ). Galaxy colors are randomized and positions are randomly distributed in such a way that galaxies still lie within  $5'$  of their original positions. This is intended to include large-scale structure without galaxy clusters. The results are shown in Figure 20. The solid line

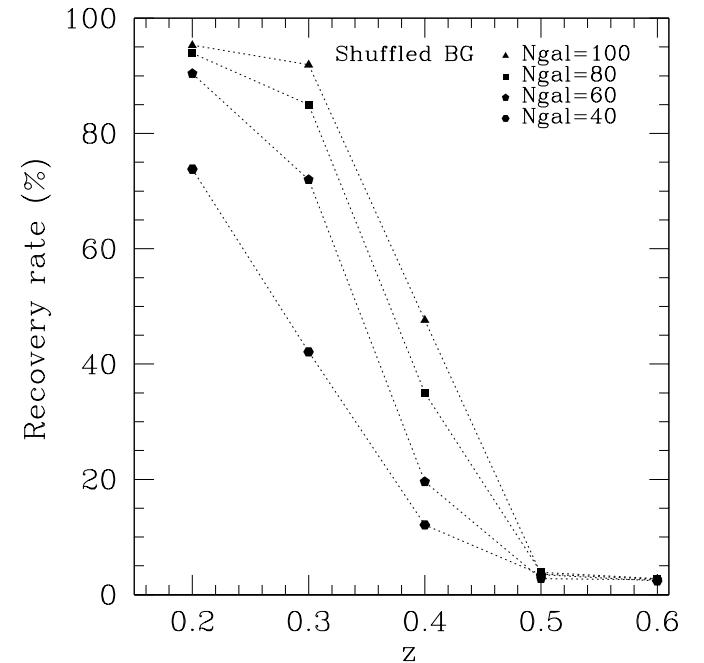


FIG. 18.—Same as Fig. 16, but for the shuffled background. Even at  $z = 0.5$ , an  $N_{\text{gal}} = 50$  cluster is detected with better than 80% probability.

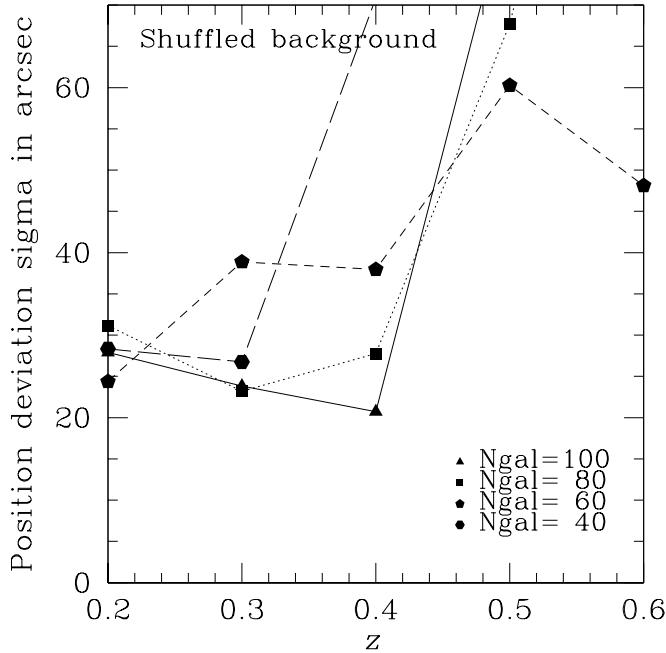


FIG. 19.—Same as Fig. 17, but for the shuffled background. Positional accuracy of  $\sim 0^{\circ}01$  is good, considering that the mesh size of the enhancement method is  $30''$  ( $= 0^{\circ}0083$ ).

represents the results with real data. The dotted line represents the results with position-shuffled data. The long-dashed line is for color-shuffled data. For color-shuffled data, we subtracted the detections in the real data because they still contain real clusters there. The fact that the color-shuffled data still detect many clusters is consistent with the

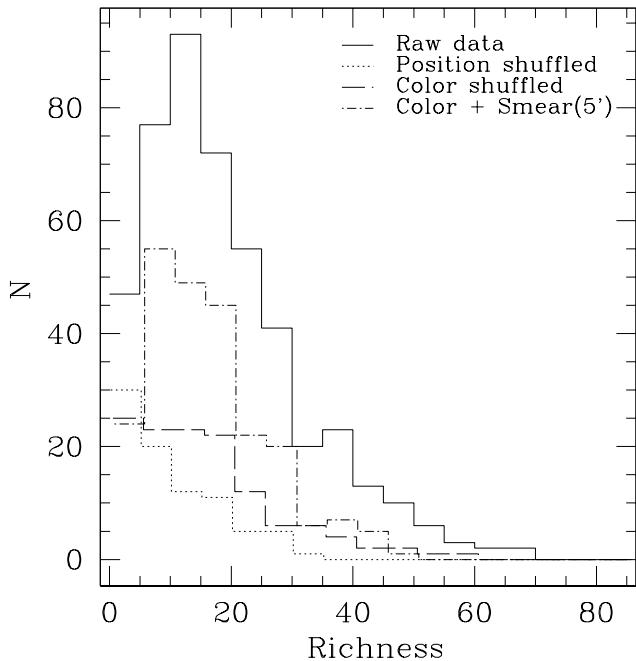


FIG. 20.—False-positive test. The detection test is performed using  $25 \text{ deg}^2$  of SDSS commissioning data. Solid line, results with real data; dotted line, results with position-shuffled data; long-dashed line: color-shuffled data subtracting the detection from the real data; dot-dashed line, color-shuffled smearing data.

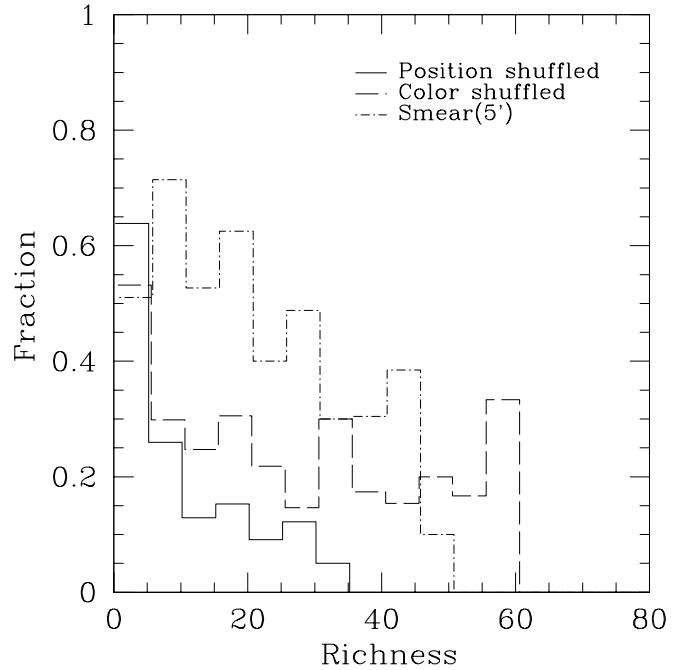


FIG. 21.—Same as Fig. 20, but with each line representing the ratio to the real data at the richness bin.

generous color cuts of the CE method. The dot-dashed line is for color-shuffled smearing data. In Figure 21, the ratio to the real data is plotted against CE richness. The promising fact is that not so many sources are detected from position-shuffled data. The rate with respect to the real data is below 20% at CE richnesses over 20. More points are detected from color-shuffled data and smearing data, but this does not mean that the false-positive rate of the CE method is as high as those values. Smearing data still have a structure bigger than  $5'$ , and they can be real clusters. Overall, our simulations show that for richness greater than 10, over 70% of CE clusters are likely to be real systems (as shown by the color- and position-shuffled simulations).

## 5. VISUAL INSPECTION

To investigate whether the detected clusters are true clusters or spurious detections, spectroscopic observations are necessary. Although large spectrometers that can observe the spectra of many galaxies at one time are becoming available (e.g., SDSS, 2dF), it is still time-consuming. Since the SDSS CE cluster catalog will have more than 100,000 galaxy clusters when the survey is complete, it is in fact impossible to spectroscopically confirm all the clusters in the catalog. As a preliminary check of our method, we visually inspect all the CE clusters within a given area (right ascension between  $16^{\circ}$  and  $25.5^{\circ}$  and declination between  $-1^{\circ}25$  and  $1^{\circ}25$ , totaling  $23.75 \text{ deg}^2$ ; the region in Fig. 9). A total of 278 CE galaxy clusters are located within this area (after removing clusters touching the region's borders). Out of the 278 CE galaxy clusters, we estimate that 10 are false detections. Since the strategy of the CE method is to detect every clustering of galaxies, we call every angular clustering of galaxies with the same color a successful detection here. (As we show in § 4.3, 30% of the clusters could be false detections, such as chance projections.)

TABLE 5  
10 FALSE CUT-AND-ENHANCE DETECTIONS

$\sigma$	CE Richness	$z$	Comment
4.80 .....	11	0.10	Looks like field
7.06 .....	1	0.04	Eight blue galaxies
7.85 .....	21	0.18	Looks like field
9.35 .....	7	0.00	Looks like field
11.0 .....	14	0.04	Looks like field
12.39 .....	31	0.22	Looks like field
16.74 .....	16	0.18	Looks like field
25.20 .....	7	0.00	Large galaxy
56.85 .....	8	0.04	Large galaxy
4848.39 .....	17	0.12	Large galaxy

NOTE.—The region used is  $16^\circ < \text{R.A.} < 25^\circ.5$ ,  $-1^\circ.25 < \text{decl.} < 1^\circ.25$  ( $23.75 \text{ deg}^2$ );  $\sigma$  is the significance of the detection, CE richness is the richness of the detection, and  $z$  is the color-estimated redshift of the detection.

Among the 10 false detections, three are large, bright galaxies deblended into several pieces. In the other cases, a few galaxies are seen but not an apparent cluster or group. (In one case, a rich cluster exists about  $6'$  from the false detection.) The 10 false detections are summarized in Table 5.

As the successful examples, we show two typical examples of clusters detected only with the CE method but not with the other methods (discussed below). One is a clustering of blue galaxies. Since the CE method does not reject blue spiral galaxies, it can detect clustering of several blue spiral gal-

axies. Indeed, some of the detected clusters that we visually inspected are clusterings of blue galaxies. The other is a clustering of numerous faint elliptical galaxies; in these regions, faint elliptical galaxies spread out over a large area ( $\sim 0.01 \text{ deg}^2$ ) but with no bright cluster galaxies. The CE method detects these regions successfully with a large radius. Figure 22 shows a true-color image of one of these clusters with numerous faint elliptical galaxies. Figure 23 shows a typical galaxy cluster successfully detected with the CE method.

## 6. COMPARISON WITH OTHER METHODS

At the time of this writing, the SDSS collaboration had implemented several independent cluster-finding methods and had run these algorithms on the SDSS commissioning data. These methods include the matched filter (MF) technique (Kim et al. 2002), the Voronoi tessellation (VT) technique (Kim et al. 2002), and the “maxBCG” technique (Annis et al. 2002). Therefore, we have the unique opportunity to compare the different catalogs these algorithms produce to further understand each algorithm and possible differences between them (see also Bahcall et al. 2002 for comparisons of SDSS cluster catalogs).

Here we provide a comparison between the CE method and the MF, VT, and maxBCG techniques using a small subregion of the SDSS data,  $23.75 \text{ deg}^2$  of commissioning data with right ascension between  $16^\circ$  and  $25^\circ.5$  and declination between  $-1^\circ.25$  and  $1^\circ.25$  (the region in Fig. 9). We first



FIG. 22.—Successful example of the CE method. The image is a  $6' \times 13'$  true-color image of the SDSS commissioning data. There are many faint galaxies in the region. The CE method has the ability to detect the region in the sky where many faint galaxies are clustering. This cluster was found only with the CE method.



FIG. 23.—Another successful example of the CE method. The cluster position and radius are shown with a yellow circle.

matched the CE catalog with each of the other three catalogs using a simple positional match criterion of less than 6'. The number of matches between the CE and other catalogs varies significantly because each cluster-finding algorithm has a different selection function. At present, the selection functions for all these algorithms are not fully established, so we have not corrected for them in this comparison. Although each algorithm measures cluster richness and redshift in its own way, the scatter between the measurements is large and makes the comparison difficult. Therefore, we remeasured richness and redshift of the MF, VT, and

maxBCG clusters using the CE method to see the richness and redshift dependence of the comparison.

In Table 6, we list the number of clusters each method finds in our test region (second column) and the number of clusters found in common between the CE method discussed herein and each of the other methods discussed above (third column). In Table 7, we also compare the number and percentage of matches found between the VT, MF, and maxBCG techniques. These two tables illustrate that the overlap between all four algorithms is between 20% and 60%, which is simply a product of their different selection functions. Furthermore, we note that we have used a simplistic matching criterion that does not account for the cluster redshift or the errors on the cluster centroids. Future SDSS papers will deal with these improvements (Bahcall et al. 2002). Tables 6 and 7 show that the CE and maxBCG

TABLE 6  
COMPARISON OF DETECTION METHODS

Method	Detections	Common Detections	Rate to CE (%)	Rate to the Method (%)
MF .....	152	116	32.0	76.3
maxBCG.....	438	183	50.4	41.8
VT .....	130	96	26.4	73.8
CE .....	363	...	...	...

NOTE.—The region used is  $16^\circ < \text{R.A.} < 25^\circ 5$ ,  $-1^\circ 25 < \text{decl.} < 1^\circ 25$  ( $23.75 \text{ deg}^2$ ). The third column gives the number of clusters detected by both the specified method and the CE method. The fourth column gives the proportion of the number of detections with each method to the number of detections with the CE method. The fifth column gives the proportion of the number of detections with the CE method to the number of detections with each method.

TABLE 7  
COMPARISON OF DETECTED CLUSTERS

Method	MF (152)	VT (130)	maxBCG (438)
MF .....	...	39.4% (60)	59.2% (90)
VT .....	45.5% (60)	...	65.4% (85)
maxBCG.....	20.5% (90)	19.4% (85)	...

NOTE.—Total numbers of clusters detected with each method in the region  $16^\circ < \text{R.A.} < 25^\circ 5$ ,  $-1^\circ 25 < \text{decl.} < 1^\circ 25$  ( $23.75 \text{ deg}^2$ ) are listed in parentheses in the column headings. The table lists the number of clusters detected with both methods and the percentage with respect to the methods in the first column.

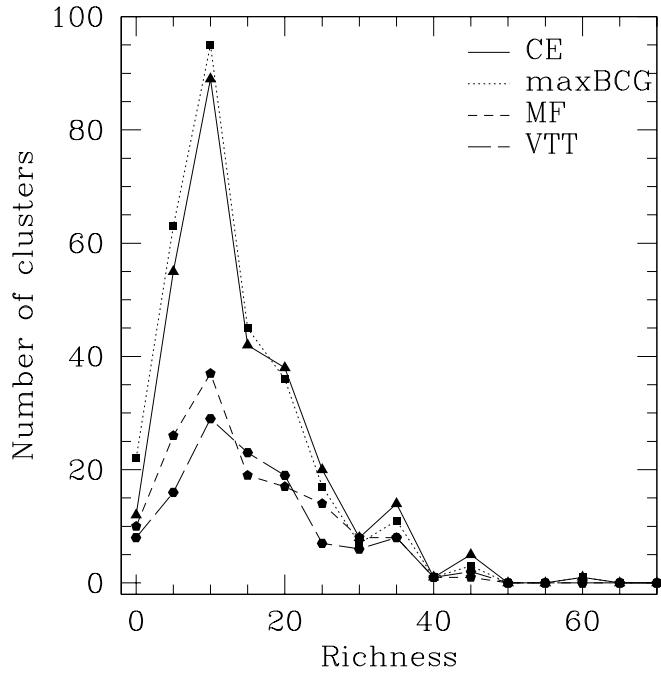


FIG. 24.—Comparison of four catalogs by richness. The abscissa is the richness of the cluster. The ordinate is the number of detected clusters. CE clusters are drawn with solid lines; maxBCG clusters are drawn with dotted lines; MF clusters are drawn with short-dashed lines; VT clusters are drawn with long-dashed lines. The CE and maxBCG methods detect poor clusters (richness less than 20) more than the MF or VT method.

methods detect more clusters overall than the other methods, i.e., 363 and 438 clusters, respectively, compared with 152 and 130 clusters for MF and VT. This difference in the number of clusters found is mainly due to differences in the thresholds used for each of these algorithms. As illustrated in Figure 24, a majority of the extra clusters in the maxBCG and CE catalogs are lower richness systems. As seen in Figure 25, these extra, lower richness systems appear to be distributed evenly over the entire redshift range of the CE catalog (i.e., out to  $z \simeq 0.4$ ).

### 6.1. Comparison of Matched Filter and Cut-and-Enhance Methods

We focus here on the comparison between the CE and MF methods (see Kim et al. 2002). In Figure 26, we show the fraction of MF clusters found in the CE catalog. We also split the sample as a function of CE richness. In Figure 27, we show the reverse relationship, the fraction of CE clusters found by the MF technique as a function of estimated redshift and CE richness. These figures show that there is almost complete overlap between the two catalogs for the highest redshift and richness systems in both catalogs (there are, however, only five  $z > 0.3$  systems in the MF catalog). At low redshifts ( $z < 0.3$ ) the overlap decreases; for example, only 60% of MF clusters are found in the CE catalog. To better understand this comparison, we visually inspected all the clusters found by the CE method that were missing from the MF catalog. As expected, most of these systems were compact ( $\sim 1'$ ) groups of galaxies.

Finally, in Figure 28 we plot the distribution of axis ratios (the major over the minor axis of the cluster) for the whole CE catalog, as well as just the CE clusters found in the MF catalog. This plot shows that a majority of clusters in both

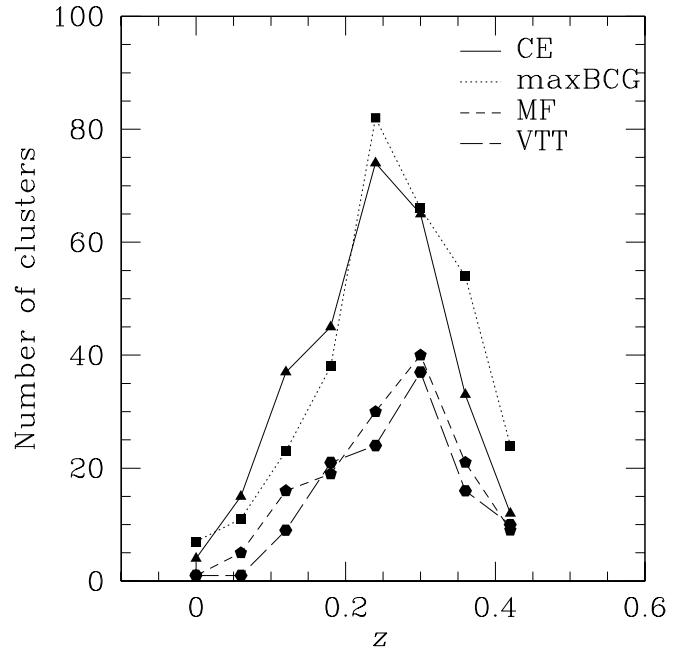


FIG. 25.—Same as Fig. 24, but for redshift. The redshift is estimated using the color (described in § 2).

samples have nearly spherical morphologies, with the two distributions in good agreement up to an axis ratio of 3:1. However, there is a tail of 11 CE clusters extending to higher axis ratios that is not seen in the CE-plus-MF subsample. However, this is only  $\sim 3\%$  of the CE clusters.

### 6.2. Comparison of maxBCG and Cut-and-Enhance Methods

In Figure 29, we show the fraction of maxBCG clusters that are found in the CE catalog, while in Figure 30 we show the reverse relationship, the fraction of CE clusters found in the maxBCG catalog. In both figures we divide the sample by estimated redshift and observed CE richness. First we note that the matching rate of maxBCG clusters to CE is  $\sim 70\%$  or better for clusters with a richness of more than 20 at all redshifts. For the lower richness systems, the matching rate decreases for all redshifts. To better understand the comparison between these two samples of clusters, we first visually inspected all clusters detected by the CE method but missing from the maxBCG sample and found them to be blue, nearby poor clusters. This is a reflection of the wider color cuts employed by the CE method, which allow the CE algorithm to include bluer, star-forming galaxies in its color criterion. The maxBCG technique, however, is tuned specifically to detect the E/S0 ridgeline of elliptical galaxies in clusters. We also visually inspected all maxBCG clusters that were not found by the CE method and found these systems to be mostly faint, higher redshift clusters whose members have mostly fallen below the magnitude limit used for the CE method ( $r^* = 21.5$ ).

### 6.3. Comparison of Voronoi Tesselation and Cut-and-Enhance Methods

In Figure 31, we show the fraction of VT clusters that were found by the CE method as a function of estimated redshift and CE richness. Figure 32 shows the fraction of CE clusters found in the VT catalog as a function of esti-

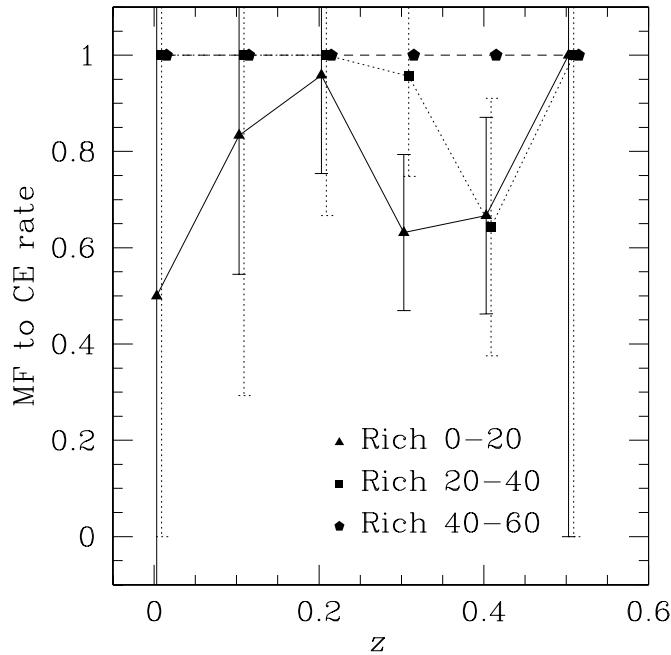


FIG. 26.—Comparison of the MF and CE methods. The abscissa is the estimated redshift. The ordinate is the ratio of the MF clusters that are found in the CE catalog to the number of the CE clusters. CE richness 0–20 is plotted with solid lines. CE richness 20–40 is plotted with dotted lines. CE richness 40–60 is plotted with dashed lines. The error bars for richness 40–60 clusters are large and omitted for clarity (at  $z = 0.3$ , the error is 80%). The data for richness 20–40 and 40–60 are shifted in the redshift direction by 0.01 for clarity.

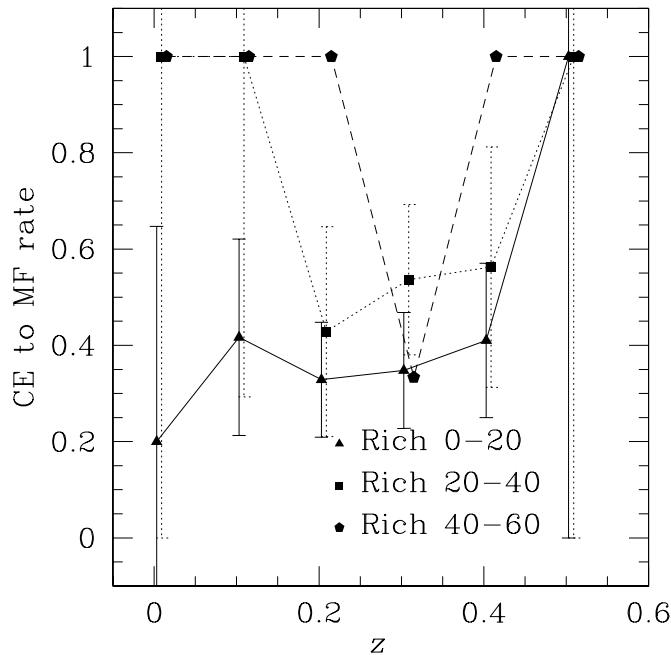


FIG. 27.—Comparison of the CE and MF methods (reverse relationship). The abscissa is the estimated redshift. The ordinate is the ratio of the CE clusters that are found in the MF catalog to the number of CE clusters. The match rate is low for poor clusters, indicating that the CE method detects poor clusters more. The error bars for richness 40–60 clusters are large and omitted for clarity (at  $z = 0.3$ , the error is 80%). The data for richness 20–40 and 40–60 are shifted in the redshift direction by 0.01 for clarity.

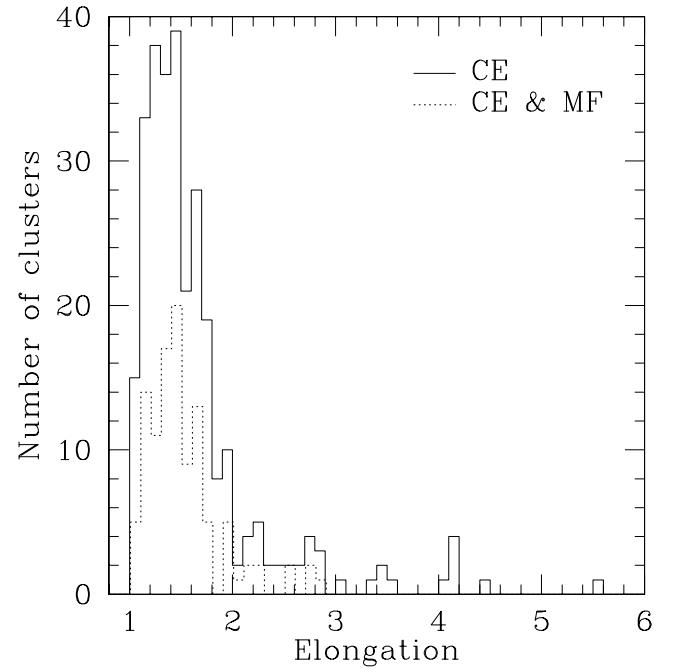


FIG. 28.—Elongation distribution of the detected clusters. The number of clusters is plotted against the elongation of clusters (ratio of the major axis to minor axis). The solid line is for the clusters detected with the CE method. The dotted line is for the clusters detected with both the MF and CE methods and is shifted by 0.01 for clarity.

estimated redshift and CE richness. Because the CE method detects twice as many clusters as the VT method, the matching rate is higher in Figure 31 than in Figure 32, showing that the CE catalog contains a high fraction of VT clusters. In Figure 32, the matching rate of low-richness clusters improves at higher redshift because the poor clusters, which VT does not detect, become fainter, and therefore both methods cannot detect these clusters at high redshift.

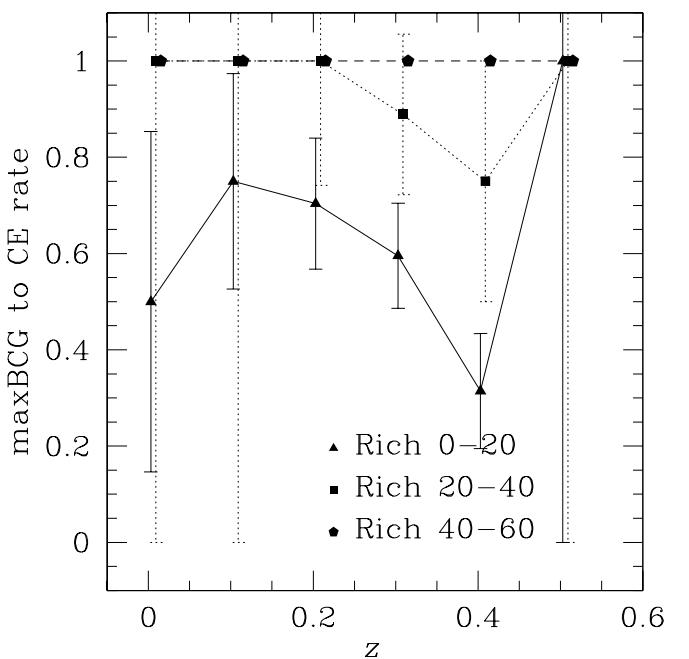


FIG. 29.—Same as Fig. 26, but for maxBCG

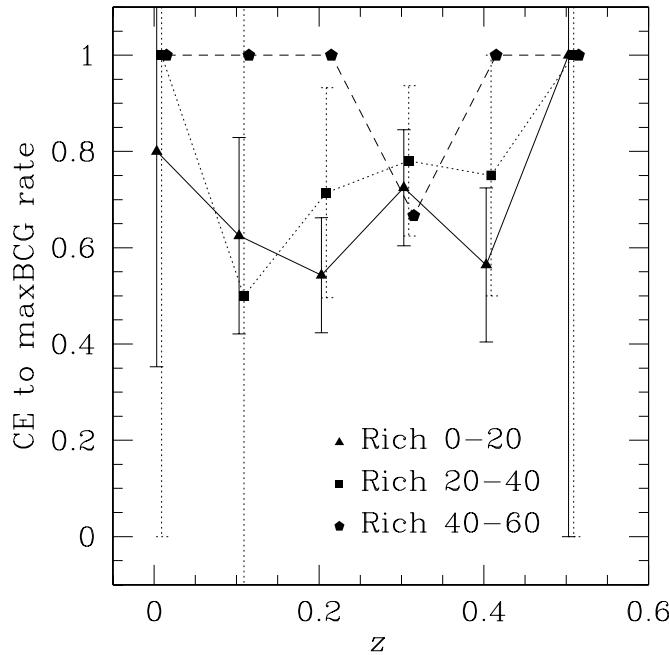


FIG. 30.—Same as Fig. 27, but for maxBCG

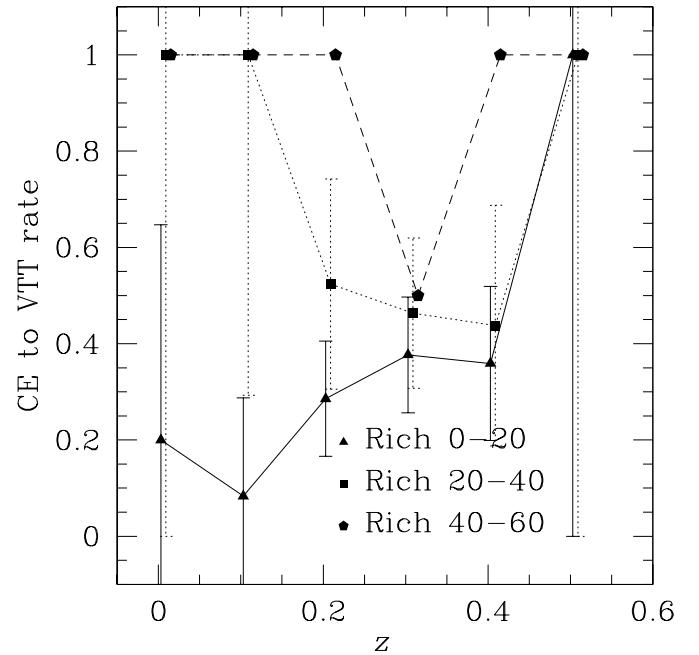


FIG. 32.—Same as Fig. 27, but for VT. Note that the CE method detects twice as many as the VT method.

## 7. SUMMARY

We have developed a new cluster-finding technique, the cut-and-enhance (CE) method. It uses 30 color cuts and four color-color cuts to enhance the contrast of galaxy clusters over the background galaxies. After applying the color and color-color cuts, the method uses the color and angular separation weight of galaxy pairs as an enhancement method to increase the signal-to-noise ratio of galaxy clusters. We use SExtractor to detect galaxy clusters from the enhanced maps. The enhancement and detection are per-

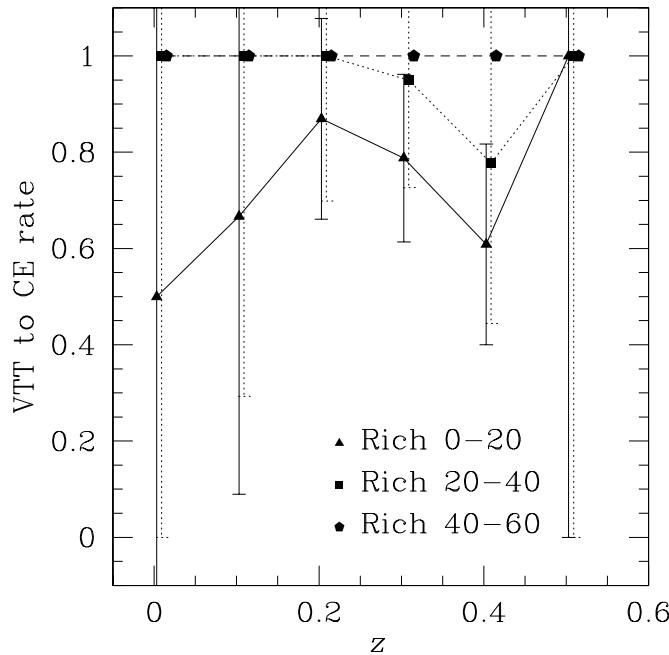


FIG. 31.—Same as Fig. 26, but for VT. Note that the CE method detects twice as many as the VT method.

formed for every color cut, producing 34 cluster lists, which are then merged into a single cluster catalog.

Using Monte Carlo simulations with a real SDSS background as well as a shuffled background, the CE method is shown to have the ability to detect rich clusters ( $N_{\text{gal}} = 100$ ) to  $z \sim 0.3$  with  $\sim 80\%$  probability. The probability drops sharply at  $z = 0.4$  because of the flux limit of the SDSS imaging data. The positional accuracy is better than  $40''$  for all richnesses examined at  $z \leq 0.3$ . The false-positive test shows that over 70% of clusters are likely to be real systems for CE richness greater than 10. We have applied the CE method to the SDSS commissioning data and produced an SDSS CE cluster catalog containing 4638 galaxy clusters in  $\sim 350 \text{ deg}^2$ . We compared the CE clusters with other cluster detection methods: MF, maxBCG, and VT. The SDSS CE cluster catalog developed in this work is a useful tool for studying both cosmology and properties of clusters and cluster galaxies.

We would like to express our sincere gratitude to Wataru Kawasaki, Kazuhiro Shimasaku, Sadanori Okamura, Mamoru Doi, Chris Miller, Francisco Castander, and Shang-Shan Chon for their suggestions and discussions to improve the work. We would also like to thank the anonymous referee, whose comments significantly improved the paper. We are grateful to all the people who helped to build the Sloan Digital Sky Survey. T. G. acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) through JSPS Research Fellowships for Young Scientists. The Sloan Digital Sky Survey is a joint project of the University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Max-Planck-Institut für Astronomie, the Max-Planck-Institut für Astrophysik, New Mexico State University, Princeton University, the US Naval Observatory, and the University of Washington. Apache Point Observatory, site of the SDSS telescopes, is operated

by the Astrophysical Research Consortium. Funding for the project has been provided by the Alfred P. Sloan Foundation, the SDSS member institutions, the National Aeronautics and Space Administration, the National Science

Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max-Planck-Gesellschaft. The SDSS Web site is <http://www.sdss.org/>.

## REFERENCES

- Abell, G. O. 1958, ApJS, 3, 211  
 Abell, G. O., Corwin, H. G., Jr., & Olowin, R. P. 1989, ApJS, 70, 1  
 Annis, J., et al. 2002, in preparation  
 Bahcall, N. A. 1988, ARA&A, 26, 631  
 Bahcall, N. A., Fan, X., & Cen, R. 1997, ApJ, 485, L53  
 Bahcall, N., et al. 2002, in preparation  
 Benítez, N., Broadhurst, T., Rosati, P., Courbin, F., Squires, G., Lidman, C., & Magain, P. 1999, ApJ, 527, 31  
 Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393  
 Bode, P., Bahcall, N. A., Ford, E. B., & Ostriker, J. P. 2001, ApJ, 551, 15  
 Borgani, S., & Guzzo, L. 2001, Nature, 409, 39  
 Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, MNRAS, 254, 601  
 Bramel, D. A., Nichol, R. C., & Pope, A. C. 2000, ApJ, 533, 601  
 Butcher, H., & Oemler, A., Jr. 1978, ApJ, 219, 18  
 —. 1984, ApJ, 285, 426  
 Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abraham, R., Gravel, P., Morris, S., & Pritchet, C. J. 1996, ApJ, 462, 32  
 Clowe, D., Trentham, N., & Tonry, J. 2001, A&A, 369, 16  
 da Costa, L., et al. 2001, in Deep Fields, ed. S. Cristiani, A. Renzini, & R. E. Williams (Berlin: Springer), (astro-ph/0012254)  
 Dressler, A. 1980, ApJ, 236, 351  
 —. 1984, ARA&A, 22, 185  
 Dressler, A., & Gunn, J. E. 1992, ApJS, 78, 1  
 Dressler, A., et al. 1997, ApJ, 490, 577  
 Eke, V. R., Cole, S., & Frenk, C. S. 1996, MNRAS, 282, 263  
 Ellis, R. S., Smail, I., Dressler, A., Couch, W. J., Oemler, A., Jr., Butcher, H., & Sharples, R. M. 1997, ApJ, 483, 582  
 Evrard, A. E., et al. 2001, ApJ, submitted  
 Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 326, 950  
 Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748  
 Fukugita, M., Shimasaku, K., & Ichikawa, T. 1995, PASP, 107, 945  
 Garilli, B., Maccagni, D., & Andreon, S. 1999, A&A, 342, 408  
 Gladders, M. D., & Yee, H. K. C. 2000, AJ, 120, 2148  
 Gunn, J. E., et al. 1998, AJ, 116, 3040  
 Hattori, M., et al. 1997, Nature, 388, 146  
 Henry, J. P. 1997, ApJ, 489, L1  
 —. 2000, ApJ, 534, 565  
 Ichikawa, S. 1986, Ann. Tokyo Astron. Obs., 21, 77  
 Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372  
 Katgert, P., Mazure, A., den Hartog, R., Adami, C., Biviano, A., & Perea, J. 1998, A&AS, 129, 399  
 Kawasaki, W., Shimasaku, K., Doi, M., & Okamura, S. 1998, A&AS, 130, 567  
 Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R., & Xu, G. 1999, ApJ, 517, 78  
 Kim, R. S. J., et al. 2002, AJ, 123, 20  
 King, I. R. 1966, AJ, 71, 64  
 Kodama, T., Arimoto, N., Barger, A. J., & Aragón-Salamanca, A. 1998, A&A, 334, 99  
 Lobo, C., Iovino, A., Lazzati, D., & Chincarini, G. 2000, A&A, 360, 896  
 Lupton, R. H., et al. 2002, in preparation  
 Nichol, R. C. 2002, in ASP Conf. Ser., Tracing Cosmic Evolution with Galaxy Clusters, ed. S. Borgani, M. Mezzetti, & R. Valdarnini (San Francisco: ASP), in press  
 Oukbir, J., & Blanchard, A. 1992, A&A, 262, L21  
 Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P., & Christensen, J. A. 1996, AJ, 111, 615  
 Reichart, D. E., Nichol, R. C., Castander, F. J., Burke, D. J., Romer, A. K., Holden, B. P., Collins, C. A., & Ulmer, M. P. 1999, ApJ, 518, 521  
 Romer, A. K., Viana, P. T. P., Liddle, A. R., & Mann, R. G. 2001, ApJ, 547, 594  
 Schuecker, P., & Böhringer, H. 1998, A&A, 339, 315  
 Shectman, S. A. 1985, ApJS, 57, 77  
 Soucail, G., Kneib, J.-P., Jaunsen, A. O., Hjorth, J., Hattori, M., & Yamada, T. 2001, A&A, 367, 741  
 Stanford, S. A., Eisenhardt, P. R., & Dickinson, M. 1998, ApJ, 492, 461  
 Stoughton, C., et al. 2002, AJ, 123, 485  
 Struble, M. F., & Rood, H. J. 1987, ApJS, 63, 555  
 —. 1999, ApJS, 125, 35  
 Viana, P. T. P., & Liddle, A. R. 1996, MNRAS, 281, 323  
 —. 1999, MNRAS, 303, 535  
 Willick, J. A., Thompson, K. L., Mathiesen, B. F., Perlmutter, S., Knop, R. A., & Hill, G. J. 2001, PASP, 113, 658  
 York, D. G., et al. 2000, AJ, 120, 1579  
 Zwicky, F., Herzog, E., Wild, P., Karpowicz, M., & Kowal, C. T. 1961–1968, Catalogue of Galaxies and of Clusters of Galaxies (6 vols.; Pasadena: Caltech)